

---


Electronic Theses and Dissertations, 2004-2019

---

2004

## Hybrid And Hierarchical Image Registration Techniques

Dongjiang Xu  
*University of Central Florida*

 Part of the [Electrical and Electronics Commons](#)  
Find similar works at: <https://stars.library.ucf.edu/etd>  
University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

---

### STARS Citation

Xu, Dongjiang, "Hybrid And Hierarchical Image Registration Techniques" (2004). *Electronic Theses and Dissertations, 2004-2019*. 264.  
<https://stars.library.ucf.edu/etd/264>

# HYBRID AND HIERARCHICAL IMAGE REGISTRATION TECHNIQUES

by

DONGJIANG XU

B.S. EE, Northeast China Institute of Electric Power Engineering, P.R. China, 1994

M.S. EE, Tsinghua University, P.R. China, 1998

A dissertation submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy  
in the Department of Electrical and Computer Engineering  
in the College of Engineering & Computer Science  
at the University of Central Florida  
Orlando, Florida

Fall Term  
2004

Advisor: Dr. Takis C. Kasparis

© Copyright 2004, Dongjiang Xu  
All Rights Reserved

## ABSTRACT

A large number of image registration techniques have been developed for various types of sensors and applications, with the aim to improve the accuracy, computational complexity, generality, and robustness. They can be broadly classified into two categories: intensity-based and feature-based methods. The primary drawback of the intensity-based approaches is that it may fail unless the two images are misaligned by a moderate difference in scale, rotation, and translation. In addition, intensity-based methods lack the robustness in the presence of *non-spatial distortions* due to different imaging conditions between images.

In this dissertation, the image registration is formulated as a two-stage hybrid approach combining both an initial matching and a final matching in a coarse-to-fine manner. In the proposed hybrid framework, the initial matching algorithm is applied at the coarsest scale of images, where the approximate transformation parameters could be first estimated. Subsequently, the robust gradient-based estimation algorithm is incorporated into the proposed hybrid approach using a multi-resolution scheme. Several novel and effective initial matching algorithms have been proposed for the first stage. The variations of the intensity characteristics between images may be large and non-uniform because of *non-spatial distortions*. Therefore, in order to effectively incorporate the gradient-based robust estimation into our proposed framework, one of the

fundamental questions should be addressed: what is a good image representation to work with using gradient-based robust estimation under *non-spatial distortions*.

With the initial matching algorithms applied at the highest level of decomposition, the proposed hybrid approach exhibits superior range of convergence. The gradient-based algorithms in the second stage yield a robust solution that precisely registers images with sub-pixel accuracy. A hierarchical iterative searching further enhances the convergence range and rate. The simulation results demonstrated that the proposed techniques provide significant benefits to the performance of image registration.

This work is dedicated to my parents.

## **ACKNOWLEDGMENTS**

First, I would like to express my sincere respect and appreciation to my research advisor, Dr. Takis C. Kasparis. His valuable guidance and great wisdom helped me overcome all problems and difficulties that I have encountered through the course of my graduate studies and work on this research. I would also like to extend my gratitude to Dr. Mubarak Shah, Dr. Linwood Jones, Dr. Xiaogang Su, and Dr. Michael G. Haralambous.

## TABLE OF CONTENTS

CHAPTER ONE : INTRODUCTION	1
CHAPTER TWO : IMAGE REGISTRATION FUNDAMENTALS	9
Definition .....	9
Transformations .....	10
Image Variations .....	14
Rectification .....	17
CHAPTER THREE : PROBLEM FORMULATION	20
Difficulties in Image Registration .....	20
Assumptions .....	27
CHAPTER FOUR : FEATURE EXTRACTION	28
Introduction .....	28
RCBS-based Edge Contour Extraction Algorithm .....	32
Gradient and Orientation Estimation .....	33
Edge Detection by Regularized Cubic B-Spline Fitting .....	37
Edge Contours: Post-processing of RCBS Edge Maps .....	43
Experiment Results .....	44
CHAPTER FIVE : INITIAL MATCHING	48
Phase Correlation: A FFT-based Image Registration .....	48



A Point Pattern Matching Algorithm.....	56
A Contour-based Algorithm .....	62
A Point-based Algorithm using Gradient Vector Field .....	65
Simulation Results .....	66
CHAPTER SIX : GRADIENT-BASED MOTION ESTIMATION .....	70
Model-based Registration .....	70
Coarse-to-Fine Iterative Estimation.....	73
CHAPTER SEVEN : ROBUST ESTIMATION UNDER NON-UNIFORM BRIGHTNESS VARIATION .....	80
Introduction.....	80
A Robust Image Registration Algorithm .....	82
Image Representation.....	84
Robust Estimation using a Direct Method .....	85
The Hierarchical Algorithm and Selective Data Sampling .....	92
Simulation Results .....	94
Experiment Scheme I.....	95
Experiment Scheme II.....	97
Experiment Scheme III .....	101
Experiment Scheme IV .....	102
CHAPTER EIGHT : CORRELATION-BASED IMAGE REGISTRATION .....	109
Matching Algorithm.....	109
Simulation Results .....	112

CHAPTER NINE : CONCLUSIONS	117
LIST OF REFERENCES	120

## LIST OF FIGURES

<b>Figure 1</b> A Registration Example.....	5
<b>Figure 2.</b> Rotation effects on correlation coefficient. ....	22
<b>Figure 3.</b> Rotation effects on correlation coefficient. ....	22
<b>Figure 4.</b> The proposed image registration framework.....	26
<b>Figure 5.</b> Block Diagram of Edge Contour Extraction Algorithm.....	33
<b>Figure 6.</b> Comparison on synthetic images using different gradient operators. ....	36
<b>Figure 7.</b> Comparison between three gradient operators ( $\sigma = 20$ ).....	38
<b>Figure 8.</b> Comparison between three gradient operators ( $\sigma = 50$ ).....	39
<b>Figure 9.</b> (a) Cubic B-spline; (b) RCBS mask. ....	39
<b>Figure 10.</b> Block Diagram of a B-spline differentiator using the smoothing B-spline kernels. ....	42
<b>Figure 11.</b> Examples of edge contour extraction by (a) LoG ( $\sigma=1$ ), (b) LoG ( $\sigma=3$ ), (c) RCBS ( $\lambda=1.0E-7$ ), (d) RCBS ( $\lambda=1.0E-1$ ).....	45
<b>Figure 12.</b> Block diagram of phase correlation for angle and scale estimation. ....	51
<b>Figure 13.</b> (a)(c): Phase correlation for rotation and scale between log-polar mapping at the levels of 64x64 and 128x128, respectively; (b)(d): Phase correlation for translation after compensation at the levels of 64x64 and 128x128, respectively....	52
<b>Figure 14</b> Comparison of rotation-scale peak between 64x64 (sub 64) and 128x128 (sub	

128) levels, respectively.....	53
<b>Figure 15.</b> Comparison of translation peak between the 64x64 (sub 64) and 128x128 (sub 128) levels, respectively.....	53
<b>Figure 16.</b> Two point patterns.....	59
<b>Figure 17.</b> Edge contours at the coarsest scale.....	63
<b>Figure 18.</b> Image pair for gradient vector field.....	66
<b>Figure 19.</b> Gradient vector fields. ....	67
<b>Figure 20.</b> 2D histogram of gradient vector fields.....	68
<b>Figure 21.</b> Estimate rotation by correlation. ....	69
<b>Figure 22.</b> PSNR for synthetic images.....	77
<b>Figure 23.</b> PSNR for aerial images. ....	77
<b>Figure 24.</b> Image registration examples.....	79
<b>Figure 25.</b> Robust estimation block diagram. ....	83
<b>Figure 26.</b> Affine Registration Example I.....	95
<b>Figure 27.</b> Extracted feature points at the coarsest scale marked by “+”. ....	96
<b>Figure 28.</b> Image Difference .....	97
<b>Figure 29.</b> Affine Image Registration .....	98
<b>Figure 30.</b> Image Differences.....	99
<b>Figure 31.</b> Video Frame Registration Example .....	101
<b>Figure 32.</b> Affine Registration Example (Landsat TM images) .....	104
<b>Figure 33.</b> Extracted feature points at the coarsest scale marked by “+”. ....	105
<b>Figure 34.</b> Affine Image Registration (Landsat TM Images) .....	106

<b>Figure 35:</b> An aerial video frame pairs. ....	113
<b>Figure 36.</b> Image registration based on initial matching.....	114
<b>Figure 37.</b> Image registration using the projective model .....	115

## LIST OF TABLES

<b>Table 1.</b> Registration results using PCME at the coarsest level of 64x64. ....	54
<b>Table 2.</b> Success ratio of some experiments. ....	60
<b>Table 3.</b> NCC Comparison between the BVM-based approach and the proposed method. .....	100
<b>Table 4.</b> NCC Comparison between the BVM-based approach and the proposed method. .....	102
<b>Table 5.</b> NCC Comparison using our method. ....	107
<b>Table 6.</b> NCC comparison based on the proposed method. ....	116

## CHAPTER ONE: INTRODUCTION

Image registration [1,2] is the process of spatially matching two or more images of the same scene, taken at different times, from different viewpoints and/or by different sensors. The registration problem is to find the transformation, which bring images into geometric alignment so that the points in one image can be related to their corresponding points in the other.

The present *differences* between images are introduced due to different imaging conditions. Image registration is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources like in image fusion, change detection, and multi-channel image restoration. Typically, the need to register images has arisen in the fields of computer vision and pattern recognition (target localization, automatic quality control), medical image analysis (combing computer tomography and NMR data to obtain more complete information about the patient, monitoring tumor growth, treatment verification, comparison of the patient's data with anatomical atlases), and remotely sensed data processing (multi-spectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images, integrating information into geographic information systems), etc. [1,2].

During the last decades, image acquisition devices have undergone rapid

development and growing amount and diversity of obtained images invoked the research on automatic image registration. A comprehensive survey of image registration techniques was published in 1992 by Brown [1].

Image registration, as it was mentioned above, is widely used in remote sensing, medical imaging, computer vision etc. In general, its applications can be divided into four main groups according to the manner of the image acquisition [2]:

*Different viewpoints* (multi-view analysis): images of the same scene are acquired from different viewpoints. The aim is to gain larger a 2D view or a 3D representation of the scanned scene.

Examples of applications: Remote sensing — mosaicing of images of the surveyed area. Computer vision — shape recovery (shape from stereo).

*Different times* (multi-temporal analysis): images of the same scene are acquired at different times, often on regular basis, and possibly under different conditions. The aim is to find and evaluate changes in the scene, which appeared between the consecutive image acquisitions.

Examples of applications: Remote sensing — monitoring of global land usage, landscape planning. Computer vision — automatic change detection for security monitoring, motion tracking. Medical imaging — monitoring of the healing therapy, monitoring of the tumor evolution.

*Different sensors* (multi-modal analysis): images of the same scene are acquired by different sensors. The aim is to integrate the information obtained from different source streams to gain more complex and detailed scene representation.



Examples of applications: Remote sensing — fusion of information from sensors with different characteristics like panchromatic images, offering better spatial resolution, color/multi-spectral images with better spectral resolution, or radar images independent of cloud cover and solar illumination. Medical imaging — combination of sensors recording the anatomical body structure like magnetic resonance image (MRI), ultrasound or CT with sensors monitoring functional and metabolic body activities like positron emission tomography (PET), single photon emission computed tomography (SPECT) or magnetic resonance spectroscopy (MRS). Results can be applied, for instance, in radiotherapy and nuclear medicine.

*Scene to model registration:* images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance maps or digital elevation models (DEM) in GIS, another scene with similar content (another patient), ‘average’ specimen, etc. The aim is to localize the acquired image in the scene/model and/or to compare them.

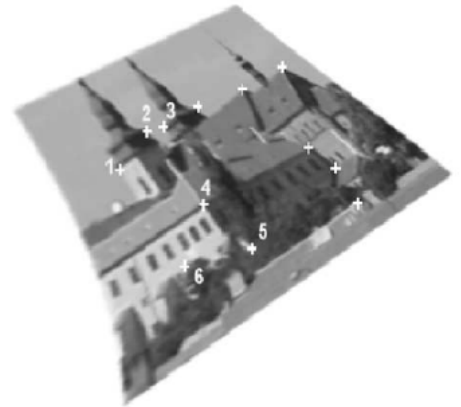
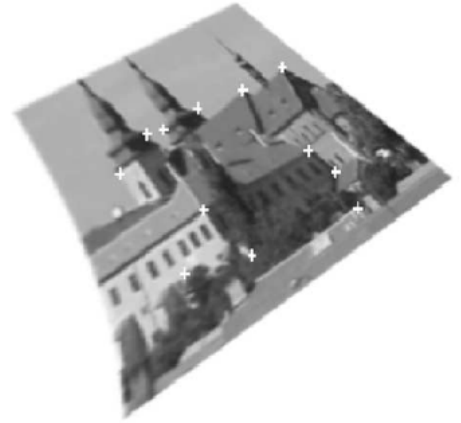
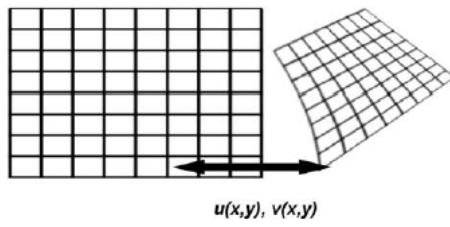
Examples of applications: Remote sensing — registration of aerial or satellite data into maps or other GIS layers. Computer vision — target template matching with real-time images, automatic quality inspection. Medical imaging — comparison of the patient’s image with digital anatomical atlases, specimen classification.

Due to the diversity of images to be registered and due to various types of degradations it is impossible to design a universal method applicable to all registration tasks. Every method should take into account not only the assumed type of geometric deformation between the images but also radiometric deformations and noise corruption,

required registration accuracy and application-dependent data characteristics.

Nevertheless, the majority of the registration methods consist of the following four components [1] (see Fig. 1):

1. *Feature Space*. It extracts the information in the images that will be used for matching. For example, consider the problem of registering the two images taken of the same building at different times shown in Fig. 1. A standard approach to registration for the images might be as follows: feature detection is first performed. Salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or, preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), which are called control points (CPs) in the literature. This removes extraneous information and reduces the amount of data to be evaluated.
2. *Search Strategy*. It decides how to choose the next transformation from this space, to be tested in the search for the optimal transformation. For feature matching, the correspondence between the features detected in the sensed image and those detected in the reference image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose.



**Figure 1** A Registration Example.

3. *Search Space*, i.e., transform model estimation. It is the class of transformations that is capable of aligning the images. The type and parameters of the so-called mapping functions, aligning the sensed image with the reference image, are estimated. The parameters of the mapping functions can be computed by means of the established feature correspondence.
4. *Similarity metric*. It determines the relative merit for each test. Search continues according to the search strategy until a transformation is found whose similarity measure is satisfactory. The sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique.

As we shall see, the types of variations present in the images will determine the selection for each of these components. The implementation of each registration step has its typical problems. For feature extraction (*feature space*), we have to decide what kind of features is appropriate for the given task. The features should be distinctive objects, which are frequently spread over the images and which are easily detectable. Usually, the physical interpretability of the features is demanded. The detected feature sets in the reference and sensed images must have enough common elements, even in situations when the images do not cover exactly the same scene or when there are object occlusions or other unexpected changes. The detection methods should have good localization accuracy and should not be sensitive to the assumed image degradation. In an ideal case, the algorithm should be able to detect the same features in all projections of the scene regardless of the particular image deformation.

In the feature-matching step, problems caused by incorrect feature detection or by image degradations can arise. Physically corresponding features can be dissimilar due to the different imaging conditions and/or due to the different spectral sensitivity of the sensors. The choice of the feature description and similarity measure (*search strategy*) has to consider these factors. The feature descriptors should be invariant to the assumed degradations. Simultaneously, they have to be discriminable enough to be able to distinguish among different features as well as sufficiently stable so as not to be influenced by slight unexpected feature variations and noise. The matching algorithm in the space of invariants should be robust and efficient. Single features without corresponding counterparts in the other image should not affect its performance.

The type of the mapping functions (*search space*) should be chosen according to the *a priori* known information about the acquisition process and expected image degradations. If no *a priori* information is available, the model should be flexible and general enough to handle all possible degradations that might appear. The accuracy of the feature detection method, the reliability of feature correspondence estimation, and the acceptable approximation error need to be considered too. Moreover, the decision about which differences between images have to be removed by registration has to be done. It is desirable not to remove the differences we are searching for if the aim is a change detection. This issue is very important and extremely difficult. Finally, the choice of the appropriate type of re-sampling technique depends on the trade-off between the demanded accuracy of the interpolation and the computational complexity. The nearest-neighbor or bilinear interpolation is sufficient in most cases; however, some applications

require more precise methods.

Because of its importance in various application areas as well as because of its complicated nature, image registration has been the topic of much recent research. The historically first survey paper covers mainly the methods based on image correlation. Probably the most exhaustive review of the general-purpose image registration methods is in [1]. Registration techniques applied particularly in medical imaging are summarized in [3-6]. In [7] the surface-based registration methods in medical imaging are reviewed. Volume-based registration is reviewed in [8]. The registration methods applied mainly in remote sensing are described and evaluated in [9-11].

Registration methods can be categorized with respect to various criteria. The ones usually used are the application area, dimensionality of data, type and complexity of assumed image deformations, computational cost, and the essential ideas of the registration algorithm. Here, the classification according to the essential ideas is chosen, considering the decomposition of the registration into the described four components.

## CHAPTER TWO: IMAGE REGISTRATION FUNDAMENTALS

In this chapter, image registration fundamentals will be introduced, including definition, transformation, image variation, and rectification.

### **Definition**

Image registration can be defined as a mapping between two images both spatially and with respect to intensity [1]. If we define these images as two 2D arrays of a given size denoted by  $I_1$  and  $I_2$ , where  $I_1(x,y)$  and  $I_2(x,y)$  each map to their respective intensity (or other measurement) values, then the mapping between images can be expressed as:

$$I_2(x,y) = g(I_1(f(x,y))) \quad (2.1)$$

where  $f$  is a 2D spatial-coordinate transformation, i.e., is a transformation in Eq. (2.1), which maps two spatial coordinates,  $x$  and  $y$ , to new spatial coordinates  $x'$  and  $y'$ ,

$$(x',y') = f(x,y) \quad (2.2)$$

and  $g$  is a 1D intensity or radiometric transformation. The registration problem is to find the optimal spatial and intensity transformations so that the images are matched either for

the purposes of determining the parameters of the matching transformation or to expose differences of interest between the images. The intensity transformation is not always necessary.

### **Transformations**

The fundamental characteristic of any image registration technique is the type of spatial transformation or mapping used to properly overlay two images. Although many types of variations may be present in each image, the registration technique must select the class of transformation that will remove only the spatial distortions between images due to differences in acquisition and scene characteristics that affect acquisition. Other differences in scene characteristics that are to be exposed by registration should not be used to select the class of transformation. In this section, we will define several types of transformations and their parameters.

The most common general transformations are rigid, affine, projective, perspective, and global polynomial. Rigid transformations account for object or sensor movement in which objects in the images retain their relative shape and size. A rigid-body transformation is composed of a combination of a rotation, a translation, and a scale change. Affine transformations are more general than rigid and can therefore tolerate more complicated distortions while still maintaining some nice mathematical properties. Projective transformations and the more general perspective transformations account for distortions due to the projection of objects at varying distances to the sensor onto the



image plane. In order to use the perspective transformation for registration, knowledge of the distance of the objects of the scene relative to the sensor is needed. Polynomial transformations are one of the most general global transformations, and can account for many types of distortions so long as the distortions do not vary too much over the image. For example, distortions due to moderate terrain relief can often be corrected by a polynomial transformation.

In this section we will briefly define the different transformation classes and their properties. A transformation  $T$  is linear if,

$$T(x_1 + x_2) = T(x_1) + T(x_2) \quad (2.3)$$

and for every constant  $c$ ,

$$cT(x) = T(cx) \quad (2.4)$$

A transformation is *affine* if  $T(x) - T(0)$  is linear. Affine transformations are linear in the sense that they map straight lines into straight lines. The most commonly used registration transformation is the affine transformation, which is sufficient to match two images of a scene taken from the same viewing angle but from a different position, i.e., the camera can be moved, and it can be rotated around its optical axis. This affine transformation is composed of the Cartesian operations of a scaling, a translation, and a rotation. It is a global transformation, which is *rigid* since the overall geometric relationships between points do not change, i.e., a triangle in one image maps into a

similar triangle in the second image. It typically has four parameters,  $t_x$ ,  $t_y$ ,  $s$ ,  $\theta$ , which map a point  $(x_1, y_1)$  of the first image to a point  $(x_2, y_2)$  of the second image as follows:

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = s \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (2.5)$$

Since the rotation matrix is orthogonal (the rows or columns are perpendicular to each other), the angles and lengths in the original image are preserved after the registration. Because of the scalar scale factor  $s$ , the rigid-body transformation allows changes in length relative to the original image, but it is the same in both  $x$  and  $y$ . Without the addition of the translation vector, the transformation becomes linear.

The general 2D affine transformation

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} \quad (2.6)$$

does not have the properties associated with the orthogonal rotation matrix. Angles and lengths are no longer preserved, but parallel lines do remain parallel. The general affine transformation can account for more general spatial distortions such as shear (sometimes called skew) and changes in aspect ratio. The perspective transformation accounts for the distortion, which occurs when a 3D scene is projected through an idealized optical image system. This is a mapping from 3D to 2D. In the special case, where the scene is a flat plane such as in an aerial photograph, the distortion is accounted for by a projective transformation. Perspective distortions cause imagery to appear smaller the farther it is

from the camera and more compressed the more it is inclined away from the camera. The latter effect is sometimes called *foreshortening*. If the coordinates of the objects in the scene are known, say  $(x_0, y_0, z_0)$ , then the corresponding point in the image  $(x_i, y_i)$  is given by

$$x_i = \frac{-fx_0}{z_0 - f}, \quad y_i = \frac{-fy_0}{z_0 - f} \quad (2.7)$$

where  $f$  is the position of the center of the camera lens. (If the camera is in focus for distant objects,  $f$  is the focal length of the lens.) In the special case where the scene is composed of a flat plane tilted with respect to the image plane, the projective transformation is needed to map the scene plane into an image that is tilt free and of a desired scale. This process, called *rectification*, is described in more detail in the next Section. The projective transformation maps a coordinate on the plane  $(x_p, y_p)$  to a coordinate in the image  $(x_l, y_l)$  as follows:

$$\begin{aligned} x_l &= \frac{a_{11}x_p + a_{12}y_p + a_{13}}{a_{31}x_p + a_{32}y_p + a_{33}} \\ y_l &= \frac{a_{21}x_p + a_{22}y_p + a_{23}}{a_{31}x_p + a_{32}y_p + a_{33}} \end{aligned} \quad (2.8)$$

where the **a** terms are constants, which depend on the equations of the scene and image plane. If these transformations do not account for the distortions in the scene or if not enough information is known about the camera geometry, global alignment can be determined using a polynomial transformation. For perspective distortion of complex 3D

scenes, or nonlinear distortions due to the sensor, object deformations and movements and other domain-specific factors, local transformations are necessary. These can be constructed via piecewise interpolation, e.g., splines when matched features are known, or model-based techniques such as elastic warping and object/motion models.

### **Image Variations**

Since image registration deals with the removal of distortions and the detection of changes between images, knowledge about the types of variations between images plays a fundamental role in any registration problem. We have found it useful to categorize these variations in the images into three groups based on their different roles in registration problems.

First, it is important to distinguish between distortions and other variations. Distortions are variations, which are the source of mis-registration. By this, we mean they are variations, which have caused the images to be misaligned and have obscured the true measurement values. It is the distortions between images, which we would like to remove by registration. The other variations are usually changes that we are interested in detecting after registration has been performed; they are therefore not distortions. Distortions may be due to a change in the sensor viewpoint, noise introduced by the sensor or its operation, changes in the subject's position, and other undesirable changes in the scene or sensor. They almost always arise from differences in the way or the circumstances under which the images are acquired. This is in contrast to variations of

interest, which stem from intrinsic differences in the scene, such as physical growths or movements.

Second, we distinguish two categories of distortions. In any registration problem, we would like to remove all the distortions possible. However, this is seldom possible or practical. What is typically done instead is remove the primary *spatial discrepancies* and to limit the influence of volumetric and small local errors. This is accomplished by choosing a viable spatial transformation class and by ignoring other variations by choosing the appropriate feature space, similarity measure, and search strategy. This effectively splits the distortions into two categories. The first category is the spatial distortions that can be satisfactorily modeled by a practical transformation class. We call these the *corrected distortions*. The remaining distortions are often caused by lighting and atmospheric changes. This is because their effects depend on the characteristics of the physical objects in the scene, and hence they are difficult to model effectively.

In summary, there are three categories of variations that play important roles in the registration of images. The first type (Type I) is the variations, usually spatial, which are used to determine an appropriate transformation. Since the application of an optimal transformation in this class will remove these distortions, they are called corrected distortions. The second type of variations (Type II) are also distortions, usually volumetric, but distortions which are not corrected by the registration transformation. We call these uncorrected distortions. Finally, the third type (Type III) is variations of interest, differences between the images, which may be spatial or volumetric but are not to be removed by registration. Both the uncorrected distortions and the variations of interest,

which together we call uncorrected variations, affect the choice of feature space, similarity measure, and search strategy that make up the final registration method. The distinction between uncorrected distortions and variations of interest is important, especially in the case where both the distortions and the variations of interest are local, because the registration method must address the problem of removing as many of the distortions as possible while leaving the variations of interest intact.

Not surprisingly, the more that is known about the type of distortions present in a particular system, the more effective registration can be. In computer vision, images with different viewing geometries, such as stereo image pairs, are “registered” to determine the depth of objects in the scene or their 3D shape characteristics. This requires matching features in the images and finding the disparity between them; this is often called the correspondence problem. In this case, the majority of the variations are corrected by the mapping between images, but on the other hand the resulting mapping is highly complex. Consider the problems of occlusion, the different relative position of imaged objects and the complete unpredictability of the mapping because of the unknown depths and shapes of objects in the scene. Hence, problems of stereo matching and motion tracking also have a real need to model the source of mis-registration. By exploiting camera and object model characteristics such as viewing geometry, smooth surfaces, and small motions, these registration-like techniques become very specialized. For example, in stereo mapping, images differ by their imaging viewpoint, and therefore the source of mis-registration is due to differences in perspective. This greatly reduces the possible transformations and allows registration methods to exploit the properties of stereo

imagery. Because of the geometry imposed by the camera viewpoints, the location of any point in one image constrains the location of the point in the other image, which represents the same point in the 3D scene, to a line. This is called the epipolar constraint, and the line in which the matching point must lie is called the epipolar line. If the surfaces in the scene are opaque, continuous and if their scanlines (the rows of pixels in the image) are parallel to the baseline (the line connecting their two viewpoints), then an ordering constraint is also imposed along corresponding epipolar lines. Furthermore, the gradient of the disparity (the change in the difference in position between the two images of a projected point) is directly related to the smoothness of surfaces in the scene. By using these constraints instead of looking for an arbitrary transformation with a general registration method, the stereo correspondence problem can be solved more directly, i.e., search is more efficient and intelligent. When sufficient information about the mis-registration source is available, it may be possible to register images analytically and statically.

### **Rectification**

One of the simplest types of registration can be performed when the scene under observation is relatively flat and the viewing geometry is known. The former condition is often the case in remote sensing if the altitude is sufficiently high. This type of registration is accomplished by rectification, i.e., the process that corrects for the perspective distortion in an image of a flat scene. Rectification is often performed to

correct images so that they conform to a specific map standard such as the Universal Transverse Mercator projection. But it can also be used to register two images of a flat surface taken from different viewpoints. Given an imaging system in which the image center  $O$  is at the origin and the lens center  $L$  is at  $(0,0,f)$ , any scene point  $P_0 = (x_0, y_0, z_0)$  can be mapped to an image point  $P_l = (x_l, y_l)$  by the scale factor  $f/(z_0 - f)$ . If the scene is a flat plane, which is perpendicular to the camera axis (i.e.,  $z$  is constant) it is already rectified since the scale factor is now constant for all points in the image. For any other flat plane  $S$ , given by

$$Ax_0 + By_0 + z_0 = C \quad (2.9)$$

where  $A$ ,  $B$ , and  $C$  are constants, rectification can be performed by mapping the intensity of the image point at  $(x_l, y_l)$  into the new rectified image point location  $(fx_l/Z, fy_l/Z)$  where  $Z = f - Ax_l - By_l$  [12]. This is because the scene plane can be decomposed into lines  $Ax_0 + By_0 = C'$  each at a constant distance  $(z_0 = C - C')$  from the image plane. Each line then maps to a line in the image plane, and since its perspective distortion is related to its distance from the image, all points on this line must be scaled accordingly by  $f/(C - C' - f)$ . The following steps can register two pictures of the flat plane, taken from different viewpoints. First, the scene coordinates  $(x_1, y_1, z_1)$  are related to their image coordinates in image 1 of a point with respect to camera 1 by a scale factor  $(z_1 - f)/f$  dependent on their depth (the  $z_1$  coordinate) and the lens center  $f$  because of



similar triangles. This gives us two equations. Since they must also satisfy the equation of the plane, we have three equations from which we can derive the three coordinates of each scene point using its corresponding image point with respect to coordinate system of camera 1. The scene coordinates are then converted from the coordinate system with respect to camera 1 to a coordinate system with respect to camera 2 to obtain  $(x_2, y_2, z_2)$ . Lastly, these can be projected onto image 2 by the factor  $f/(z_2 - f)$ , again by similar triangles. Of course, if these are discrete images, there is still the problem of interpolation if the registered points do not fall on grid locations.

## CHAPTER THREE: PROBLEM FORMULATION

This chapter first reviews current difficulties in image registration. A hybrid and hierarchical image registration framework is proposed, which consists of two stages: *initial matching* and *final matching*. The purpose of the initial matching is to provide a good initial estimate to the second stage of final matching. In the second stage, gradient-based algorithms are incorporated to precisely register images using a multi-resolution method. This formulation is used in subsequent Chapters as a starting point to examine the performance of image registration algorithms under different conditions.

### **Difficulties in Image Registration**

A large number of image registration techniques have been developed for various types of sensors and applications, with the aim to improve the accuracy, computational complexity, generality, and robustness. They can be broadly classified into two categories:

1. *Intensity-based* Methods
2. *Feature-based* Methods

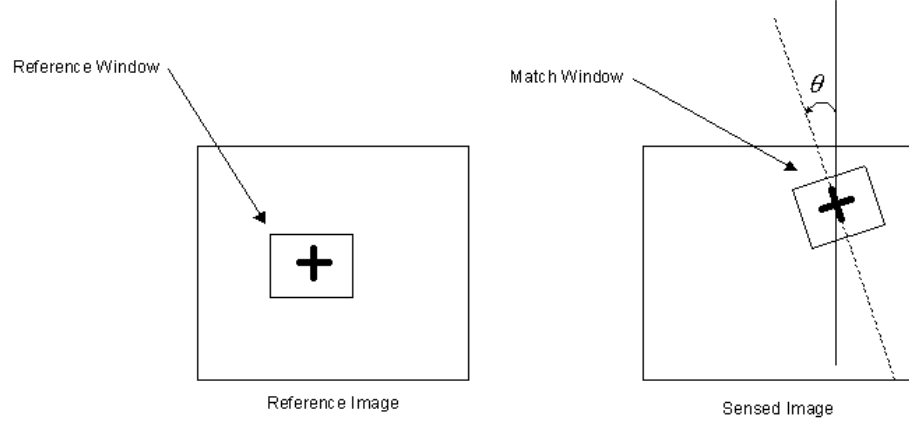
Conventional intensity-based methods use the correlation function as a measure of match, where a small patch of rectangular window in one image is statistically compared with a patch of the same size in the other image. The correlation process generally

produces a correlation surface that is irregular in shape and usually has multiple peaks. The centers of the matched windows are control points, which can be used to solve for the transformation parameters between the two images. By itself, the cross-correlation coefficient given in Eq. (3.1) is not a registration method. It is a similarity measure or match metric, i.e., it gives a measure of the degree of similarity between an image and a template. The main advantage of the correlation function is that the correlation based methods deals with the images without attempting to detect salient features. But the value of correlation is mainly governed by the multiplication between intensity pixels, and so has less robustness for image distortions.

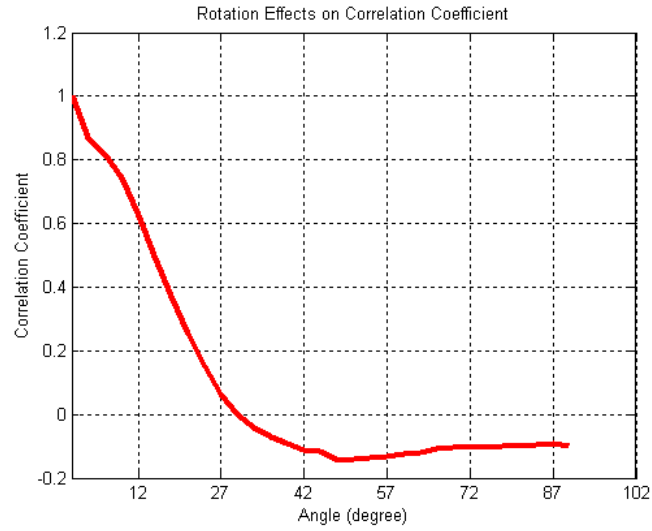
$$Corr = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}} \quad (3.1)$$

Another difficulty in correlation-based methods originates in their basic idea. Firstly, the rectangular window, which is most often used, suits the registration of images that locally differ only by a translation. If images are deformed by more complex transformations, this type of the window is not able to cover the same parts of the scene in the reference and sensed images (the rectangle can be transformed to some other shape). Several researchers proposed to use circular shape of the window for mutually rotated images. However, the comparability of such simple-shaped windows is violated too if more complicated geometric deformations (similarity, perspective transformations, etc.) are present between images. For similarity transformations characterized by four

parameters (rotation  $\theta$ , translations  $\Delta x$ ,  $\Delta y$ , and scale  $s \geq 0$ ), the value of correlation will be greatly influenced and the correspondence is hard to establish if the rotation  $\theta$  or scale  $s$  parameters between images are large, which is shown in Fig. 2 and Fig. 3.



**Figure 2.** Rotation effects on correlation coefficient.



**Figure 3.** Rotation effects on correlation coefficient.

In summary, classical intensity-based methods like the normalized cross-correlation exploit for matching directly image intensities, without any structural analysis. Consequently, they are sensitive to the intensity changes, introduced for instance by noise, varying illumination, and/or by using different sensor types.

In contrast, feature-based methods [13,14] extract and match the corresponding structures from images. There are generally two critical procedures involved in the feature-based methods, namely feature extraction and correspondence establishment. Feature maps, which include salient points, edge segments, boundaries of objects and regions, etc., provide a concise and accurate representation of an image. Such correspondence-based methods first employ feature matching techniques to determine corresponding feature pairs from the two images, and then compute the geometric transformation relating them, typically using a least squares approach. Their primary advantage is that the transformation parameters can be computed in a single step, and are accurate if the feature matching is reliable. Features represent information on a higher level, which makes feature-based methods suitable for situations under *non-spatial distortions*. However, obtaining correct matches of features is a hard problem, and could be computationally expensive if there involves a large number of feature candidates to be matched. Some correspondence-less registration methods assume the geometric information contained in the features is sufficient to establish the correspondence between the images and determine the geometric transformation that aligns the reference image with the sensed image [13]. However, the presence of textural patterns and noise in real images make the tasks of the detection of contour information and the curvilinear

linking process unnecessarily harder. Sophisticated heuristics are always needed to estimate geometric properties of image contours and perform edge linking for unbroken contours.

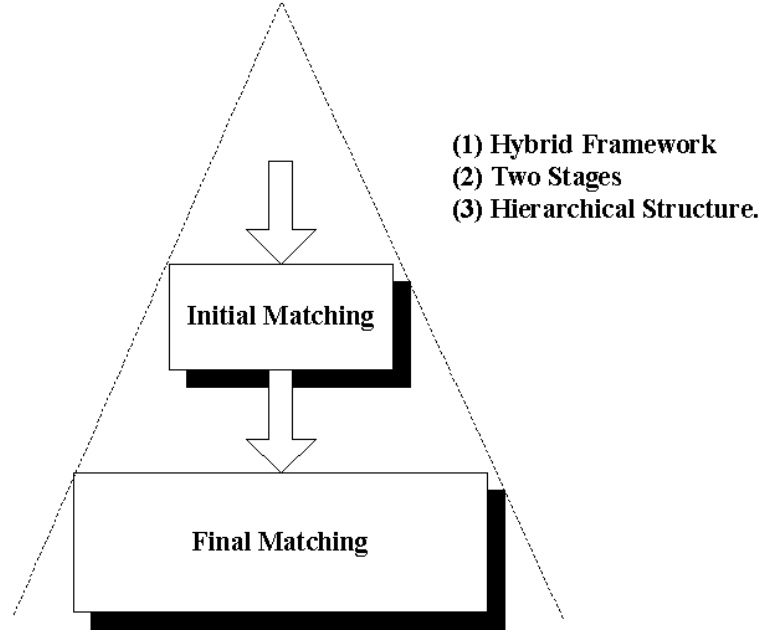
For example, Li *et al.* [14] proposed a contour-based approach to register image from multiple sensors. The success of their method depends on the assumption that the common structures of images must be preserved well. Therefore, their method is efficient but works well only on cases where the contour information is well preserved. In [15], Zheng and Chellappa proposed a novel method for determining the rotation parameter. They used a *Lambertian* model to model an image. Under the assumption that the illumination source is stationary, they use a shape-from-shading technique to estimate the illuminant directions of images. By taking the difference between the illuminant directions, the rotation angle between images is obtained. After obtaining the rotation angle, one of the two images is then rotated such that the orientation difference between the two images becomes very small. By adopting the method proposed by Manjunath et al. [16], a number of feature points are matched by using an area-based method in a hierarchical image structure. In Zheng and Chellappa's approach, the technique for estimating the rotation angle will fail due to the fact that the illumination conditions in one image are different to those in the other. Further, their approach requires a Gabor function decomposition in the feature extraction process. This decomposition is computationally intensive.

Methods for estimating motion that are based on the optical flow equation (OFE) [17-20] assume that the illumination of the scene is uniform. Recently, some researchers

tried to relax this brightness constancy assumption and developed algorithms to estimate the optical flow in the presence of illumination variations [21]. Hager and Belhumeur [22] proposed an efficient region matching and tracking algorithm based on robust estimation framework. They modeled the illumination changes into the sum of squared differences (SSD) formulation by using a low-dimensionality linear subspace determined from several images of the same object under different illumination conditions. The main disadvantage of this algorithm is the need of several images of the same object under different illumination conditions to compute the linear subspace before the tracking process. Lai [23] explicitly modeled spatial illumination variations by low-order polynomial functions in an energy minimization framework. Altunbasak et al. [24] proposed a similar model for time-varying illumination and imperfect optics, where the resulting optimization framework estimates the motion, illumination and camera parameters simultaneously. Haussecker and Fleet [25] used several physical models that describe brightness variations to compute the optical flow.

The BVM-based approach [23,24] basically accounts for smoothly varying illumination changes. However, the situations resulting in the brightness variation between the reference and the sensed images are very complex, and the effectiveness of the BVM-based approach for image registration is sometimes limited. This prompts the necessity to identify an appropriate image representation, on which the OFE-based estimation in a coarse-to-fine manner can be incorporated. The primary drawback of the OFE-based estimation is that it may fail unless the two images are misaligned by a moderate difference in scale, rotation, and translation. In order to overcome this problem,

an effective initial matching algorithm is required to bring the images into approximate alignment, even in the presence of large rotation angles and a wide range of scale changes.



**Figure 4.** The proposed image registration framework.

In this research, we focused on a hybrid and hierarchical image registration framework, which consists of two stages: *initial matching* and *final matching*. The purpose of the initial matching is to provide a good initial estimate to the second stage of final matching. The first-stage algorithm is applied to the coarsest level of both images. In the second stage, gradient-based algorithms are incorporated to precisely register images using a multi-resolution method. Fig. 4 describes the proposed block diagram of the hybrid approach.



### **Assumptions**

We are concerned with the problem of registering the sensed image from the video camera in nadir viewing on an aircraft with the reference imagery. The proposed algorithm is based on the following assumption. Since the distance between the camera and the target objects on the ground is very large, it is reasonable to assume that the scene under observation is relatively flat. Under weak perspective conditions, the perspective transformation can be well approximated by affine or projective transformations. Therefore, we limit our discussion here on affine or projective transformations. For the domain of the images under consideration, there are generally two types of distortions between the images to be registered. The first type is called *spatial* distortions that cause the images spatially misaligned in relation to each other. These distortions are typically geometric, and can be satisfactorily modeled by a practical transformation class. The second type of distortions can be attributed to illumination conditions, weather, seasonal variations, etc., which is called *non-spatial* distortions. These distortions are usually volumetric, and can't be corrected by the registration transformation. Actually, the non-spatial distortions make the registration more difficult.

## CHAPTER FOUR: FEATURE EXTRACTION

Edge contour extraction plays an important role in computer vision because edge contours are relatively invariant to the changes of illumination conditions, sensor characteristics, etc. In particular, edge contours can be used as matching primitives for correspondence determination, an important step in video geo-registration. In this chapter, we present a new approach for edge contour extraction based on a three-step procedure that using a RCBS-based scheme, inherently more accurate results can be produced, even though the edge model used for edges is relatively simple. We also present recursive filters that can efficiently smooth splines by approximating a signal with a complete set of coefficients subject to certain regularization constraints. We demonstrate our method on both synthetic and real images.

### **Introduction**

Edge detection is a fundamental operation in low-level computer vision with a plethora of techniques and several distinct paradigms that have been proposed [26]. Despite these efforts, the solutions to the edge contour extraction problem are still unsatisfactory for some application, where the subsequent processing stages depend primarily on the performance of the edge detector. In an image, an edge refers to an

occurrence of rapid image intensity transition [27]. An edge detector is to locate these transitions, and its resulting edge map identifies the location of them, together with some gradient and direction information. In order to use such an edge map in higher-level processes such as stereo and motion analysis, the next step is to identify those edge points that should be grouped together into edge segments [28]. However, real image data are often very diverse, and edges occur over a wide range of scales. As a result, the edge maps are cluttered with discontinued edge segments with degraded accuracy of location and isolated edge points. This type of edge map carries unexpected errors into the later stages of image processing tasks.

Edge maps provide a concise and accurate representation of the boundaries of objects and regions in an image. The geometric information contained in these stable edge contours is often sufficient to spatially match two images so that corresponding pixels in the two images correspond to the same physical region of the scene being imaged. Actually, the goal of image matching is to establish the correspondence between two images and determine the geometric transformation that aligns one image with the other. The major advantage of contour-based approach over the correlation-based is its insensitivity to scaling, rotation, and intensity changes as well as its low computational low cost. Video geo-registration, a process of associating 3D world coordinates with videos, is different from image registration, since most approaches of the latter employ only simple image-to-image mappings, which can't correctly model the projections between the 3D world and 2D frames [29]. However, after projecting the reference and video imagery to a common coordinate frame, the subsequent matching between them

can be largely 2D in nature, and a reliable and robust correspondence determination is essential to successfully register video images to a reference image [29,30]. At present, major video geo-registration schemes still choose normalized cross-correlation as a matching measure for correspondence determination, which is a computationally intensive process.

Torre and Poggio [31] have suggested that edge detection is a problem in numerical differentiation and showed that numerical differentiation of images is an ill-posed problem in the sense of Hadamard. Differentiation amplifies high frequency components. In practical situations, image data are contaminated by noise and differentiation will enhance high frequencies of the noise. However, differentiation is a mildly ill-posed problem [31], which can be transformed into a well-posed problem by several methods. Marr and Hildreth [32] proposed convolving the signal with a rotationally symmetric Laplacian of Gaussian (LoG) mask and locate zero-crossings of the resulting output, where the amount of smoothing is controlled by the variance of the Gaussian. There is a good theoretical foundation for using such a method because the Gaussian is an optimal filter for edge detection due to its localization properties in both the spatial and frequency domains. Additionally, Gaussian smoothing is also computationally inexpensive. However, the fundamental conflict encountered in edge detection is to eliminate noise without distorting its localization, namely noise immunity and accurate localization. The Gaussian filters, while smoothing the noise, also remove genuine high-frequency edge features to be sought, degrade edge localization and the detection of low-contrast edges. Edges can be weak but well localized. Such edges, also

representing a major fraction of information content in an image like the strong edges, maybe severely blurred by effective noise-reduction with circularly symmetric operators.

The LoG operator has been applied to extract edge contours [14]. As explained above, the main objection to the use of this operator has been that it has poor localization properties. The proposed scheme also chooses good control points on the non-closed edge contours, where uncertain errors will be introduced. To overcome these problems, many authors have proposed multi-scale techniques in order to reach the conflicting goals of detecting the existence of intensity discontinuities and of determining their exact location. The whole edge detection scheme requires that irrelevant details and noise being suppressed. Multiple scale algorithms can improve detection of weak edges and those that are close together. But, there are two major problems with multi-scale techniques, namely how to select the step between operator's scales for 2D images and effectively combine the edge contours from different scales. Williams and Shah [28] present a theoretical analysis of the movement of idealized edges, two adjacent step edges with the same or opposite polarity. However, it is very difficult to find a robust scale space combining approach in 2D case since edges obtained from the 2D filtered signals behave in a much more complex way in scale space than those from 1D filtered signals. In addition, combination of edge contours itself is an image matching problem, and computationally expensive. Although Canny edge detector has better SNR and edge location accuracy than that of LoG-based approach, the local extrema of its output may have unconstrained behaviors in the scale space [33,34]. Moreover, the 2D Canny detector, where is obtained by simply extending the 1D result based on the constant

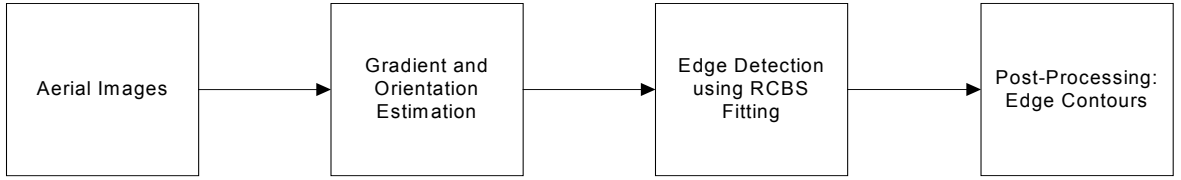
cross-section 2D edge model, is not optimal in 2D cases.

### **RCBS-based Edge Contour Extraction Algorithm**

The surface fitting method is an effective method used to detect edges based on the assumption that a 2D image is a discrete array of intensity values obtained from sampling a real-valued function  $f$  defined on the domain of the image [35,36]. To this end, some parametric form of the underlying function  $f$  is assumed and the discrete sampled intensity values in the finite size of neighborhood are used to estimate those parameters. Edge decisions are made based on the estimated underlying function  $f$ . Because such an edge detection scheme involves a process of fitting the basis functions to the sampled image data, the chosen basis function set must be complete in the sense of approximating all the edge features being sought in a scene. Otherwise, some edges, not in the feature space spanned by the selected basis function set, can't be detected.

Our proposed method applies regularized cubic B-spline fitting [37] (RCBS) for edge detection based on a one-dimension surface model from Nalwa-Binford [38], where 2D image data can be reduced into 1D by projecting the data in the direction of least local intensity changes. It is effective because the original image data can be smoothed to reduce noise while preserving discontinuities. The primary advantage of this one-dimension surface model is that it explicitly exploits the directional characteristics of edges, which differs from Haralick's 2D surface fitting approach [36]. Fig. 5 shows the block diagram of the basic structure of the proposed edge contour extraction scheme.

First, gradient and orientation at each pixel are estimated. Subsequently, RCBS fitting is applied for edge detection. In the end, edge contours of aerial images are generated after post-processing of edge maps from edge detection. Our proposed scheme generates better results than LoG-based methods, and the detected contours, open or closed, can directly be used for the matching stage.



**Figure 5.** Block Diagram of Edge Contour Extraction Algorithm

### **Gradient and Orientation Estimation**

Directionality in edge detection is not a new concept. First, a number of templates that could match an ideal edge model at various orientations are applied, and then edges and their directions are detected from results of the largest search and the thresholding of the outputs. Some simple and popular examples are Roberts' operator, Prewitt operator, and Sobel operator. On the use of these operators, however, there seems to have been considerable confusion between gradients and edges [39], where the magnitude of the gradient is used to locate the positions of edges. Actually, there is no such a direct correspondence between them. If we take noise and edge imperfection into consideration, weak but well localized edges are difficult to be detected using such simple thresholding

schemes, since edges are not at the locations of high gradient, but at locations of spatial gradient maxima. In our proposed scheme, a pixel is marked as an edge pixel, if there is a zero crossing of second derivative of the underlying function  $f$  taken in the direction of the estimated gradient.

Although there is a correspondence between the continuous and the discrete image, this is not the case between the continuous gradient and the discrete gradient due to inherent errors involved in gradient operators. Shigeru [39] presents optimal gradient operators using a newly derived consistency criterion, which is based on an orthogonal decomposition of the difference between a continuous gradient and discrete gradients into the intrinsic smoothing effect and the self-inconsistency involved in the operator. To obtain accurate gradient information, the author suggests that reduction of the self-inconsistency is of the primary importance, where the exact shape of the smoothing filter can be determined secondarily. The power spectrum of the consistent gradient image is expressed as:

$$|G_x(u, v)|^2 + |G_y(u, v)|^2 = \frac{1}{u^2 + v^2} |uH_i(u, v) + vH_j(u, v)|^2 |F(u, v)|^2 \quad (4.1)$$

The power spectrum of the inconsistent gradient image is express as:

$$|\tilde{G}_x(u, v)|^2 + |\tilde{G}_y(u, v)|^2 = \frac{1}{u^2 + v^2} |uH_j(u, v) - vH_i(u, v)|^2 |F(u, v)|^2 \quad (4.2)$$

Therefore, optimal gradient operators can be obtained by minimizing the inconsistent power



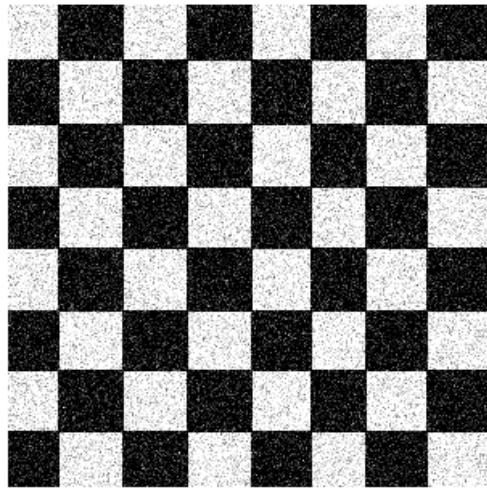
$$J \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \left| \tilde{G}_x(u, v) \right|^2 + \left| \tilde{G}_y(u, v) \right|^2 \right) dudv = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \Psi(u, v) P(u, v) dudv \rightarrow \mathbf{min} \quad (4.3)$$

where

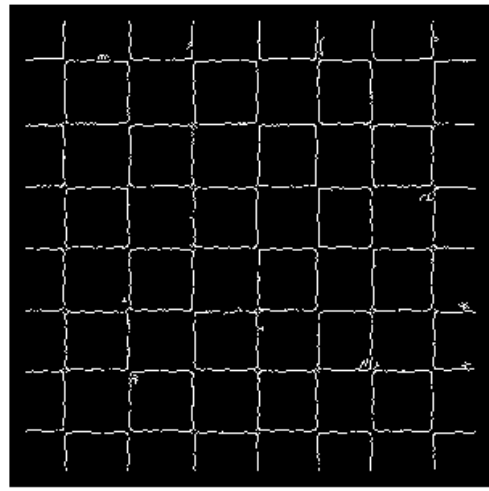
$$\Psi(u, v) = \frac{1}{u^2 + v^2} \left| uH_j(u, v) - vH_i(u, v) \right|^2 \quad (4.4)$$

$$P(u, v) = E[|F(u, v)|^2] \quad (4.5)$$

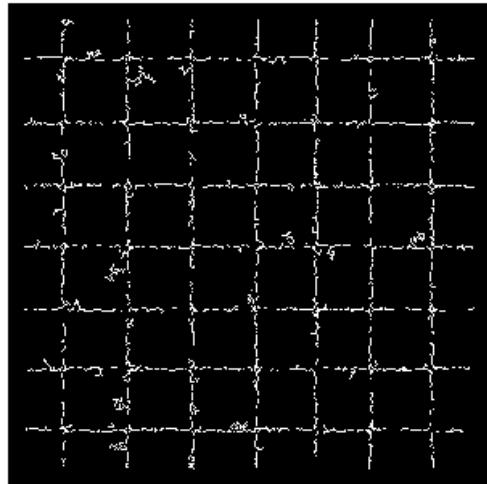
To obtain the best result, we choose Shigeru's optimum 5x5 gradient operators, and compare them with Prewitt operators with a size of 3x3 and Zuniga-Haralick's integrated directional derivative gradient operator (IDDG) with a size of 5x5. Fig. 6 presents the results of the above three operators on a synthetic checkerboard image with added Gaussian noise. Fig. 6(a) shows the synthetic checkerboard image with noise having  $\sigma=50$ . Fig. 6(b) presents edges produced by RCBS based on Shigeru's 5x5 gradient operators, Fig. 6(c) presents edges produced by RCBS based on IDDG operators, and Fig. 6(d) is the result of RCBS based on Prewitt operators.



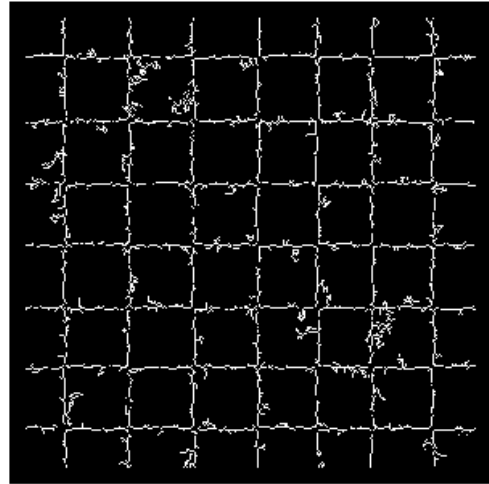
(a)



(b)



(c)



(d)

**Figure 6.** Comparison on synthetic images using different gradient operators.

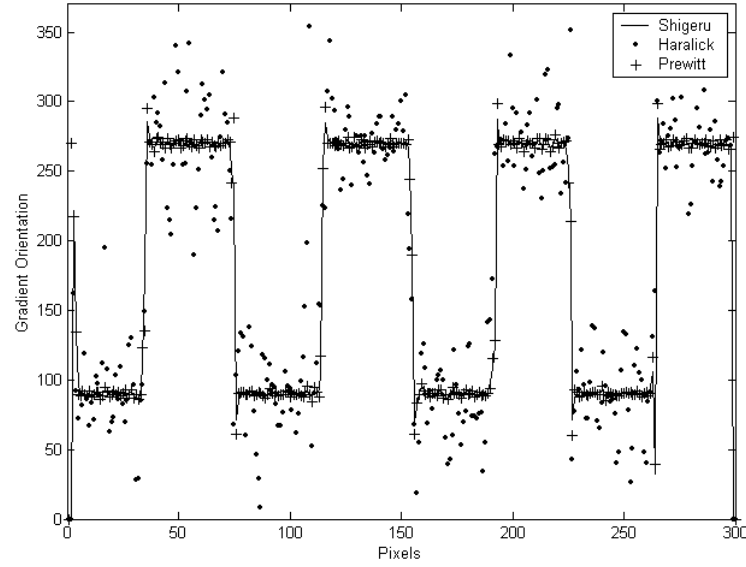
It can be seen that, when the noise level is high, the Prewitt operator produces the noisiest results because of errors in estimating gradient orientation. This is expected since operators with small support are always more sensitive to noise. Fig. 6(c) displays the results from IDDG operators, which are less noisy than that of the Prewitt operators, however, more true edges are missed, which could be from bigger errors in estimating the edge orientation. To this end, we took a slice near the position of a potential edge in the original image and tested with two levels of Gaussian noise ( $\sigma=20, 50$ ), and plotted the estimated gradient orientations from these three operators along the slice in Figures 7 and 8 respectively. We notice that orientation errors from IDDG operators are at some points worse than those of Prewitt operators, and these errors generate more missed detections at true edge pixels. Therefore, using a bivariate cubic polynomial to model the image would also be noise sensitive when the noise level is high. As is evident in these figures, RCBS based on Shigeru's gradient operators gives the best result (Fig. 6(a)) regarding to both noise immunity and the accuracy of localization.

### **Edge Detection by Regularized Cubic B-Spline Fitting**

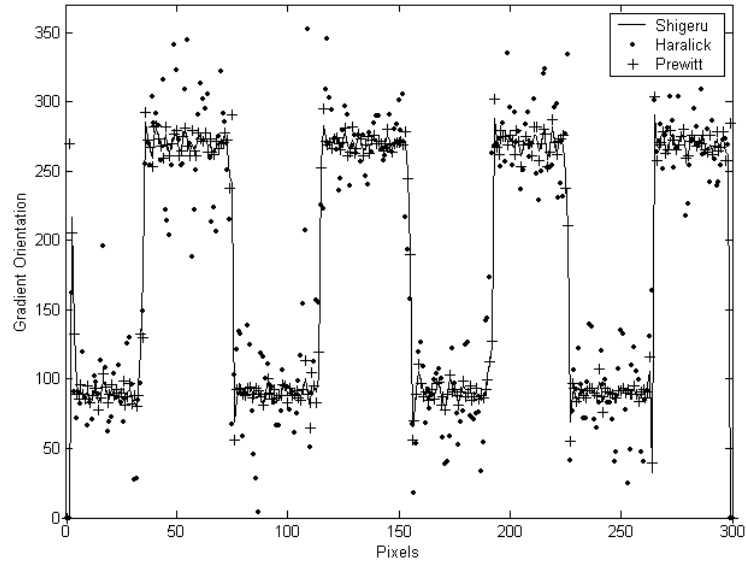
Splines are piecewise polynomials with pieces that are smoothly connected together. B-splines are splines that have smallest possible support, in other words, they are zero on a large set. The essential property of B-splines of order  $n$  is to provide a basis of the subspace of all continuous piecewise polynomial functions of degree  $n$  with derivatives up to  $n-1$  that are continuous everywhere on the real line [40]. In the case of equally spaced integral knot points, any uncton  $\phi^n(x)$  of this space can be represented as:

$$\phi^n(x) = \sum_{i=-\infty}^{+\infty} c(i) \beta^n(x-i) \quad (4.6)$$

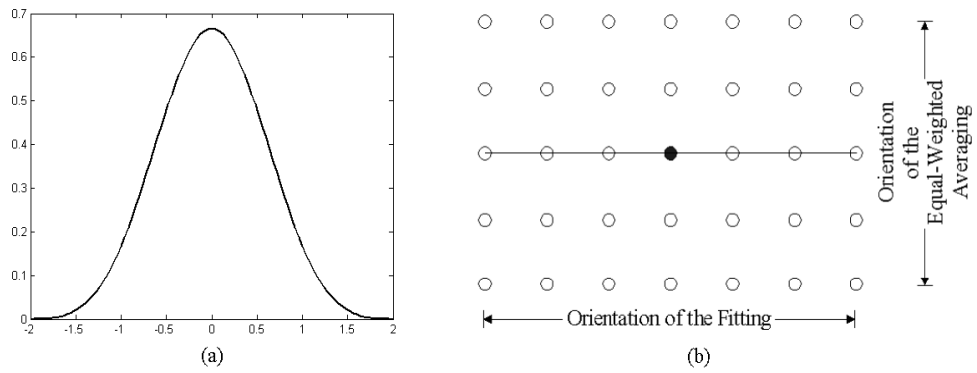
where  $\beta^n(x)$  denotes the normalized B-spline of order  $n$ . Because the function  $\phi^n(x)$  is uniquely determined by its B-spline coefficients  $\{c(i)\}$ , the crucial step in B-spline interpolation is to determine the coefficients of this expansion such that  $\phi^n(x)$  matches the values of some discrete sequence  $\{g(k)\}$  at the knot points:  $\phi^n(k) = g(k)$  for  $\{k = -\infty, \dots, +\infty\}$ .



**Figure 7.** Comparison between three gradient operators ( $\sigma = 20$ ).



**Figure 8.** Comparison between three gradient operators ( $\sigma = 50$ ).



**Figure 9.** (a) Cubic B-spline; (b) RCBS mask.

In Fig. 9(a) we show the following closed form representation of the cubic B-spline:

$$\beta^3(x) = \begin{cases} 2/3 - |x|^2 + |x|^3/2, & 0 \leq |x| < 1 \\ (2 - |x|)^3/6, & 1 \leq |x| < 2 \\ 0, & 2 \leq |x| \end{cases} \quad (4.7)$$

Chen and Yang [37] proposed a RCBS-based edge detection scheme, which uses a set of cubic B-splines to approximate the underlying 3D intensity surface along the gradient direction in Fig 9(b). Because real image data are corrupted by noise, a regularization term is introduced to suppress its effect. Reinsh [41] and Schoenberg [42] have proposed the use of smoothing spline. Given a set of discrete signal values  $\{g(k)\}$ , the smoothing spline  $\hat{g}(x)$  of order 3 is defined as the function that minimizes:

$$\varepsilon_s^2 = \sum_k [g(x_k) - \hat{g}(x_k)]^2 + \lambda \int \left[ \frac{d^2 \hat{g}(x)}{dx^2} \right]^2 dx = \varepsilon_A^2 + \lambda \varepsilon_r^2 \quad (4.8)$$

$$\hat{g}(x) = \sum_{i=0}^{M+1} c(i) \beta^3(x - i) \quad (4.9)$$

where  $\lambda$  is a given positive parameter. The choice of  $\lambda$  depends on which of these two conflicting goals is accorded the greater importance. A set of cubic B-splines basis function  $\beta^3(x)$  is used for the fitting between the interval  $[1, M]$ , where  $M$  is the number of the grid pixels to be fitted. The minimization of  $\varepsilon_s^2$  can be achieved by setting the

partial derivatives of  $\varepsilon_s^2$  with respect to  $\{c(i)\}, k = 0, \dots, M + 1$  to zero. This leads to the solution of a system of  $M + 1$  linear equations, and can be solved by matrix operations.

In the following, we show how to efficiently determine the coefficients of the smoothing spline by digital filtering instead of conventional matrix operations. Applying the properties of the B-splines, a simpler form of  $\varepsilon_r^2$  is therefore:

$$\varepsilon_r^2 = \sum_{k \in Z} (d^{(2)} * c)(k) (d^{(2)} * b_1^3 * c)(k) \quad (4.10)$$

Then,  $\varepsilon_s^2$  can be expressed in terms of discrete convolutions

$$\varepsilon_s^2 = \sum_{k \in Z} (g(k) - (b_1^3 * c(k)))^2 + \lambda \sum_{k \in Z} (d^{(2)} * c)(k) (d^{(2)} * b_1^3 * c)(k) \quad (4.11)$$

which, using the inner product notation, is also equivalent to:

$$\varepsilon_s^2 = \langle g, g \rangle - 2 \langle g, b_1^3 * c \rangle + \langle b_1^3 * c, b_1^3 * c \rangle + \lambda \langle d^{(2)} * c, d^{(2)} * c * b_1^3 \rangle \quad (4.12)$$

The smoothing spline coefficients are found by setting to zero the derivative of this expression with respect to  $c(k)$ . By using the properties of the inner product calculus, we find that

$$(b_1^3)' * g(k) = (b_1^3)' * b_1^3 * c(k) + \lambda (d^{(2)'} * d^{(2)} * (b_1^3)' * c(k)) \quad (4.13)$$

Applying z transform on both sides, we have

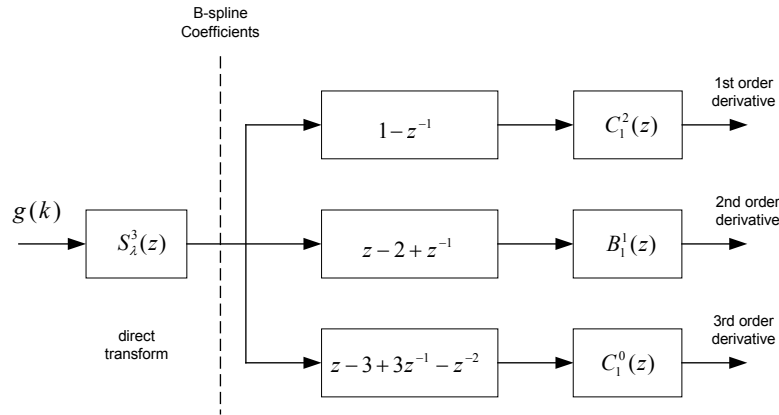
$$B_1^3(z^{-1})G(z) = B_1^3(z^{-1})B_1^3(z)C(z) + \lambda D^{(2)}(z)D^{(2)}(z^{-1})B_1^3(z^{-1})Y(z) \quad (4.14)$$

By solving for  $C(z)$ , we have

$$C(z) = \frac{1}{B_1^3(z) + \lambda(z - 2 + z^{-1})^2} G(z) = S_\lambda^3 G(z) \quad (4.15)$$

$$B_1^3(z) = \frac{1}{6}(z + 4 + z^{-1}) \quad (4.16)$$

This above expression clearly shows that the coefficients of the smoothing spline can be determined by digital filtering [43], as illustrated in Fig. 10. The transfer function of the smoothing spline filter  $S_\lambda^3(z)$  corresponds to a IIR filter, which can be most efficiently implemented recursively. After we determine the cubic B-spline coefficients, we can directly obtain a general cubic B-spline differentiator (first-, second-, and third-order derivatives) shown in Fig. 10.



**Figure 10.** Block Diagram of a B-spline differentiator using the smoothing B-spline



kernels.

We mark the center pixel of the operator support as an edge pixel if for some  $r$ ,  $|r| < r_0$ , where  $r_0$  is smaller than half the length of the side of a pixel,  $\hat{g}_\alpha''(r) = 0$ ,  $\hat{g}_\alpha'''(r) \neq 0$ , and  $|\hat{g}_\alpha'(r)|$  is greater than its neighbors', namely a non-maximum suppression step. For every marked edge point, the edge strength is defined as the slope at each zero-crossing of  $\hat{g}_\alpha''(r)$  in the estimated gradient direction.

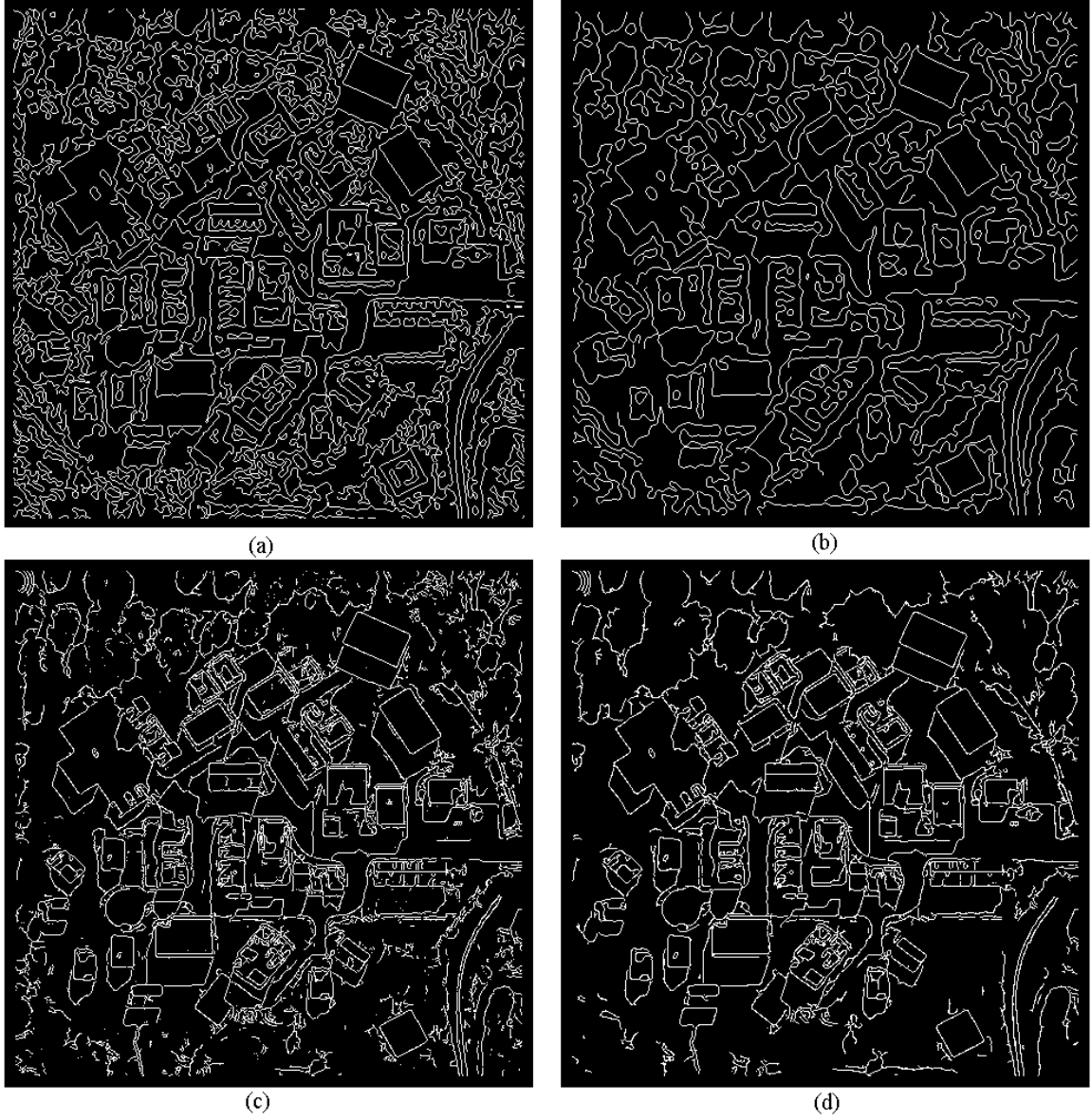
### **Edge Contours: Post-processing of RCBS Edge Maps**

The results from RCBS edge maps maybe “noisy”, but the edges are continuous and thin. The noisy edge points can be discarded easily in the post-processing stage using the edge strength defined above. First, a hysteresis thresholding [45] is applied to the edge map. This algorithm is basically the same as the one used in the Canny algorithm. The low threshold is set to preserve the whole contour around the region boundary without incurring discontinuities at weak edge points. The high threshold is chosen large enough to avoid spurious edges. This two-threshold scheme is implemented by scanning the 2D edge strength array. Contour search is initiated wherever one point with a value greater than the high threshold is scanned. The same search operation continues until the whole edge strength array has been scanned. The contours are then divided into two categories, closed contours and open contours. Second, different thinning rules are applied. For example, if the point is adjoining a diagonal edge, then remove it. These rules are applied to each identified edge pixel in the edge map sequentially left to right

and top to bottom, which can be achieved using only one pass of the algorithm.

### **Experiment Results**

Because of space limitations, only results from one aerial image are presented. Fig. 11 shows the results of the LoG-based and RCBS-based schemes with different scales. Fig. 11(a) is the output of the LoG-based scheme ( $\sigma=1$ ). Fig. 11(b) is the output of the LoG-based scheme ( $\sigma=3$ ), Fig. 11(c) the output of RCBS-based scheme ( $\lambda=0.0000001$ ) at a fine scale, and Fig. 11(d) the output of RCBS-based scheme ( $\lambda=0.01$ ) at a medium scale.



**Figure 11.** Examples of edge contour extraction by (a) LoG ( $\sigma=1$ ), (b) LoG ( $\sigma=3$ ), (c) RCBS ( $\lambda=1.0E-7$ ), (d) RCBS ( $\lambda=1.0E-1$ ).

Note that the size of LoG operators is different at different scale and the size is bigger at coarse scales. On the other hand, in the RCBS-based scheme the operator's support is fixed. Our experiments use a rectangular window of size  $5 \times 7$ . For LoG operators, the detection of edges with accurate position depends on the scales we choose [46]. When we choose a smaller scale ( $\sigma=1$ ), the resulted edge contours by LoG operators are shown in Fig 11(a). Obviously, the edges are very noisy and include a lot of false edges. In order to suppress the effect of noise, we try a bigger scale having  $\sigma=3$  and the result is shown in Fig. 11(b). As it is expected, the noise is greatly suppressed, but we have further degraded the localization of edge contours. In addition, note also the large influence of each edge on one another at a coarse scale for the LoG operator because of its corresponding bigger operator size [46]. Figures 11(c) and 11(d) show the results from our proposed RCBS-based scheme. Obviously, it produces much better results compared with those of the LoG-based scheme. We test it using two different scales ( $\lambda=1.0E-7$  and  $1.0E-1$ ). From the figures we can see that it can effectively suppress the noise, even at the fine scale ( $\lambda=1.0E-7$ ). At medium scale ( $\lambda=1.0E-1$ ), the localization of edge contours only shift very slightly. In conclusion, our proposed RCBS-based scheme is an effective way to control the balance between the two conflicting performance requirements for real images, namely noise immunity and accurate localization.

We have presented a new approach for edge contour extraction based on a three-step procedure. In the first step we obtain greatly improved estimated gradient information from Shigeru's operators. The second step applies RCBS edge detection,

which is an effective way to control the balance between the two conflicting performance requirements, namely noise immunity and accurate localization. The third step post-processes the resulted edge map using some strategies, which generate qualified edge contours for higher visual processing tasks. The experiment results indicate that our proposed scheme has better performance in both noise immunity and localization than that of LoG-based schemes. For some applications with time constraints, digital filtering techniques have been applied for solving the problem of regularized cubic B-spline fitting instead of the matrix approaches [47].

The balance of the operator's support size and the regularization parameter with the noise immunity and the localization requirements should be further researched. Most existing algorithms in the literature treat edge detection a purely local process, which can't guarantee the connections between the detected edge points. As a result, the output edge map often contains many broken segments. Accordingly, research work should focus on taking a global view towards edge detection using its related properties, such as position, curvature, orientation, and contrast, in order to suppress the effect of noise and generate more manageable 2D edge contours for the stage of matching process.

## CHAPTER FIVE: INITIAL MATCHING

In this chapter, we present several effective algorithms for initial matching, which aims to handle a wide range of motion between images.

### **Phase Correlation: A FFT-based Image Registration**

The hierarchical registration method consists of two stages. The first stage is a coarse phase correlation registration [48,49] on the coarsest level of the pyramids, where the translation, scale, and rotation parameters of the similarity transformation are determined. The idea behind phase correlation method is simple and is based on the Fourier shift property that states that a shift in the coordinate frames of the function becomes a linear phase difference in the Fourier domain. This can be described as follows:

Let  $f_1(x,y)$  and  $f_2(x,y)$  are the images that differ only by a displacement  $(\Delta x, \Delta y)$  i.e.,

$$f_2(x,y) = f_1(x - \Delta x, y - \Delta y) \quad (5.1)$$

According to the Fourier shift property, their corresponding Fourier transforms  $F_1$  and  $F_2$  will be related by:

$$F_2(\xi, \eta) = F_1(\xi, \eta) e^{-j2\pi(\xi\Delta x + \eta\Delta y)} \quad (5.2)$$

Hence, the normalized cross power spectrum (NCPS) of two images with their Fourier transform is given by:

$$\frac{F_1(\xi, \eta) F_2(\xi, \eta)^*}{|F_1(\xi, \eta) F_2(\xi, \eta)^*|} = e^{j2\pi(\xi\Delta x + \eta\Delta y)} \quad (5.3)$$

where \* indicates the complex conjugate.

The practical approach to solve Eq. (5.3) for  $(\Delta x, \Delta y)$  is to first take inverse Fourier transform of NCPS.

$$NCPS(x, y) \equiv \mathfrak{F}^{-1} \{ e^{j2\pi(\xi\Delta x + \eta\Delta y)} \} = \delta(x - \Delta x, y - \Delta y) \quad (5.4)$$

Eq. (5.4) is a Dirac delta function centered at  $(\Delta x, \Delta y)$  and it is a simple matter to determine  $\Delta x, \Delta y$  from Eq. (5.4).

If the image  $f_2(x, y)$  is a rotated, scaled, and translated version of the image  $f_1(x, y)$ , we can write the relationship between them as follows:

$$f_2(x, y) = f_1(s(x \cos \theta_0 + y \sin \theta_0) - \Delta x, s(-x \sin \theta_0 + y \cos \theta_0) - \Delta y) \quad (5.5)$$

The Fourier transform of  $f_1(x, y)$  and  $f_2(x, y)$  are respectively  $F_1(\xi, \eta)$  and  $F_2(\xi, \eta)$ , and their magnitudes are related by:

$$M_2(\xi, \eta) = s^{-2} M_1(s^{-1}(\xi \cos \theta_0 + \eta \sin \theta_0), s^{-1}(-\xi \sin \theta_0 + \eta \cos \theta_0)) \quad (5.6)$$

Since Eq. (5.6) is a Fourier transform magnitude, it is independent of the translation parameters  $(\Delta x, \Delta y)$ , which affect only the phase component according to Eq. (5.3).

Then, we can rewrite Eq. (5.6) using log-polar coordinates:

$$M_2(\rho, \theta) = s^{-2} M_1(\rho - \log s, \theta - \theta_0) \quad (5.7)$$

Eq. (5.7) indicates that the amplitude of the log-polar spectrum is scaled by  $s^{-2}$ , that image scaling results in a shift of  $\log s$  along the log-radius  $\rho$  axis, that image rotation results in a cyclical shift of  $\theta_0$  along the angle  $\theta$  axis, and that image translation has no effects in the log-polar domain.

According to the shift property of the Fourier transform, the Fourier transforms of  $M_1$  and  $M_2$  are related by:

$$\Phi_2(\xi_\rho, \eta_\theta) = s^{-2} e^{-j2\pi(\xi_\rho \cdot \log s + \eta_\theta \cdot \theta_0)} \Phi_1(\xi_\rho, \eta_\theta) \quad (5.8)$$

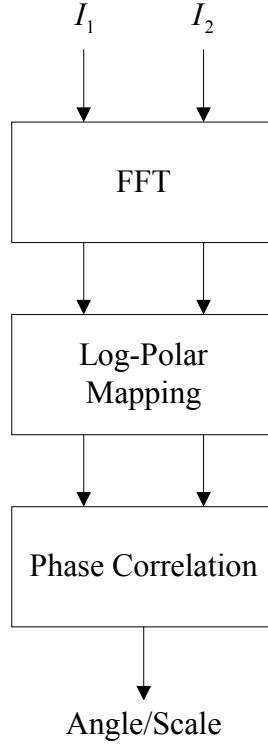
$$\Phi_1(\xi_\rho, \eta_\theta) \equiv \Im\{M_1(\rho, \theta)\} \quad (5.9)$$

$$\Phi_2(\xi_\rho, \eta_\theta) \equiv \Im\{M_2(\rho, \theta)\} \quad (5.10)$$

Thus, the NCPS of two magnitude images with their Fourier transform is given by

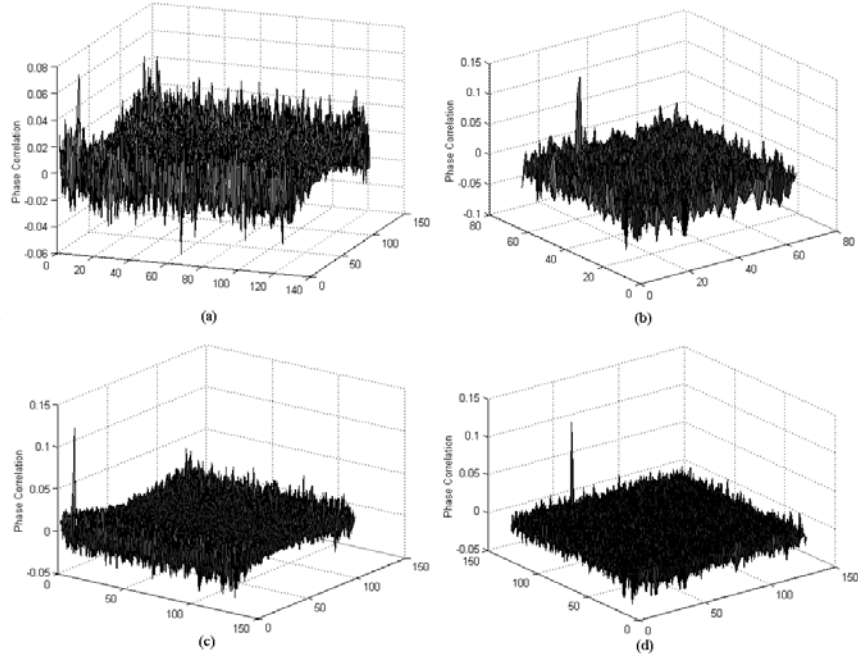
$$\frac{\Phi_1(\xi_\rho, \eta_\theta) \Phi_2^*(\xi_\rho, \eta_\theta)}{|\Phi_1(\xi_\rho, \eta_\theta) \Phi_2^*(\xi_\rho, \eta_\theta)|} = e^{j2\pi(\xi_\rho \log s + \eta_\theta \theta_0)} \quad (5.11)$$





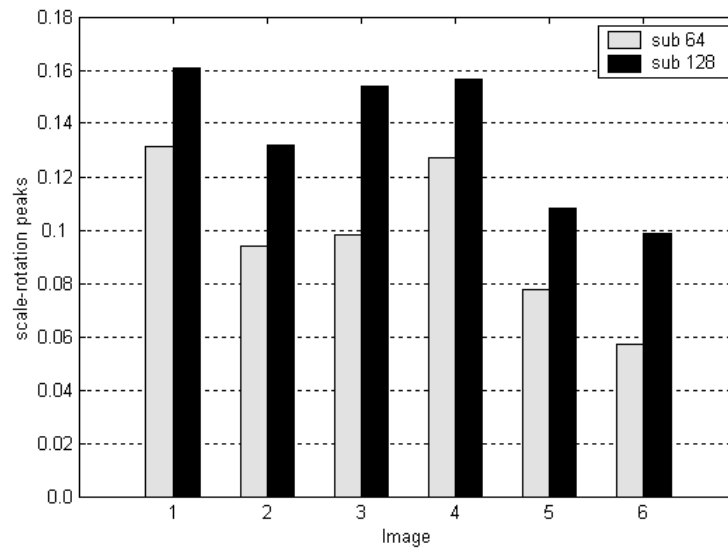
**Figure 12.** Block diagram of phase correlation for angle and scale estimation.

According to Eq. (5.11), the scale  $s$  and rotation  $\theta_0$  could be obtained by taking inverse Fourier transform of the NCPS, which is a Dirac delta function centered at  $(\log s, \theta_0)$ . Once the scale and rotation angle are determined, the image is scaled and rotated by amounts  $s$  and  $\theta_0$ , respectively, and the amount of translation can be found out using the phase correlation method. Based on the above description, Fig. 12 shows the block diagram of the procedures to determine the scale  $s$  and rotation  $\theta_0$  using the phase correlation method (PCME).

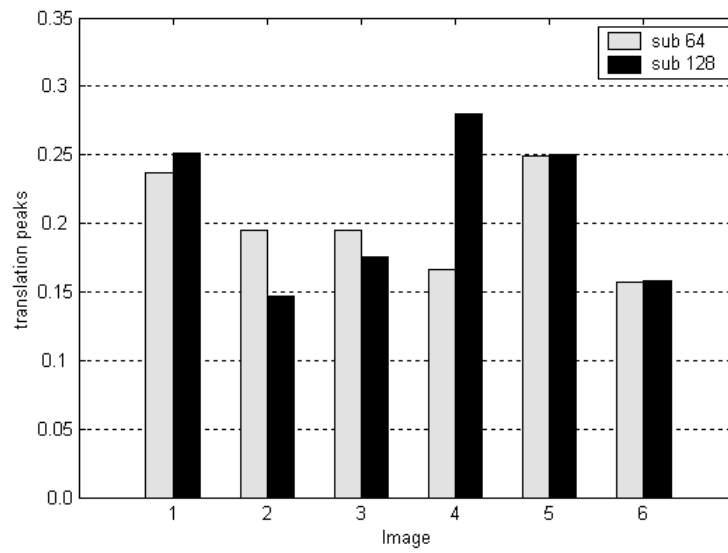


**Figure 13.** (a)(c): Phase correlation for rotation and scale between log-polar mapping at the levels of 64x64 and 128x128, respectively; (b)(d): Phase correlation for translation after compensation at the levels of 64x64 and 128x128, respectively.

In this section, we illustrate the procedures, and evaluate the performance of the proposed hierarchical approach using different sets of aerial images. Using PCME for similarity transformation, the estimated translations, rotations, and scales can be measured to a good accuracy.



**Figure 14** Comparison of rotation-scale peak between 64x64 (sub 64) and 128x128 (sub 128) levels, respectively.



**Figure 15.** Comparison of translation peak between the 64x64 (sub 64) and 128x128 (sub 128) levels, respectively.

For our proposed hierarchical registration algorithm, PCME is applied at the highest level of decomposition of the Gaussian pyramid. The decomposition level is chosen so that at the coarsest scale, the image still retains enough overlapping area in size. In the experiment, we limit our decomposition level to be 4, so that the coarsest image level uses  $64 \times 64$  images for the  $512 \times 512$  test images we consider.

**Table 1.** Registration results using PCME at the coarsest level of  $64 \times 64$ .

		1	2	3	4	5	6
Original	Tx	111	148	185	222	259	296
	Ty	-54	-64	-90	-108	-126	-144
	Rotation	$-5^0$	$-10^0$	$-15^0$	$-20^0$	$-25^0$	$-30^0$
	Scale	0.997	1.270	1.100	0.997	1.200	0.970
Estimated	Tx	110.2	142.4	179.7	221.2	263.2	300.7
	Ty	-51.3	-64.4	-98.6	-106.5	-136.9	-149.0
	Rotation	$-4.39^0$	$-8.82^0$	$-14.36^0$	$-19.06^0$	-25.09	-29.89
	Scale	0.988	1.289	1.077	0.974	1.193	0.958

We have experimentally found that a size of  $64 \times 64$  is appropriate for coarse matching using PCME. In the log-polar plane, the size of representation is  $128 \times 128$ . Therefore, logarithmic conversion of  $\rho$ -axis is done with based 1.033. This base value is chosen because  $\log_{1.033}^{64} = 128$ , i.e., 64 rows will be mapped to 128 rows in polar plane, where the based value is chosen based on the required level of accuracy. First, we compare the phase correlation results for two image levels in the Gaussian pyramid (Level 2:  $128 \times 128$ , and level 3:  $64 \times 64$ ). To estimate scales and rotations, Figures 13(a) and 13(c) in show their cross power spectrum in the spatial domain after applying FFT and log-polar mapping to both test image pairs, where the positions of peaks correspond

to the rotations and scales. Once the rotation and scale are known, the target images at two levels, sub 64 and sub 128, can be compensated, after which PCME can be applied again to directly estimate translation parameters. Figures 13(b) and 13(d) shows their cross power spectrum in the spatial domain, where the positions of peaks can be easily located to obtain the translations. For phase correlation, matches are considered valid only if the peak value of the cross power spectrum in the spatial domain is greater than 0.03. Theoretically, for exact matches this value should be equal to 1.0. However, the presence of dissimilar parts and the noise in image reduce the peak value. From Fig. 13, we find the cross power spectrum in the spatial domain is much noisy for the coarsest image level (sub 64) due to the small overlapping area.

Estimation accuracy was assessed when global motion was manually induced and hence known a priori. In this way, we can study the accuracy of the registration since we know the actual transformation. We use image “Desert” as the original test image. In this experiment, various translation, scale, and rotation values are used to generate the target images, where the geometric distortions are not so large as to be irrelevant for practical applications. Table 1 summarizes the results based on registration of the “Desert” image to the generated target images using PCME. Row “original” show the exact transformation parameter sets applied to the original test image. Row “estimated” shows the computed transformation parameters using PCME at the coarsest pyramid level. With the same procedures as above, we compare in Fig. 14 the phase correlation peaks for estimation of scales and rotations at two image levels in the Gaussian pyramid (Level 2:  $128 \times 128$ , and level 3:  $64 \times 64$ ). Fig. 15 shows the peak values of their cross power

spectrum in the spatial domain for translations after compensation.

### **A Point Pattern Matching Algorithm**

In this section, we have pursued a feature point based method using common points in both images. Several researchers [50-53] have addressed the problem of control-point matching. In general, there are two types of information that can be used in control-point matching; feature properties associated with each point, and relative distances between points. Ranade and Rosenfeld [50] proposed an iterative point-matching algorithm based on the relative distance information between points. However, their algorithm can only handle translation. Ton and Jain's algorithm [51] can handle both translation and rotation. Goshtasby and Stockman [52] used the convex hull property to choose subsets of points for matching. The convex hull property is not suitable to reduce the computational cost when there are many extra or missing point patterns. Wang and Chen [53] exploited the invariant relations between line segments of reference and target images respectively.

The problem considered here can be stated as follows: We are given two control point sets in the reference image  $Q_1 = \{a_{1j} = (x_{1j}, y_{1j}) | j = 1, \dots, m\}$  and in the target image  $Q_2 = \{b_{2k} = (x_{2k}, y_{2k}) | k = 1, \dots, n\}$ , respectively. The goal is to find the maximum matching pairs  $l$  between  $Q_1$  and  $Q_2$ , where  $l$  is unknown and  $l \leq \min(m, n)$ . Our basic assumption is that enough control points are available in both images. It is important to observe that since our matching algorithm is designed to deal with missing and spurious points, we

don't require foolproof performance of the control point operator.

A 2D affine transformation is a mapping  $b_{2k}^T = A \cdot a_{1j}^T + t$  from coordinate system  $a_{1j} = (x_{1j}, y_{1j})$  to coordinate system  $b_{2k} = (x_{2k}, y_{2k})$ , where  $A$  is a nonsingular matrix and  $t$  is a translation vector. An important observation is that the area  $S$  of any object is a relative affine invariant, i.e.  $S' = \det(A) \cdot S$ . For correct and efficient mapping it is necessary to identify the invariance properties of the transform in order to characterize the mapping between two sets of control points. Based on triangles in the two images, two invariance properties relate the area and perimeter ratios of any triangle pair. Let  $R_{jk}$  be the area ratio of a triangle  $\Delta_{1j}$  in the reference, and  $\Delta_{2k}$  in the target image respectively:

$$R_{jk} = \gamma \log(S(\Delta_{1j})/S(\Delta_{2k})), \quad j = 1, \dots, C_3^m, \quad k = 1, \dots, C_3^n \quad (5.12)$$

where  $S(\Delta)$  denotes the area of triangle  $\Delta$  and  $\gamma$  a given constant. Similarly let  $P_{jk}$  be the perimeter-length ratio:

$$P_{jk} = \beta \log(L(\Delta_{1j})/L(\Delta_{2k})), \quad j = 1, \dots, C_3^m, \quad k = 1, \dots, C_3^n \quad (5.13)$$

where  $L(\Delta)$  denotes the perimeter and  $\beta$  a given constant.

The proposed algorithm consists of three phases; (1) geometric invariance properties between randomly selected triangles in the two images are evaluated, (2) an accumulator is formed where votes on a particular match are tallied, and (3) a procedure

of scanning the accumulator to identify corresponding point pairs. In phase (1) we create a 2D Area-Perimeter (  $AP$  ) histogram based on the area and perimeter-length ratios. By using a 2D histogram, all possible matching pairs of triangles can be identified. The 2D histogram is formed by first computing all  $R_{jk}$  and  $P_{jk}$ ,  $j = 1, \dots, C_3^m$ ,  $k = 1, \dots, C_3^n$ . Then we construct a Table  $T_s(j, k)$ ;  $j = 1, \dots, C_3^m$ ,  $k = 1, \dots, C_3^n$ , where  $T_s(j, k) = (R_{jk}, P_{jk})$  is evaluated from triangle pair  $\Delta_{1j}$  and  $\Delta_{2k}$  from Eqs. (5.12) and (5.13). The  $AP$  histogram is created from  $T_s(j, k)$ .

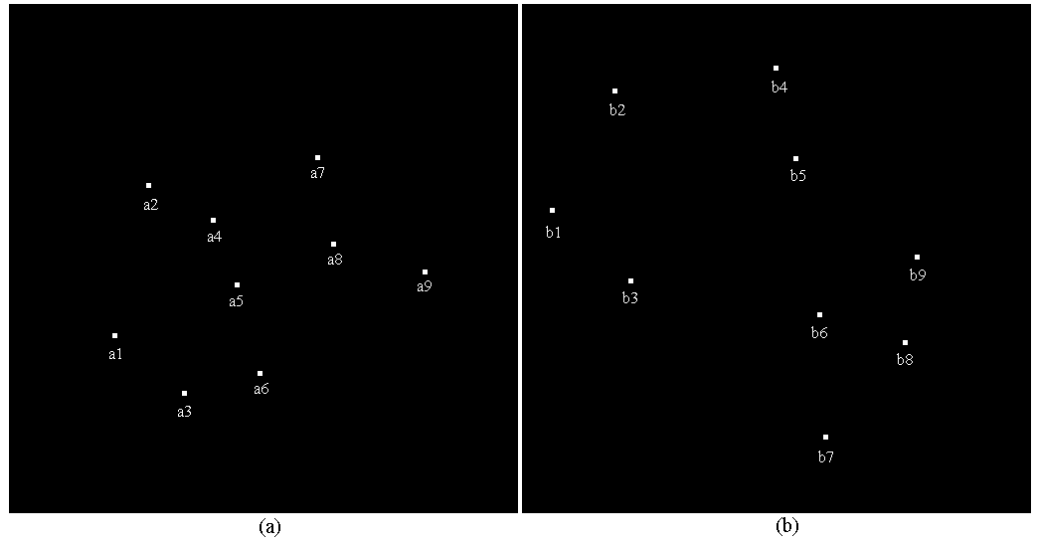
In phase (2) we form a matching Table  $\{T_m(r, c) | r = 1, \dots, m; c = 1, \dots, n\}$  of control points as follows: From the 2D histogram we find the maximum value and its corresponding  $R_{jk} = \hat{R}$  and  $P_{jk} = \hat{P}$ . Any triangle pair  $(\Delta_{1j}, \Delta_{2k})$  corresponding to  $T_s(j, k) = (\hat{R}, \hat{P})$  is selected as a candidate of a possible correct pair. Suppose  $\Delta_{1j} = (a_{1j_1}, a_{1j_2}, a_{1j_3})$ ,  $\Delta_{2k} = (b_{2k_1}, b_{2k_2}, b_{2k_3})$  is a candidate triangle pair. Let  $v_1 = \|\overrightarrow{a_{1j_1}a_{1j_2}}\| / \|\overrightarrow{b_{2k_1}b_{2k_2}}\|$ ,  $v_2 = \|\overrightarrow{a_{1j_2}a_{1j_3}}\| / \|\overrightarrow{b_{2k_2}b_{2k_3}}\|$ ,  $v_3 = \|\overrightarrow{a_{1j_1}a_{1j_3}}\| / \|\overrightarrow{b_{2k_1}b_{2k_3}}\|$ . If all three conditions

$$|v_1 - v_2| < \varepsilon, \quad |v_2 - v_3| < \varepsilon, \quad |v_3 - v_1| < \varepsilon \quad (5.14)$$

are satisfied ( $\varepsilon$  is a small threshold), the three cells of  $T_m$  accumulate one vote, i.e.,  $T_m(r, c) = T_m(r, c) + 1$ ,  $r = j_1, j_2, j_3$ ,  $c = k_1, k_2, k_3$ , respectively. The value in cell  $(r, c)$  denotes the possibility that the  $(r)$ th control point  $a_{1r}$  of reference image and the  $(c)$ th



control point  $b_{2c}$  of target image could be a matching pair. After Table  $T_m (m \times n)$  is formed, a “scanning algorithm” can determine reasonable pairs. The pairing procedure is described as follows: First find the maximum value in each row. If that value is also the maximum in the corresponding column then we keep it and set all other values in the same row and column to zero. If not, then we set the entire row and column to zero. After this scanning, non-zero remaining values exceeding another threshold are used to determine matching pairs.



**Figure 16.** Two point patterns.

**Table 2.** Success ratio of some experiments.

Point set 1 ( $m$ )	Point set 2 ( $n$ )	Matching pairs ( $l$ )	Displacement (pixel)	Success Ratio
35	30	25	1	100%
35	30	25	2	100%
35	30	25	3	100%
20	15	12	1	100%
20	15	12	2	100%
20	15	12	3	100%
15	12	6	0	23%
15	12	6	1	20%
15	12	6	2	20%
15	12	6	3	16%

A large number of control points will produce a lot of triangles, and increase the possibility of false pairing and the amount of computations. In our experiment,  $m$  and  $n$  range from 4 to 35. After corresponding pairs are found, affine transformation parameters can be easily estimated using the method proposed by Lamdan et al. [54]. Figure 16 shows two sets with 9 control points extracted from two aerial images using the Harris corner detector, where 6 in each set are correct pairs and the other 3 are randomly inserted points. For this example, the matching index  $MI$  defined in [51] as:

$$MI = \frac{(k-1)}{(m-1)} \cdot \frac{(k-1)}{(n-1)} \quad (5.15)$$

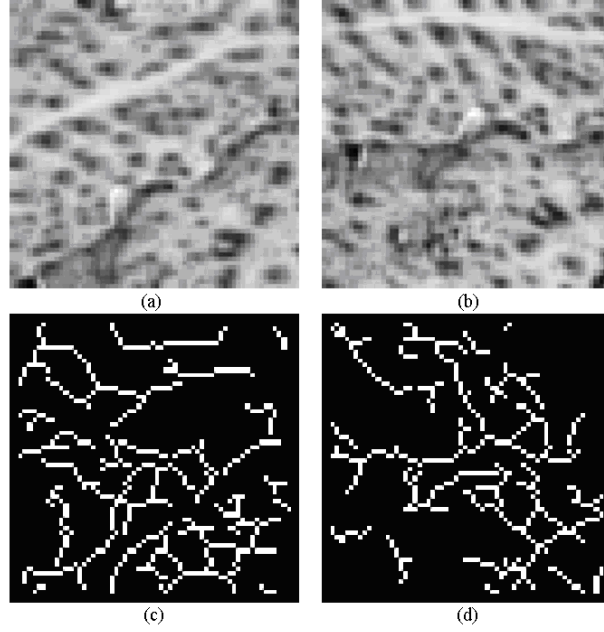
has a value of 0.39. We consider the matching to be successful if at least two-thirds of the  $l$  true control-point pairs are detected. The success ratio is the ratio of successful trials to the total number of trials. In this experiment, our algorithm correctly identified the 6 matching pairs. To test the robustness of the proposed algorithm, we performed a Monte Carlo simulation similar to [51], where for each combination of parameters, the previous experiment is repeated 40 times using two 512x512 images containing both matching and randomly inserted points. The positions of matching and incorrect points change each time. Furthermore, the matching points are slightly displaced to simulate extraction errors. The simulation results are shown in Table 2. It is noticed that slight point displacement has little effect on the success ratio, but the number of incorrect points does.

We have proposed a new algorithm for a more robust matching of point patterns

under a similarity transformation. The performance of our proposed algorithm is experimentally demonstrated using Monte Carlo simulations.

### **A Contour-based Algorithm**

Our goal is to define fast and accurate methods for automatic registration of aerial images. As mentioned above, feature matching could also be computationally expensive if there involves a large number of feature candidates from the feature extraction stage. To address that problem in the matching process, we propose to solve the registration problem in two stages: 1) the transformation space match methods with the Hausdorff distance measure [55], which aims to provide a good estimate of the mapping function, is applied to the coarsest scale of both edge images, 2) salient and well-distributed candidates of feature points are extracted throughout the original images at the fine scale, and the initial transformation is utilized to speedup the process of feature point matching.



**Figure 17.** Edge contours at the coarsest scale.

In the first stage, the edges images should be much less noisy at the highest level of decomposition, and often correspond to salient features in the original images. Fig 17 shows two aerial photos at the coarsest scale and the edge maps detected based on the proposed contour extraction model. Note that the images in Fig. 17 have been enlarged for the purpose of illustration. The computational cost at the first stage is much less because only salient edges are considered, and the translations between these two edge images are smaller. The search space is composed of 2D rotations, scales, and translations. The system looks for rotations with angle included in the interval  $[0, 90^\circ]$ . At the highest level of decomposition, when only looking for rotations, the search is exhaustive over the whole search space but with an accuracy equal to  $\Delta$ . The first approximation of the best rotation  $\theta_o$  is chosen over this search space based on a strategy

of efficiently searching the parameter space by Huttenlocher *et al.* [55]. The size of the search space is reduced by partitioning the image into blocks using the quadtree partition technique and searching for translations that minimize the partial directed Hausdorff distance between corresponding blocks:

$$h_K(A, B) = K^{th} \min_{a \in A} \min_{b \in B} \|a - b\| \quad (5.16)$$

where  $A$  and  $B$  are edge point sets from both the feature maps respectively. Then  $\theta_o$  becomes the center of a new search interval of length  $2\Delta$ ,  $[\theta_o - \Delta, \theta_o + \Delta]$ . Following the same step as  $\theta_o$ , the new approximated rotation  $\theta_l$  is found within this search interval with an accuracy of  $\Delta/2$ . This process is repeated until the accuracy  $\Delta/2^n$  is less than  $1^\circ$ . Decomposing the main search into above two “sub-search” can dramatically reduce the amount of computations, because the image size is small and only salient edges are considered. For efficiency, we can also choose relatively large steps between each translation. Actually, the search space for translations has been greatly reduced after  $\theta_o$  is estimated.

### **A Point-based Algorithm using Gradient Vector Field**

In this section, we present a new method that is able to handle large rotations and translations in a computationally efficient manner. The method works by first computing the gradient field for each image. Next a two dimensional histogram representing the gradient field distribution is constructed for each image. Because histograms discard the positional information of the gradients they are invariant to translation. Thus, the rotation between the images can be found by matching the gradient field histograms of the two images, if the scale difference between them is small. Histogram matching is performed by correlation after mapping the histograms to polar coordinates, where the rotation is reduced to a 1D shift. The gradient field distribution is computed using Shigeru gradient operators.

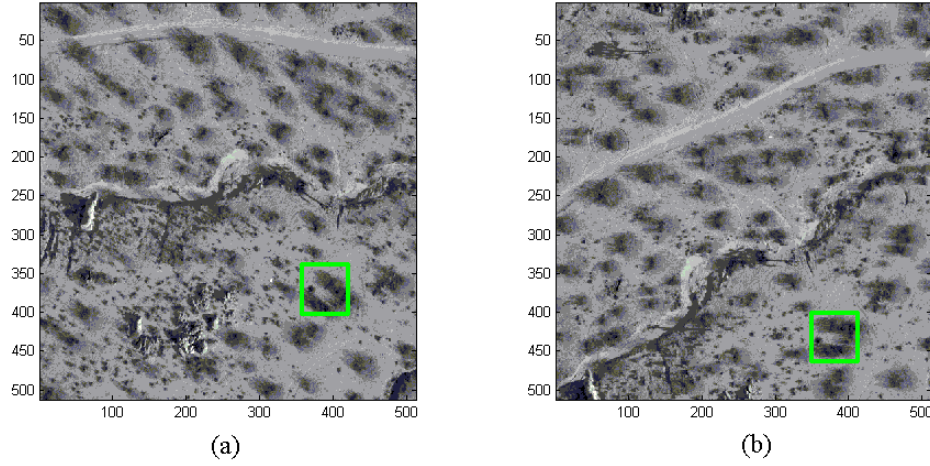
$$\begin{aligned}\nabla f(x, y) &= \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)^T \\ r &= \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2} \\ \theta &= \arctan \left( \frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial x}} \right)\end{aligned}\tag{5.17}$$

It is noticed that the following relation holds:

$$R_\alpha(\nabla f_1(r, \theta)) = \nabla f_1(r, R_\alpha(\theta)) = \nabla f_1(r, \theta + \alpha) \quad (5.18)$$

where  $R_\alpha$  denotes the operation of rotating the gradient field image by  $\alpha$  in a counter-clockwise direction.

### Simulation Results

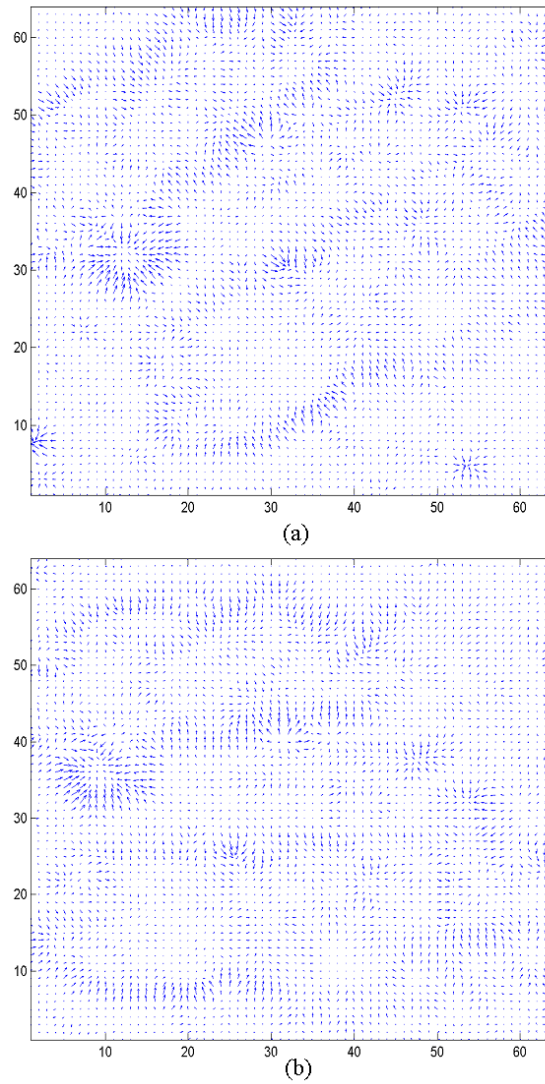


**Figure 18.** Image pair for gradient vector field.

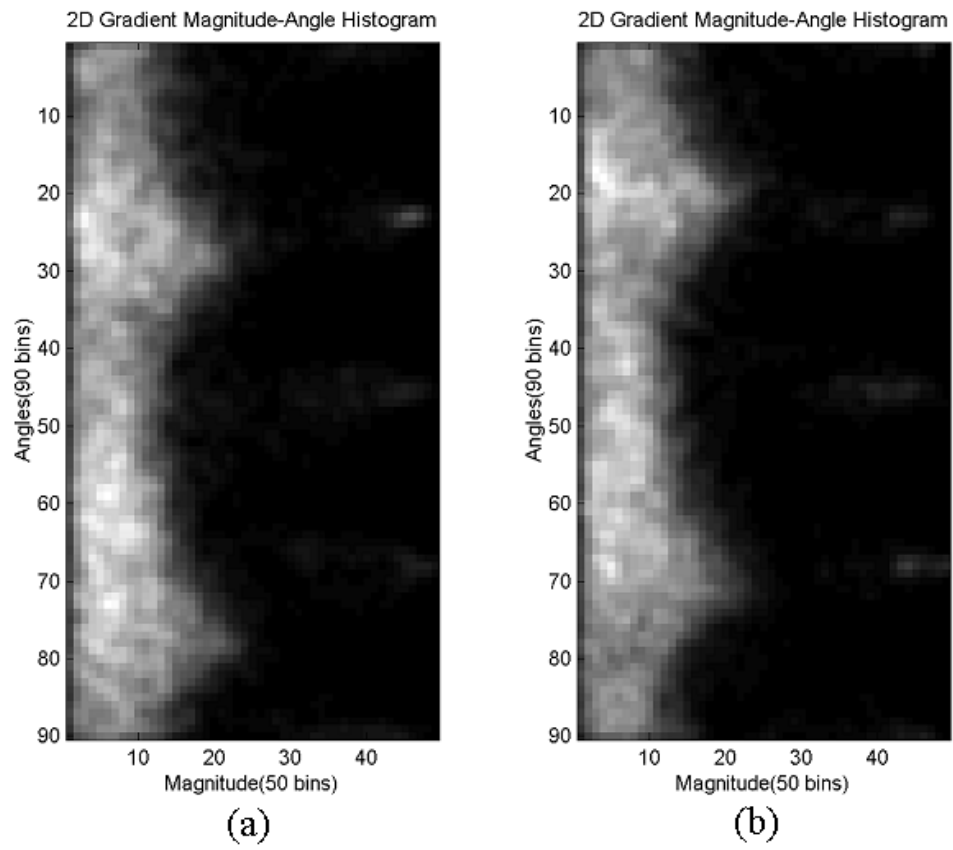
In Figures 18(a) and 18(b), two aerial images are shown. We picked two green rectangular patches of the same area from both images shown in the Figure 18, and compute the gradient field of them. The gradient vector images are shown in Figures 19(a) and 19(b), and the 2D histogram of gradient vector field of Figures 18(a) and 18(b) are shown in Figures 20(a) and 20(b). By correlation, the rotation can be first estimated based on the position of peak of the correlation function shown in Figure 21. After the rotation between images is estimated, the sensed image is compensated. Then, we apply Harris



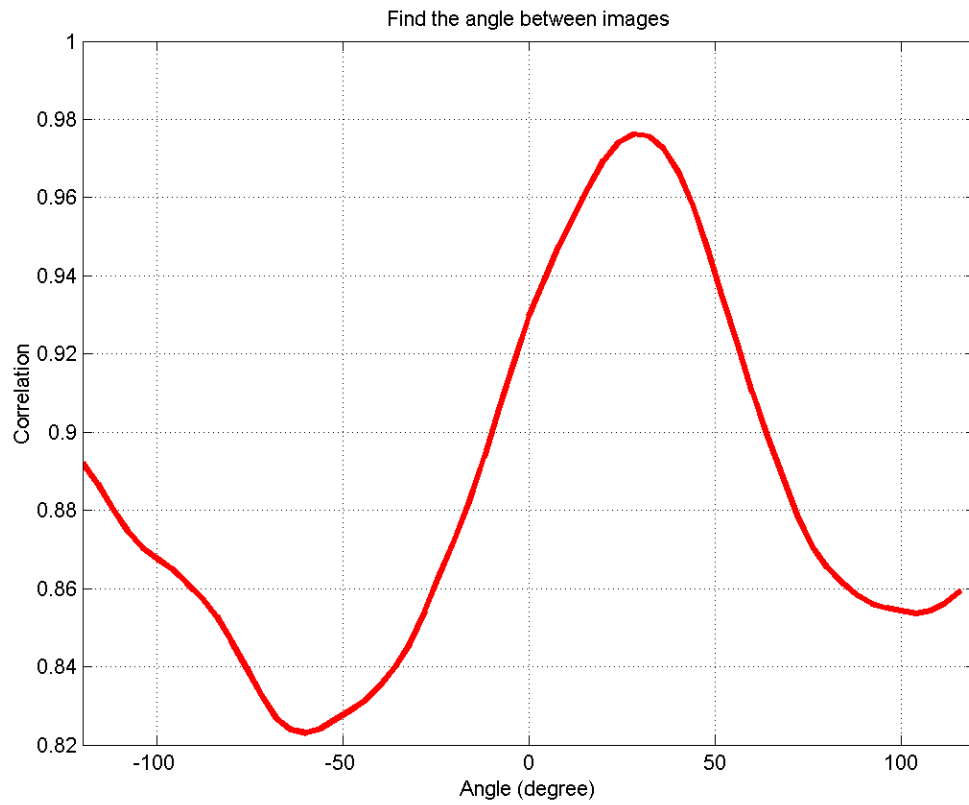
corner detector for feature extraction. The matching pairs between images can be found using correlation-based method directly.



**Figure 19.** Gradient vector fields.



**Figure 20.** 2D histogram of gradient vector fields.



**Figure 21.** Estimate rotation by correlation.

## CHAPTER SIX: GRADIENT-BASED MOTION ESTIMATION

Optical flow is a 2D image motion measure that has wide range of applications in computer vision, video coding and computer graphics, which mainly come from different applications than those considered here. This chapter presents a model-based registration algorithm using optical flow estimation.

### Model-based Registration

The usual starting point for velocity estimation is to assume that the intensities are shifted from one frame to the next, and that the shifted intensity values are conserved. Actually, the intensity conservation assumption is only approximately true in practice, because it ignores possible changes in intensity due to varying illuminant changes. For the AIR applications, the influences of this approximation will be discussed later.

Assuming that  $(u, v)$  are small, we can linearize  $I(x + u\Delta t, y + v\Delta t, t + \Delta t)$  around  $(x, y)$  by a first-order Taylor series expansion:

$$I(x + u\Delta t, y + v\Delta t, t + \Delta t) \approx I(x, y, t) + u\Delta t I_x(x, y, t) + v\Delta t I_y(x, y, t) + \Delta t I_t(x, y, t) \quad (6.1)$$

where  $I_x$ ,  $I_y$  and  $I_t$  are the spatial and temporal partial derivatives of image intensity.

Based on Eq. (6.1), we can obtain the following well-established constraints:

$$I_x(x, y, t)u + I_y(x, y, t)v + I_t(x, y, t) = 0 \quad (6.2)$$

All the quantities in these equations are functions of the position  $(x, y)$  within the image. Thus, every pixel provides one such equation that constrains the displacement of that pixel. However, since the displacement of each pixel is defined by  $u$  and  $v$ , the brightness constraint alone is not sufficient to determine the displacement of a pixel. The second constraint is provided by a global motion model, which describes the variation of the image motion across the whole image. The 2D affine motion model is a very good approximation for the induced image motion when the camera is imaging distant scenes, such as AIR applications. It is described by the equations:

$$u(x, y) = p_1 + p_2x + p_3y \quad (6.3)$$

$$v(x, y) = p_4 + p_5x + p_6y \quad (6.4)$$

We can substitute the affine motion of Eqs. (6.3) and (6.4) into the optical flow constraint equation in Eq. (6.2) to obtain,

$$I_x(p_1 + p_2x + p_3y) + I_y(p_4 + p_5x + p_6y) + I_t = 0 \quad (6.5)$$

The motion vector  $\vec{u}(x, y) = (u(x, y), v(x, y))^T$  can be written as:

$$\vec{u}(x, y; \vec{p}) = X(x, y) \cdot \vec{p} \quad (6.6)$$

where  $\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_6)^T$  and  $X = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix}$ .

Let  $\vec{I}_s = (I_x, I_y)^T$ , the gradient constraint of Eq. (6.5) can be expressed in a matrix form by substituting Eq. (6.6) into Eq. (6.5) as:

$$\vec{I}_s^T \cdot X(x, y) \vec{p} + I_t = 0 \quad (6.7)$$

Hence, every pixel provides one such equation on the six unknown global parameters  $\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_6)^T$ . Since these parameters are global, therefore, theoretically, six independent constraints from six different pixels are adequate to recover these parameters. In practice, however, the constraints from all the pixels within the overlapped region of analysis are combined to minimize the square error:

$$E(\vec{p}) = \sum_{x,y} (\vec{I}_s^T \cdot X(x, y) \vec{p} + I_t)^2 \quad (6.8)$$

$E(\vec{p})$  can be minimized by taking derivatives with respect to  $\vec{p}$ , which gives the least-square solution:

$$\vec{p} = M^{-1} \vec{b} \quad (6.9)$$

where

$$M = \sum_{x,y} X^T \vec{I}_s \vec{I}_s^T X, \quad \vec{b} = - \sum_{x,y} X^T \vec{I}_s I_t \quad (6.10)$$

The well-established gradient constraint equation of Eq. (6.1) is derived by taking

the Taylor expansion of the image constancy up to the first-order terms. The higher-order terms are neglected under the assumption that the motion between consecutive frames is arbitrarily small. Therefore, the lack of higher-order terms becomes the main source of errors in the data constraint. The reliability of the image flow constraint equation depends on the magnitudes of the higher order derivatives of image brightness function. In the next Section, we present a method to overcome this limitation via coarse-to-fine processing, using iterative refinement within a multi-resolution pyramid.

### **Coarse-to-Fine Iterative Estimation**

The basic observation behind coarse-to-fine estimation is that given proper filtering and sub-sampling, the induced image motion decreases as we go from full resolution images (fine pyramid levels) to small resolution images (coarse pyramid levels). Approximation errors in computing the partial derivatives are inevitable due to inaccurate numerical approximation as well as the temporal and spatial aliasing in the brightness function. This hierarchical iterative-refine estimation process can alleviate these errors due to approximation. In order to apply the coarse-to-fine hierarchical approach, we need a method for updating the affine parameters when propagating all the way up to a finer resolution level. Here we propose a reasonably general method that is also used to produce improved estimates through iterative refinement at a given scale. For a given pair of images, we usually select one as the reference image, and warp the other image to register it with the reference image. The registration process is actually the

estimation of affine warping parameters.

When an improved estimate for the affine transformation parameters  $\bar{p} = (p_1, p_2, p_3, p_4, p_5, p_6)^T$  and an initial estimate  $\bar{p}^0$  are available, we consider the first step in the iteration using the following notation

$$\bar{u} = X\bar{p}, \quad \bar{u}^0 = X\bar{p}^0, \quad \Delta\bar{u} = (\bar{u} - \bar{u}^0) = X\Delta\bar{p} = X(\bar{p} - \bar{p}^0) \quad (6.11)$$

Note that the  $\Delta\bar{u}$  and  $\Delta\bar{p}$  allow us to apply the notation to each step in the iteration. By modifying the generalized brightness equation, we create a motion corrected, or warped image  $h(x, y, t)$  to obtain:

$$h(x, y, t) = T[I(x, y, t), \bar{p}^0] = I(x + \bar{u}^0 \Delta t, x + \bar{v}^0 \Delta t, t + \Delta t) \quad (6.12)$$

where  $T[I(x, y, t), \bar{p}^0]$  denotes the warping operation that warps the image  $I(x, y, t)$ . Applying the gradient constraint on the warped image, similar to Eq. (6.7) with  $\bar{p} \rightarrow \Delta\bar{p}$ , we have:

$$\vec{h}_s^T X \Delta\bar{p} + h_t = 0 \quad (6.13)$$

where  $\vec{h}_s$  and  $h_t$  are computed using data from the last step in the iterative process.

Similar to the solution in Eq. (6.9) for the square error  $E$ , we can obtain  $\Delta\bar{p}$  in Eq. (6.13) as:

$$\Delta\bar{p} = \bar{p}^{(k+1)} - \bar{p}^{(k)} = M^{-1} \vec{b} \quad (6.14)$$



where  $M$  and  $\vec{b}$  are as in Eq. (6.10), except with all occurrences of  $I$  replaced with  $h$  to give

$$M = \sum_{x,y} X^T \vec{h}_s \vec{h}_s^T X, \quad \vec{b} = -\sum_{x,y} X^T \vec{h}_s h_t \quad (6.15)$$

Thus, we can iteratively refine the parameters of the affine transformation using

$$\vec{p}^{(k+1)} = \vec{p}^{(k)} + M^{-1} \vec{b} = \vec{p}^{(k)} + \Delta \vec{p} \quad (6.16)$$

where  $k$  is the index of iterations.

The updated affine transformation  $\vec{p}^{(k+1)}$  is a composite transformation of  $\vec{p}^{(k)}$  and  $\Delta \vec{p}$ . Let the affine transformation  $\Delta \vec{p}$  be written as

$$\Delta \vec{p} : \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} 1 + \Delta p_2 & \Delta p_3 \\ \Delta p_5 & 1 + \Delta p_6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \Delta p_1 \\ \Delta p_4 \end{pmatrix} \quad (6.17)$$

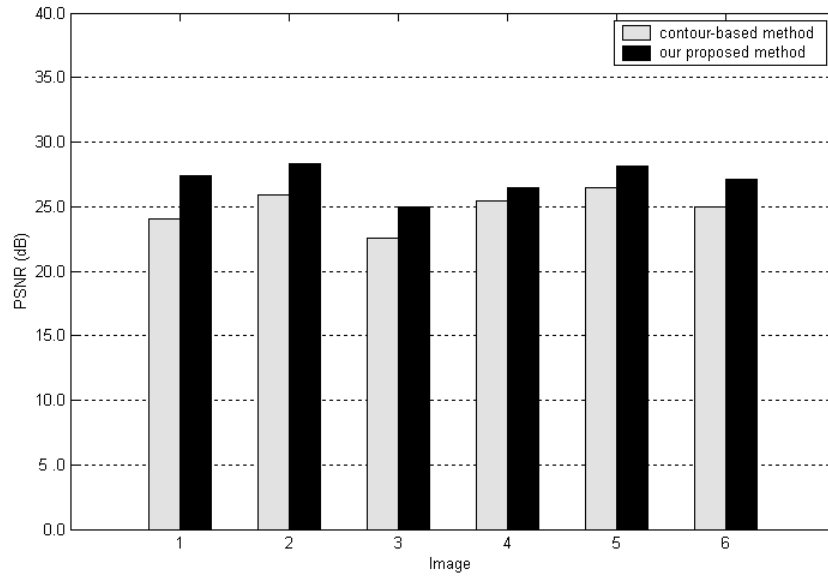
with  $\Delta \vec{p} = (\Delta p_1, \Delta p_2, \dots, \Delta p_6)$  being the affine parameter vector. Then the affine parameters can updated as follows:

$$\begin{pmatrix} p_2^{(k+1)} & p_3^{(k+1)} \\ p_5^{(k+1)} & p_6^{(k+1)} \end{pmatrix} = \begin{pmatrix} 1 + \Delta p_2 & \Delta p_3 \\ \Delta p_5 & 1 + \Delta p_6 \end{pmatrix} \begin{pmatrix} 1 + p_2^{(k)} & p_3^{(k)} \\ p_5^{(k)} & 1 + p_6^{(k)} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (6.18)$$

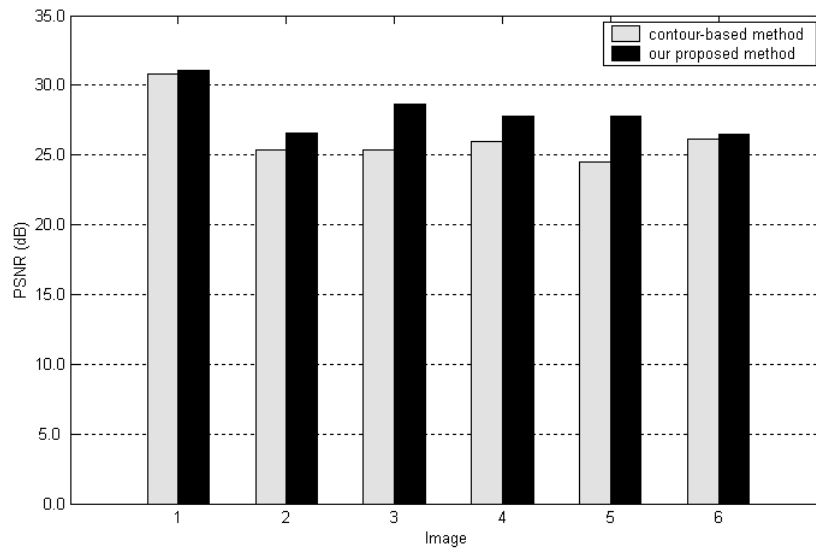
$$\begin{pmatrix} p_1^{(k+1)} \\ p_4^{(k+1)} \end{pmatrix} = \begin{pmatrix} 1 + \Delta p_2 & \Delta p_3 \\ \Delta p_5 & 1 + \Delta p_6 \end{pmatrix} \begin{pmatrix} p_1^{(k)} \\ p_4^{(k)} \end{pmatrix} + \begin{pmatrix} \Delta p_1 \\ \Delta p_4 \end{pmatrix} \quad (6.19)$$

The gradient-based global optimization algorithm starts at the coarsest resolution

level of the pyramid, then following subsequent levels in a coarse-to-fine approach. At each resolution scale, Eq. (6.16) is iterated until a maximal number of iterations are reached or the magnitude of the update of motion parameters reaches a predetermined threshold. Finally, when the procedure stops at the finest resolution scale, the final motion parameters are obtained. In the AIR applications, the induced relative motion is usually small. Therefore, less than ten iterations are needed for an accurate registration and convergence.



**Figure 22.** PSNR for synthetic images



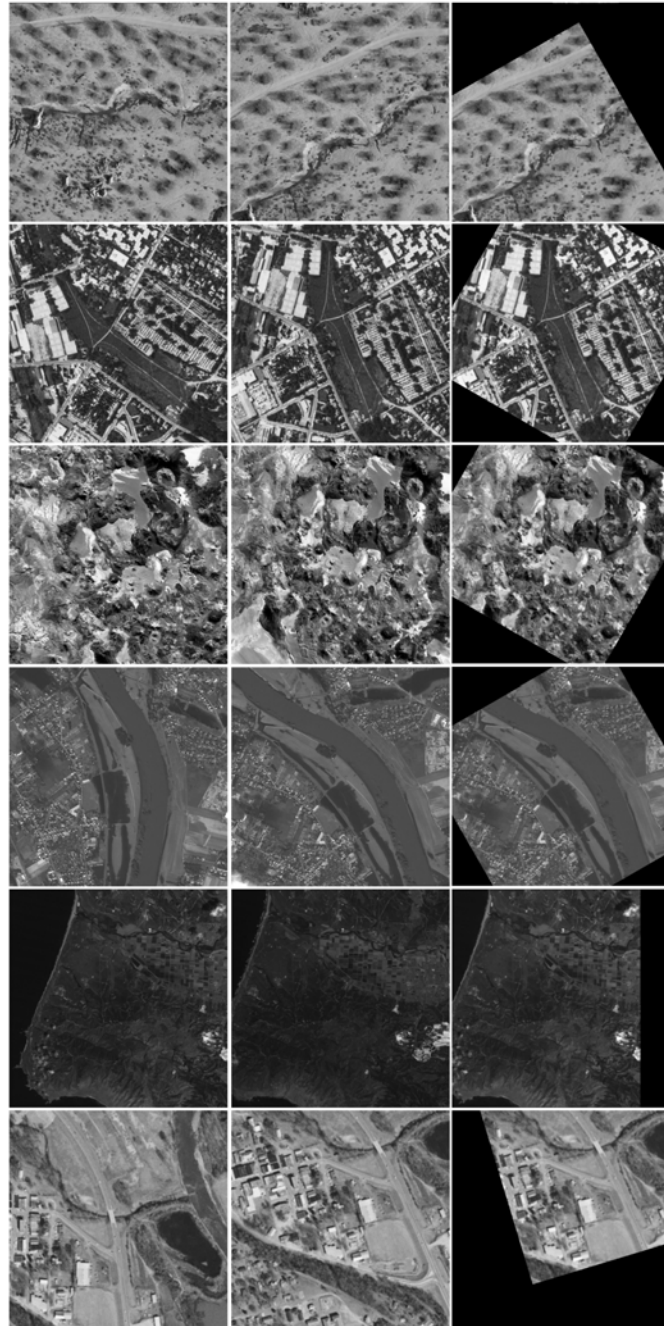
**Figure 23.** PSNR for aerial images.

Thus, the set of estimated transformation parameters in Table 1 are used as the initial  $\bar{p}^0$  to warp one of the image pairs, where every pairs are approximately aligned to each other. To compare the final registration accuracy between the contour-based method and our proposed method, we choose the peak signal to noise ratio (PSNR), which is computed by using

$$PSNR = 20 \log_{10} \left( \frac{255}{RMSE} \right) \quad (6.20)$$

where RMSE is the root mean squared error. The summation for the computation of MSE is based on the all pixels in the overlapping area between the reference image and the warped target image. Fig. 22 shows the compensated image errors at six different combined motions listed in Row “original” of Table 1. From Fig. 22, we can see that the computed values of PSNR using our method are all greater than those based on the contour-based method.

Fig. 24 shows six pairs of aerial images in the first two columns, and the registration results on the 3<sup>rd</sup> column using our proposed hierarchical registration algorithm. For comparison, we also compute the corresponding PSNR values using our method and the contour-based method. Fig. 23 demonstrates that our proposed method is superior to the contour-based approach.



**Figure 24.** Image registration examples.

## **CHAPTER SEVEN: ROBUST ESTIMATION UNDER NON-UNIFORM BRIGHTNESS VARIATION**

In this chapter, we introduce our proposed robust image registration algorithm, where the OFE framework can be extended to provide robust performance under non-spatial distortions. We experimentally demonstrate the accuracy, efficiency, and robustness of the proposed algorithm.

### **Introduction**

Methods for estimating motion that are based on the optical flow equation (OFE) [17-20] assume that the illumination of the scene is uniform. Recently, some researchers tried to relax this brightness constancy assumption and developed algorithms to estimate the optical flow in the presence of illumination variations [21]. Hager and Belhumeur [22] proposed an efficient region matching and tracking algorithm based on robust estimation framework. They modeled the illumination changes into the SSD formulation by using a low-dimensional linear subspace determined from several images of the same object under different illumination conditions. The main disadvantage of this algorithm is the need of several images of the same object under different illumination conditions to compute the linear subspace before the tracking process. Lai [23] explicitly modeled spatial illumination variations by low-order polynomial functions in an energy

minimization framework. Altunbasak et al. [24] proposed a similar model for time-varying illumination and imperfect optics, where the resulting optimization framework estimates the motion parameters, illumination parameters, and camera parameters simultaneously. Haussecker and Fleet [25] used several physical models that describe brightness variations to compute the optical flow.

The BVM-based approach [23,24] basically accounts for smoothly varying illumination changes. However, the situations resulting in the brightness changes between the reference and the sensed images are very complex, and the effectiveness of the BVM-based approach for image registration is sometimes limited. This prompts the necessity to identify an appropriate image representation, on which the OFE-based robust estimation using M-estimator in a coarse-to-fine manner can be incorporated. The primary drawback of the OFE-based estimation is that it may fail unless the two images are misaligned by a moderate difference in scale, rotation, and translation. In order to overcome this problem, we propose a novel point-based registration algorithm to bring the images into approximate alignment, even in the presence of arbitrary rotation angles and a wide range of scale changes. Its purpose is to furnish a good initial estimate to the affine registration module that is based on OFE-based robust estimation using M-estimator with a multi-resolution method. Thus, the image registration is formulated as a two-stage hybrid framework combining both a novel point-based algorithm and robust estimation with M-estimator to register aerial images with spatial and non-spatial distortions. In the first stage, the proposed point-based matching algorithm is applied to the coarsest level of both images in the Laplacian pyramid, and it can provide an initial approximate

estimation of the affine transformation parameters. In the second stage, based on the non-linear quasi-bandpassed image representation, the OFE-based robust estimation with M-estimator is incorporated to precisely register images using a multi-resolution method. Fig. 25 describes the block diagram of the proposed hybrid approach, where  $L_0, L_1, L_2$  are levels of the Laplacian pyramid, and  $G_3$  is the coarsest level of Gaussian pyramid. The basic components of this framework are: (1) pyramid construction [56], (2) non-linear image transformation, (3) feature point extraction, (4) point-based parameter estimation, (5) image warping, (6) robust motion estimation, and (7) coarse-to-fine hierarchical refinement.

### **A Robust Image Registration Algorithm**

Optical flow [17-20] is a 2D image motion measure that has a wide range of applications in computer vision, video coding, and computer graphics, which mainly come from different applications than those considered here. Optical flow formulations assume brightness constancy, i.e., they estimate the 2D velocity of points of constant image brightness. Let  $I(x, y, t)$  be the image brightness at a point  $(x, y)$  at time  $t$ .





The brightness constancy constraint can be expressed as follows:

$$I(x, y, t) = I(x + u\Delta t, y + v\Delta t, t + \Delta t) \quad (7.1)$$

Here  $(u, v)$  is the horizontal and vertical image velocity at a point and  $\Delta t$  is small.

This simply states that the image value at time  $t$ , at a point  $(x, y)$ , is the same as the value in a later image at a location offset by the optical flow. But in our framework, the reference and the sensed image may be acquired at different times under different imaging conditions. The variations of the intensity characteristics between images may be large and non-uniform because of non-spatial distortions. If brightness is not conserved, then the optical flow field estimated from Eq. (7.1) can be a severely biased approximation to the underlying 2D motion field of interest [21]. Therefore, in order to effectively incorporate the OFE-based parametric motion estimation into our proposed framework, *two fundamental questions* should be addressed in the following Sections: (i) what is a good image representation to work with using the OFE-based framework under non-spatial distortions; (ii) the spatial distortions, i.e., the misalignment between images, may exceed certain large values above which OFE-based methods can't converge to the correct result.

### **Image Representation**

The derivation of the optical flow equation assumes that the intensity of a pixel doesn't change along motion trajectory. In our proposed framework, we need to identify

an image representation under which the brightness constancy assumption is still valid, thus the motion between images can be described by the apparent motion of brightness pattern. To capture the common intensity information while suppressing the non-common brightness changes, the image transformation we have chosen is the absolute value of pixels in a Laplacian pyramid, a non-linear image representation. The advantage of using the Laplacian pyramid is that its successive levels are quasi-bandpassed versions of the original signal [56]. The quasi-bandpassed operation ensures that the low spatial frequencies containing the information about brightness changes are substantially removed.

Non-linear quasi-bandpassed representations are useful to image registration with both spatial and non-spatial distortions, because: (1) The creation of such representation images doesn't involve any thresholding, and therefore preserves all image details. This is in contrast to "invariant" representations (e.g., edge maps, edge vectors, contours, point features), which eliminate most of the detailed variations within local image regions. (2) A pyramid data structure of the non-linear quasi-bandpassed image representation facilitates a coarse-to-fine search based on signal details, which can't be directly applied by modeling spatial brightness variations with low-order polynomial functions[23,24].

### **Robust Estimation using a Direct Method**

In our research, we found gradient-based algorithms appear to be most suitable for applications under consideration. Based on the non-linear image representations of both images, the optical flow equation can be written as follows:

$$\Phi_x(x, y, t)u(x, y, t) + \Phi_y(x, y, t)v(x, y, t) + \Phi_t(x, y, t) = 0 \quad (7.2)$$

where  $\Phi(x, y, t)$  is the image function after the non-linear transformation,  $(u, v)$  is the motion vector, and  $\Phi_x$ ,  $\Phi_y$ ,  $\Phi_t$  are the partial derivatives of the image function with respect to  $x$ ,  $y$ , and  $t$ , respectively. This equation is derived based on the first-order Taylor series approximation and the assumption of brightness constancy on the non-linear image representation  $\Phi(x, y, t)$ . Dropping the terms above first order and simplifying give the following gradient-based formulation of the objective function:

$$E_D(u, v) = \sum_{(x, y) \in \mathfrak{R}} (\Phi_x(x, y, t)u(x, y, t) + \Phi_y(x, y, t)v(x, y, t) + \Phi_t(x, y, t))^2 \quad (7.3)$$

This objective function can be written as [57]

$$E_D(u, v) = \sum_{\mathfrak{R}} ((\nabla \Phi)^T \mathbf{u} + \Phi_t)^2 \quad (7.4)$$

where  $\nabla \Phi$  denotes the local gradient vector, and  $\mathbf{u} = [u, v]^T$  denotes the flow vector. Exploiting the fact that the motion field is smoothly varying, several attempts have been made to describe it using a parametric model described by a few parameters. In these cases, the motion vector of a pixel is completely represented by the model parameters and its locations. Given a vector of model parameters  $\mathbf{a}$ , the motion may be expressed as  $u = f_1(x, y; \mathbf{a})$  and  $v = f_2(x, y; \mathbf{a})$ , where  $f_1$  and  $f_2$  determine the motion model. Common parametric motion models are as follows.

1. The affine motion model

$$\begin{aligned} u(x, y) &= a_1x + a_2y + a_3 \\ v(x, y) &= a_4x + a_5y + a_6 \end{aligned} \tag{7.5}$$

2. The bilinear motion model

$$\begin{aligned} u(x, y) &= a_1 + a_2x + a_3y + a_4xy \\ v(x, y) &= a_5 + a_6x + a_7y + a_8xy \end{aligned} \tag{7.6}$$

3. The perspective model

$$\begin{aligned} u(x, y) &= \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \\ v(x, y) &= \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \end{aligned} \tag{7.7}$$

In this chapter, we focus on the estimation of affine transformation for image registration. However, the proposed framework can be easily extended to other two global transformations in Eqs (7.6) and (7.7). Our robust formulation follows on the lines of standard M-estimation techniques as is also used by Black and Anandan [57], and Odobez and Bouthemy [58]. This is employed in a direct estimation framework popularized by Bergen et al. [18]. In the M-estimation formulation, the unknown parameters are estimated by minimizing an objective function of the residual error. In particular, the following minimization problem is solved:

$$\min_{\mathbf{a}} E_D(r; \sigma) \quad \text{where} \quad E_D(r; \sigma) = \sum_{\mathfrak{R}} \rho((\nabla \Phi)^T \mathbf{u}(x, y; \mathbf{a}) + \Phi_i; \sigma) \quad (7.8)$$

where  $\rho(r; \sigma)$  is the robust  $\rho$ -function defined over the residuals,  $r$ ; with a given scale factor,  $\sigma$ . In this work, we used the *Lorentzian* function, which is given as follows:

$$\rho(r; \sigma) = \log \left( 1 + \frac{r^2}{2\sigma^2} \right) \quad (7.9)$$

The M-estimation for the parameters  $\mathbf{a}$ , based on the  $\rho(r; \sigma)$  function in the minimization problem of Eq. (7.8), is the parameter  $\mathbf{a}$  that is a solution of the  $K$  equations [59],

$$\sum_i w(r_i) r_i \frac{\partial r_i}{\partial \mathbf{a}_k} = 0, \quad w(t) = \frac{\rho(t)}{t}, \quad k = 1 \dots K, \quad (7.10)$$

where  $w(t)$  is called the weight function, and  $K$  is the number of unknown parameters, the dimension of  $\mathbf{a}$ . However, instead of solving this non-linear system of equations, we use an alternative by Sawhney et al. [60] to apply the Gauss-Newton (GN) method to the original minimization problem. With the introduction of a particular approximation, this leads to an iterated re-weighted least squares method (IRLS) [60]. This estimation is a particular form of M-estimation and is also called W-estimation [59].

GN method for parameter estimation is an approximation to the general Newton's method for problems involving minimization of the sum of some functions of the unknown parameters and the measurements, for instance the problem in Eq. (7.8). The second order terms in the Hessian of the error function are ignored. When the GN

formulation is applied for M-estimation, it is seen that a particular approximation of the weight terms leads to descent directions and iterated weighted least squares method for the  $\rho$  function under consideration. In the GN method, given a solution,  $\mathbf{a}^{(m)}$  at the  $m$ -th step, the descent direction,  $\Delta \mathbf{a}^{(m)}$ , is given by [61]

$$\Delta \mathbf{a}^{(m)} = -H^{-1}(\mathbf{a}^{(m)})g(\mathbf{a}^{(m)}), \quad (7.11)$$

and

$$\mathbf{a}^{(m+1)} = \mathbf{a}^{(m)} + \lambda \Delta \mathbf{a}^{(m)} \quad (7.12)$$

for some positive  $\lambda$ .  $H(\mathbf{a}^{(m)})$  is the approximation to the Hessian of the objective function in Eq. (7.8), involving only the first derivatives of the residuals, and  $g(\mathbf{a}^{(m)})$  is its gradient, both defined at the current  $\mathbf{a}^{(m)}$ . Writing the  $g$  and  $H$  in terms of  $\rho$  and  $r_i$ , we get

$$g_k = \sum_i \frac{\partial \rho}{\partial r_i} \frac{\partial r_i}{\partial \mathbf{a}_k} \quad H_{kl} = \sum_i \frac{\partial^2 \rho}{\partial r_i^2} \frac{\partial r_i}{\partial \mathbf{a}_k} \frac{\partial r_i}{\partial \mathbf{a}_l}, \quad (7.13)$$

as the  $k$ -th and  $kl$ -th elements of  $g$  and  $H$ , respectively. Thus,  $\Delta \mathbf{a}$  can be written in terms of these components as the solution of  $K$  linear equations

$$\sum_l H_{kl} \Delta \mathbf{a}_l = -g_k, \quad k, l = 1 \dots K. \quad (7.14)$$

For the non-quadratic  $\rho$ ,  $\frac{\partial^2 \rho}{\partial r_i^2}$  could be negative, therefore the solution to Eq.

(7.11) may not be a descent direction. For the Lorentzian function,  $\frac{\partial^2 \rho}{\partial r_i^2}$  is computed as:

$$\frac{\partial^2 \rho}{\partial r_i^2} = \frac{4\sigma^2 - 2r^2}{(2\sigma^2 + x^2)^2} \quad (7.15)$$

If we approximate  $\ddot{\rho}(r)$  with its secant approximation [62],  $\frac{\partial \dot{\rho}}{\partial r_i}$ , which is positive

everywhere, then the GN equations become,

$$\sum_l \sum_i \frac{\dot{\rho}(r)}{r_i} \frac{\partial r_i}{\partial \mathbf{a}_k} \frac{\partial r_i}{\partial \mathbf{a}_l} \Delta \mathbf{a}_l = - \frac{\dot{\rho}(r_i)}{r_i} r_i \frac{\partial r_i}{\partial \mathbf{a}_k}, \quad k, l = 1 \dots K, \quad i = 1 \dots N. \quad (7.16)$$

From Eq. (7.16), it is noticed that the corresponding equations for the robust estimator are simply weighted normal equations with the weight of each measurement  $i$  being  $\frac{\dot{\rho}(r_i)}{r_i}$ .

$$\frac{\dot{\rho}(r)}{r} = \frac{2}{2\sigma^2 + x^2} \quad (7.17)$$

Based on the plots [60] of the  $\rho$  function and the weights,  $\frac{\dot{\rho}(r_i)}{r_i}$ , it is apparent that  $\rho$  decreases the influence of large residuals on the solution rapidly. The parameter  $\sigma$  that controls the location of the inflection point in the curve, governs the point beyond



which there is a faster decrease in the influence.

There are basically two strategies for dealing with scale in the regression problem: estimate  $\sigma$  beforehand or estimate parameters  $\mathbf{a}$  and  $\sigma$  simultaneously [59]. We choose the first method to compute  $\sigma$ . This means that before each iterative step we choose a scale estimator and calculate its value  $\hat{\sigma}$ . Then, considering  $\hat{\sigma}$  as a known and fixed constant, we proceed with M-estimation for parameters  $\mathbf{a}$ . In this setting, the most commonly used resistant scale is the median absolute deviation (MAD) [59] estimate given by

$$\hat{\sigma} = 1.4825 \operatorname{median}_i \left( \left| r_i - \operatorname{median}_j(r_j) \right| \right) \quad (7.18)$$

The median based estimate has excellent resistance to outliers.

Finally, starting at the coarsest resolution level with  $\mathbf{a}^{(0)} = \mathbf{a}_0$  initially estimated from the point-based algorithm in Section 3, the following steps are performed at each resolution level:

1. Compute the residues  $r_i$  and the associated gradient vector  $\mathbf{g}_i = \frac{\partial r_i}{\partial \mathbf{a}}$  and  $r_i$  is the left-hand side of Eq. (7.2) at the current solution  $\mathbf{a}^{(m)}$ .
2. Update the scale parameter by Eq. (7.18).
3. Compute the weight  $\tau_i$  associated with each data constraint by  $\tau_i = \frac{\dot{\rho}(r_i)}{r_i}$ .
4. Form the weighted Hessian matrix  $H = \sum_i \tau_i \mathbf{g}_i \mathbf{g}_i^T$  and the weighted gradient

$$\text{vector } g = \sum_i \tau_i r_i g_i .$$

5. Update the solution by Eqs. (7.11) and (7.12).
6. Set  $m = m + 1$ .
7. If  $\|\mathbf{a}^{(m+1)} - \mathbf{a}^{(m)}\| < \varepsilon_1$ , go to next resolution level; else go back to step 1.

### **The Hierarchical Algorithm and Selective Data Sampling**

Given the GN formulation and the step for  $\sigma$  estimation, we embed these in a hierarchical coarse-to-fine direct method. Starting at the coarse level, given an initial estimate of the parameters  $\mathbf{a}^{(0)} = \mathbf{a}_0$  shown in Figure 1, the sensed image  $\Phi_1$  is warped so that  $\Phi_1^w(p; \mathbf{a}^{(0)}) = \Phi_1(p - u(p; \mathbf{a}^{(0)}))$ . At this step, the residual  $r$  at  $p$  is defined as

$$r = \Phi_2(p + \Delta \mathbf{u}(p; \mathbf{a})) - \Phi_1^w(p; \mathbf{a}^{(0)}) \quad (7.19)$$

where  $\Delta \mathbf{u}$  is a small unknown increment in  $\mathbf{u}$ . The robust  $\sigma$  estimate is computed using the residuals  $r$ 's defined over all  $p$ 's within the region of interest. Now a GN step is performed with the *Lorentzian* function  $\rho$  to compute a new GN direction  $\Delta \mathbf{a}^{(0)}$  using Eq. (7.16). A line minimization along this direction is performed to get the local minimum solution for the current iteration. These iterations at any level are repeated until the change in parameters is below a threshold or a specified number of iterations are reached. The estimated parameters are projected to the next finer level and used as an

initial estimate to warp the corresponding image  $\Phi_1$ , and the process repeated until convergence at the finest level.

In order to augment the efficiency of the proposed algorithm, a selective data-sampling scheme is applied to reduce the computational cost. The objective function to be minimized is the sum of the  $\rho$ -functions of the data constraints computed at the pixels  $(x, y)$  within the region of interest. However, this may cause very high computation cost for the estimation of a small number of model parameters. Thus, we proposed to select a sparse set of locations to form the objective function. The selection scheme is designed based on the consideration of efficiency as well as reliability. In the proposed selection scheme, we first partition the image into  $m \times n$  uniform blocks. In the experiment, we found 400 constraints are enough to provide accurate registration results based on an affine parametric model. So we set  $m \times n$  to be roughly 400. Then, we select a location in each block to form the objective function. For each pixel  $(x, y)$  in the block, we compute a local normalized-correlation surface  $NC(x, y)$  around the displacement  $\mathbf{u}(x, y; \mathbf{a})$ . In our current implementation, the correlation surface is estimated only within a radius  $d = 1$  around  $\mathbf{u}(x, y; \mathbf{a})$ , where the radius  $d$  is determined by the size of masks  $(3 \times 3)$  used for discretely estimating the first and second order derivatives of  $NC(x, y)$  at  $\mathbf{u}(x, y; \mathbf{a})$  using Beaudet's masks [63]. We define a reliability measure  $\delta(x, y)$  associated with the data constraint at pixel  $(x, y)$  as the inverse of the sum of minimum distance errors in a quadratic model fitting to the local normalized-correlation surface  $NC(x, y)$ , i.e.,

$$\delta(x, y) = \frac{NC_x^2(x, y) + NC_y^2(x, y)}{e(x, y) + \varepsilon_2} \quad (7.20)$$

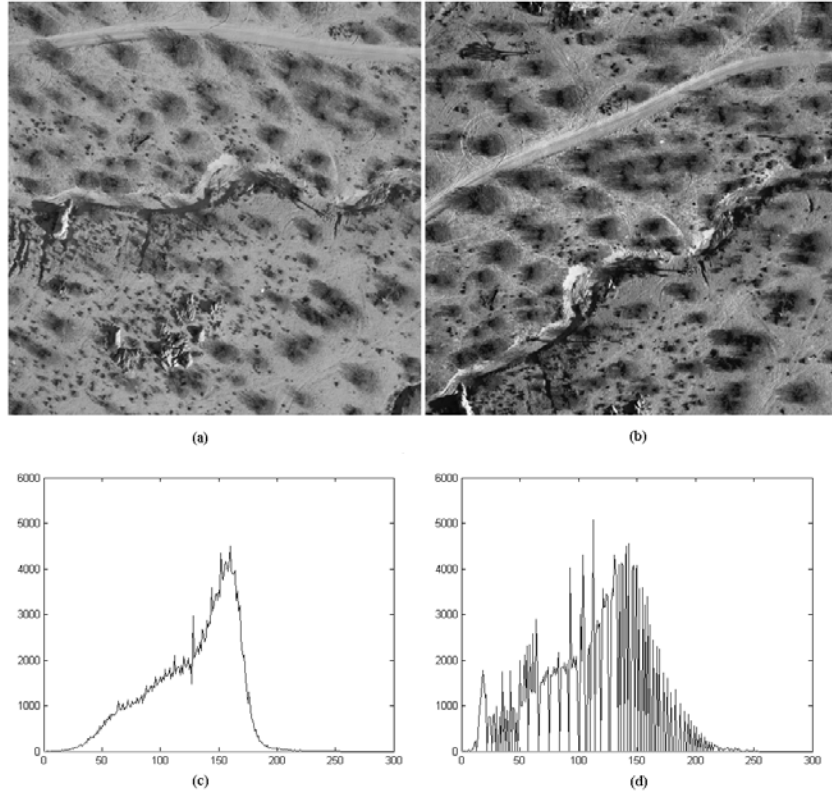
where  $e(x, y)$  is the sum of squared errors from the quadratic fitting in the local neighborhood of pixel  $(x, y)$ ,  $NC_x(x, y)$  and  $NC_y(x, y)$  the partial derivatives, and  $\varepsilon_2$  a small positive number to prevent instabilities when  $e(x, y)$  is very small. Thus, our selection of a reliable data constraint in each block is simply to find pixel  $(x, y)$  with the maximum reliability measure  $\delta_{\max}$ .

### **Simulation Results**

In order to illustrate the procedures, and evaluate the performance of the proposed robust registration algorithm, we have designed four sets of experiments: the first uses a pair of aerial images under irregular brightness changes, and the second set of 4 pairs (Test A-D) was captured with non-spatial distortions, the third set of 4 frame pairs (Test E-H) uses an aerial video sequence with small illumination changes, and the fourth set consists of two pairs of Landsat TM images (Test I-J) with salient brightness changes. To evaluate the generality, accuracy, and effectiveness, we compare its performance with that of the BVM-based approach. In our experiments, both the point-based method and robust estimation performances were addressed. The size of the test image pairs is  $512 \times 512$ .

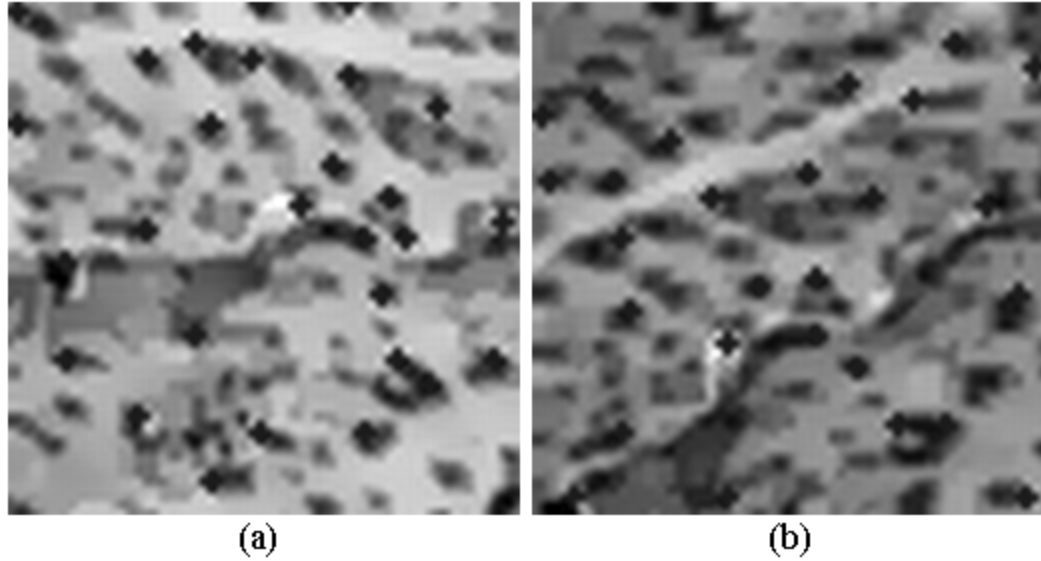
## Experiment Scheme I

For the first set of experiments, our proposed point-based registration algorithm is applied at the highest level of decomposition of the Laplacian pyramid. Figures 26(a)-(b) show two different views of an area in the Dojave Desert, and the histograms are displayed in Figures 26(c)-(d) respectively. In order to test the performance of our proposed hybrid approach under irregular brightness changes, we transformed the histogram of Figure 26(b) to a specified histogram shown Figure 26(d) using the technique of histogram matching.



**Figure 26.** Affine Registration Example I

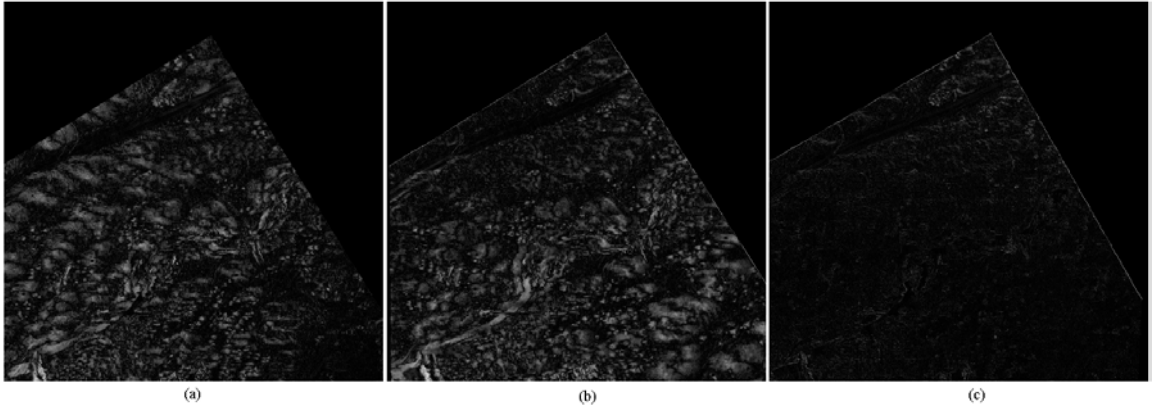
In the experiment, we limit our decomposition level to be 4, so that the coarsest image level uses  $64 \times 64$ . It has been experimentally found that the size of  $64 \times 64$  is appropriate since the images still retain adequate information for feature extraction. The computational cost at this first stage is much less because only salient features are considered. In Figure 27(a)-(b), one can see that salient feature points marked by black “+” were detected using Harris’ method. Note that the images in Figure 27 have been enlarged for the purpose of illustration. Our point-based algorithm correctly identified all true matching pairs, from which initial similarity transformation parameters can be easily estimated. In fact, only 2 matching point pairs are required to compute the similarity transformation.



**Figure 27.** Extracted feature points at the coarsest scale marked by “+”.

Next, the robustness of the proposed point-based algorithm is tested. Given the

reference and the sensed images  $A$  and  $B$  with  $m$  and  $n$  extracted feature points respectively, we performed a Monte Carlo simulation similar to [], where similarity transformation parameters, i.e. translations and rotation, are randomly selected. We ran 40 trials for each combination of parameters. Furthermore, all true point-pairs are slightly displaced to simulate feature extraction errors. A trial is a success if the translation errors are within  $\pm 10$  pixels in both horizontal and vertical directions for original images and  $\pm 5$  degrees for rotation errors. The success ratio is the ratio of successful trials to the total number of trials. The simulation results are shown in Table 1. It is noticed that slight point displacement has little effect on the success ratio, but the number of incorrect points does.

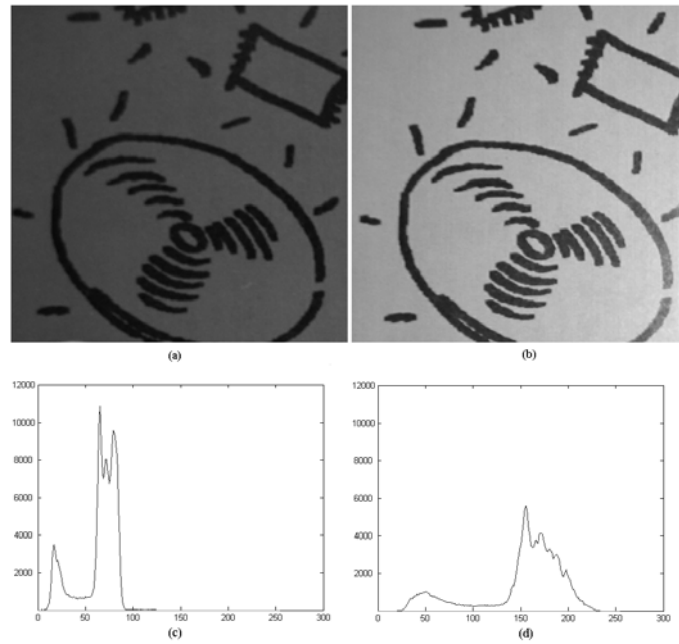


**Figure 28.** Image Difference

## Experiment Scheme II

For the second set of experiments, we first compare the final registration accuracy between our robust estimation algorithm and the BVM-based approach using the image

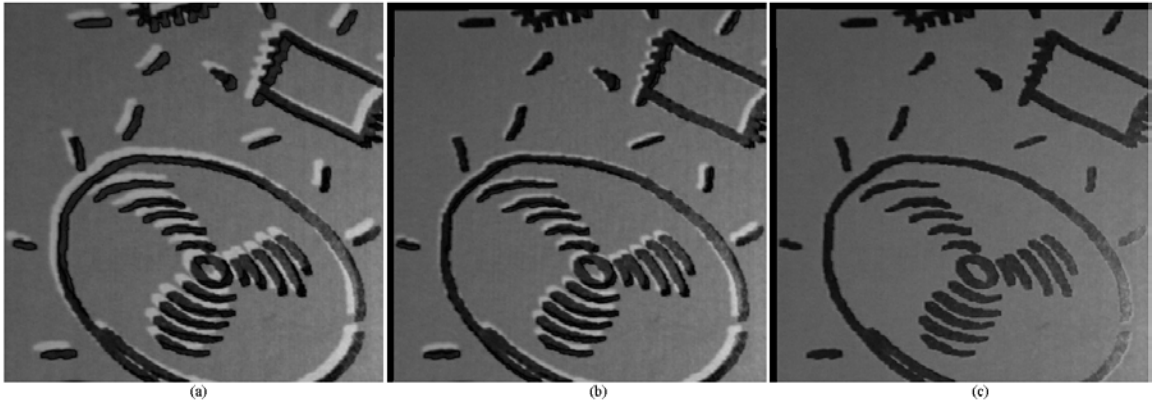
pair of Figures 26(a) and 26(b). Figure 28(a) shows the motion compensated image differences using our proposed point-based algorithm for initial matching. We can observe that the point-based initial matching can provide a good estimate although only matching control points in the coarsest level images are used. Based on this initial matching, we applied both the BVM-based approach and our robust estimation algorithm, and the results are shown in Figures 28(b) and 28(c), respectively. A close examination of Figures 28(b) and 28(c) indicates that our method greatly reduces registration errors from the previous point-based intermediate step, whereas the BVM-based approaches fail completely and its motion estimates are totally erroneous. The brightness variations between the reference and sensed images are probably not spatially varying illumination multiplication and bias factors as low-order polynomials.



**Figure 29.** Affine Image Registration



Second, we tested our proposed robust estimation algorithm and the BVM-based approach on a group of images consisting of four test image pairs captured with a digital camera, which introduces considerable non-spatial distortions. The first pair of Test A includes an image of the top cover printed on the AT89/90 Micro-controller Box, and the other is an image of the same scene under flashlight, with some object movements (translations only). In the similar manner, we obtained the remaining pairs of Test B (translations, rotation only), Test C (translations, rotation, and scale only), and Test D (small translations only). The motions were not synthetic, and induced by movement of the object being photographed. These were intended to compare the performance of our proposed algorithm and the BVM-based approach. Since the true motion is not known, we consider using the normalized correlation  $Corr2$  between the overlapping areas of the image pairs, which is defined as

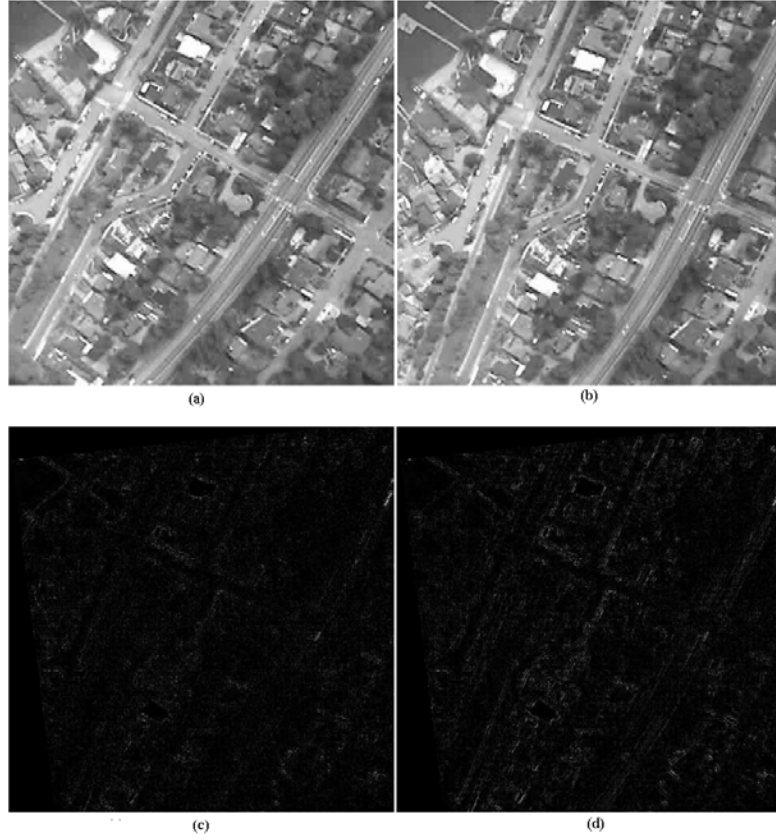


**Figure 30.** Image Differences

**Table 3.** NCC Comparison between the BVM-based approach and the proposed method.

	Test A	Test B	Test C	Test D
BVM-Based Approaches	0.7803	0.8308	0.7643	0.8575
Our Proposed Method	0.9664	0.9553	0.9556	0.9229

Table 3 shows the registration accuracy in terms of the normalized correlation  $Corr2$  for the four test image pairs. The results show that our proposed robust estimation algorithm always outperforms the BVM-based approach. The image pair of Test A is shown in Figures 29(a)-(b), and the corresponding histograms in Figures 29(c)-(d), respectively. The motion compensated image differences using BVM-based approach and our proposed method are depicted in Figures 30(b) and (c), with the direct image difference without motion compensation shown in Figure 30(a). It is noticed that the final accuracy of our proposed method is much better than that of BVM-based approaches, although the latter also reduces the error as indicated in the computed correlation coefficient  $Corr2$  of Table 2 and in Fig. 30(b).



**Figure 31.** Video Frame Registration Example

### Experiment Scheme III

With the third set of experiments (Test E-H), we would like to evaluate the performance of the BVM-based approach and our proposed method using an aerial video sequence with small illumination changes. We calculated the frame difference correlation  $Corr2$  values for each pair of frames after motion compensation. For visual comparison, we selected the frame pairs of Test E depicted in Figures 31(a) and 31(b). The motion compensated frame difference images corresponding to the BVM-based and our proposed

robust estimators are shown in Figures 31(c) and 31(d). As can be seen from the results in Table 4, our proposed robust estimator performs almost as well as the BVM-based approach under small illumination changes. The slight difference in performance results from the fact that the BVM-based approach assumes a more general model than our proposed method, and thus it needs to estimate more model parameters.

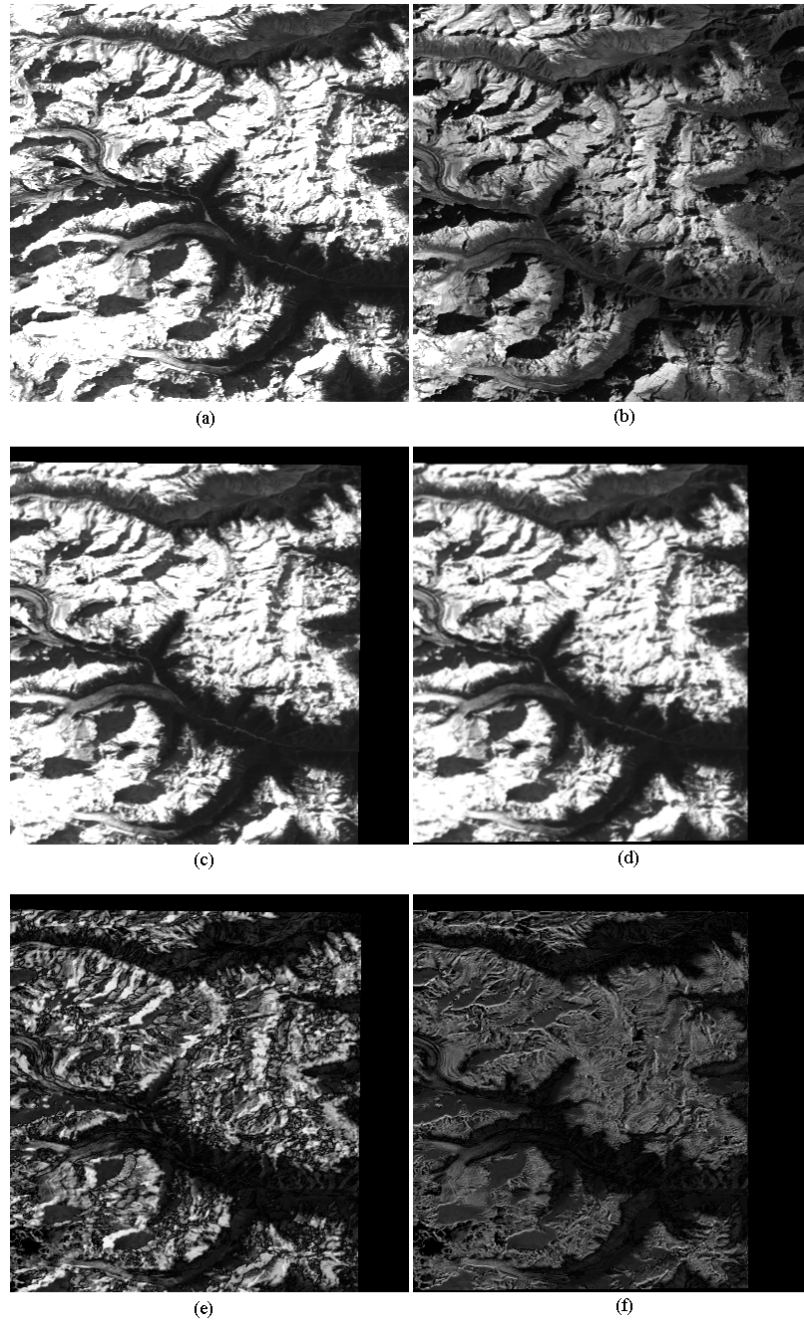
**Table 4.** NCC Comparison between the BVM-based approach and the proposed method.

	Test E	Test F	Test G	Test H
BVM-Based Approaches	0.9722	0.9745	0.9783	0.9717
Our Proposed Method	0.9818	0.9794	0.9782	0.9745

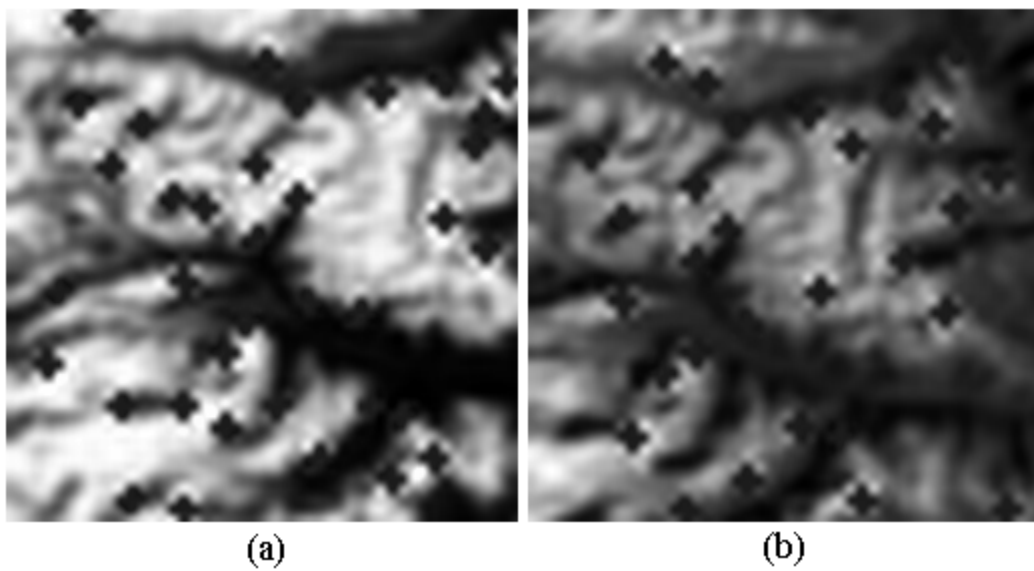
#### Experiment Scheme IV

The fourth set of experiments consists of two pairs of Landsat TM images. They were chosen to measure the effectiveness of our proposed method for image pairs under salient brightness changes. The first image pair of Test I is shown in Figures 32(a) and 32(b), respectively. Since the spatial distortion between Figures 32(a) and 32(b) is large, we first applied the proposed point-based algorithm for initial matching, where the extracted feature points are marked by black “+” in Figures 33(a) and 33(b), respectively. Based on the founded matching pairs of points, similarity transformation parameters were estimated, and thus used to approximately register the sensed image of Figure 32(a) with the reference image of Figure 32(b). The initially registered image is shown in Figure

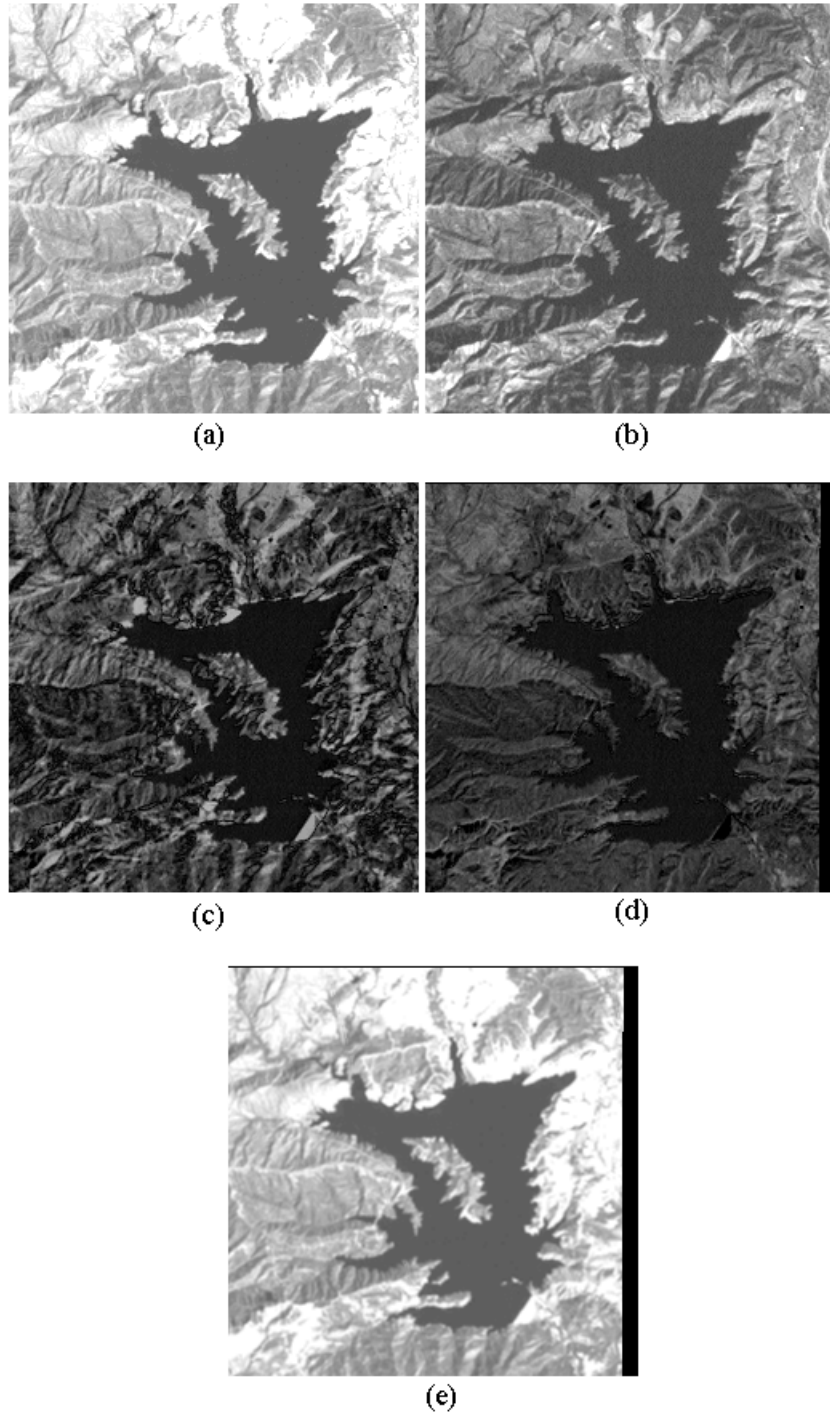
32(c), and we could observe the initial matching errors from the image difference in Figure 32(e). The final registration result using our proposed robust estimation algorithm was shown in Figure 32(d), and the compensated image difference in Figure 32(f).



**Figure 32.** Affine Registration Example (Landsat TM images)



**Figure 33.** Extracted feature points at the coarsest scale marked by “+”.



**Figure 34.** Affine Image Registration (Landsat TM Images)



**Table 5.** NCC Comparison using our method.

	Test I	Test J
Image Difference without Motion Compensation	0.6280	0.4871
Our Proposed Method	0.9133	0.8742

For Test J, we followed the same procedure as above for the second pair of images of Figures 34(a) and 34(b) except that no initial point-based matching was performed because of small misalignment between them. The direct image difference was shown in Figure 34(c), the compensated image difference using our proposed method in Figure 34(d), and the registered image in Figure 34(e). For comparison, we also applied the BVM-based approach to Test I and Test J. However, it doesn't converge to the correct result in both cases. From the correlation coefficient  $Corr2$  of Table 5, it is noticed that the final accuracy of our proposed method is much better compared with that without motion compensation.

In this chapter, we have proposed a hybrid hierarchical approach to the registration problem under spatially non-uniform brightness variations. Based on a non-linear quasi-bandpassed image representation, the image registration is formulated as a two-stage procedure combining both the point-based algorithm and the robust estimation framework in a coarse-to-fine manner. With a new point-based method applied at the highest level of decomposition, the initial affine model parameters could be first

estimated using similarity transformation. Subsequently, the robust estimation mechanism using M-estimators was incorporated into the proposed hybrid framework for completeness. Applying a point-based method at the highest level of Laplacian pyramid allows the algorithm to exhibit superior convergence range, and a hierarchical iterative searching further enhances the convergence range and speed. As it is experimentally demonstrated, our proposed framework can achieve higher accuracy than the BVM-based approach by a series of intensive experiments on real image pairs.

## **CHAPTER EIGHT: CORRELATION-BASED IMAGE REGISTRATION**

Based on the proposed hybrid framework, this chapter describes a hierarchical image registration algorithm for projective motion estimation. The parameters are computed iteratively in a coarse-to-fine hierarchical framework using a variation of the Levenberg-Marquadt nonlinear least squares optimization method. This approach yields a robust solution that precisely registers images with sub-pixel accuracy.

### **Matching Algorithm**

The primary drawback of the optimization-based approach is that it may fail unless the two images are misaligned by a moderate difference in scale, rotation, and translation. In order to address this problem, we could apply the initial matching algorithms introduced at the Chapter 5. The Levenberg-Marquardt nonlinear optimization algorithm [64] is well suited for performing registration based on a least-squares criterion. In the proposed framework, we cast Levenberg-Marquardt into a multi-resolution framework, using a coarse-to-fine iteration strategy, and propagating estimates for one level of a resolution pyramid from its prior level. Most iterations are carried out at the coarsest level, where the amount of data is so greatly reduced that the computational cost of one iteration is negligible. Once convergence has been reached at any particular level,

a switch to a finer level is made, where only a few iterations are needed because of near-optimal initial conditions. For many types of optimizers, this strategy for convergence is significantly faster than a single-stage approach. In the case of Levenberg-Marquardt, the benefits are even greater because this algorithm is super-linear and converges much faster than most other minimization schemes so long as the initial estimate is close to the correct solution. In addition, a multi-resolution strategy improves robustness, in the sense that it decreases the likelihood of being trapped at a false local optimum.

Any automatic registration method requires the choice of an objective criterion that measures the similarity of the sensed data to the reference. As the optimization criterion, we select  $\varepsilon^2$ , the integrated square difference of the intensity values, named the residue here. Let  $f_R$  be the reference data and  $f_S$  the sensed data. Then, this criterion can be written as:

$$\varepsilon^2 = \iint_{\{x\} \subset R^q} (f_R(x) - Q_p\{f_S(x)\})^2 dx = \|f_R(x) - Q_p\{f_S(x)\}\|^2 \quad (8.1)$$

where  $Q_p(f)$  is a transformation parametrized by  $p$ , and where  $q$  is the space dimension. Such a criterion lends itself well to minimization with respect to  $p$ , and is well understood. In particular, this Euclidean dissimilarity measure is known to be maximum likelihood if the noise is additive, white, and Gaussian. Its drawback is a lack of robustness in the presence of severe outliers, where its minimum may become less pronounced. In the worst case, outliers predominate and the parameter  $p$  for which  $\varepsilon^2$

reaches its minimum can be quite different from  $p_0$ , the minimum in the noiseless case.

Approximation by a finite sum of the criterion given in Eq. (8.1) leads to:

$$\varepsilon^2 \cong \chi^2(p) = \sum_{i=1}^N [I_1(u_i, v_i) - I_2(x_i, y_i)]^2 = \sum_{i=1}^N e^2 \quad (8.2)$$

$$u_i(x, y) = \frac{m_0x + m_1y + m_2}{m_6x + m_7y + 1} \quad (8.3)$$

$$v_i(x, y) = \frac{m_3x + m_4y + m_5}{m_6x + m_7y + 1} \quad (8.4)$$

$$D = m_6x + m_7y + 1 \quad (8.5)$$

Once the initial estimate of projective model parameters were given, we can solve the minimization problem of Eq. (8.2) by using Levenberg-Marquardt algorithm.

$$\begin{aligned} \frac{\partial e}{\partial m_0} &= \frac{x}{D} \cdot \frac{\partial I_1}{\partial u_i} \\ \frac{\partial e}{\partial m_1} &= \frac{y}{D} \cdot \frac{\partial I_1}{\partial u_i} \\ \frac{\partial e}{\partial m_2} &= \frac{1}{D} \cdot \frac{\partial I_1}{\partial u_i} \\ &\dots \\ \frac{\partial e}{\partial m_7} &= \frac{y}{D} \cdot \left( u_i \frac{\partial I_1}{\partial u_i} + v_i \frac{\partial I_1}{\partial v_i} \right) \end{aligned} \quad (8.6)$$

From these partial derivatives, we update the parameters by:

$$m_{k+1} = m_k + \Delta m \quad (8.7)$$

$$\Delta m = (A + \lambda I)^{-1} b \quad (8.8)$$

$$a_{kl} = \sum_i \frac{\partial e}{\partial m_k} \frac{\partial e}{\partial m_l} \quad (8.9)$$

$$b_k = -\sum_i e \frac{\partial e}{\partial m_k} \quad (8.10)$$

### **Simulation Results**

In order to illustrate the procedures, and evaluate the performance of the proposed correlation-based registration algorithm, we have designed a set of experiments using an aerial video sequence with large motion. To evaluate the generality, accuracy, and effectiveness, we compare its performance with that of the approach using affine motion parameters. The size of the test image pairs is  $480 \times 480$ .

The first image pair is shown in Figures 35(a) and 35(b), respectively. Since the spatial distortion between Figures 35(a) and 35(b) is large, we first applied any of the proposed algorithms in the Chapter 5 for initial matching, where similarity transformation parameters are used.

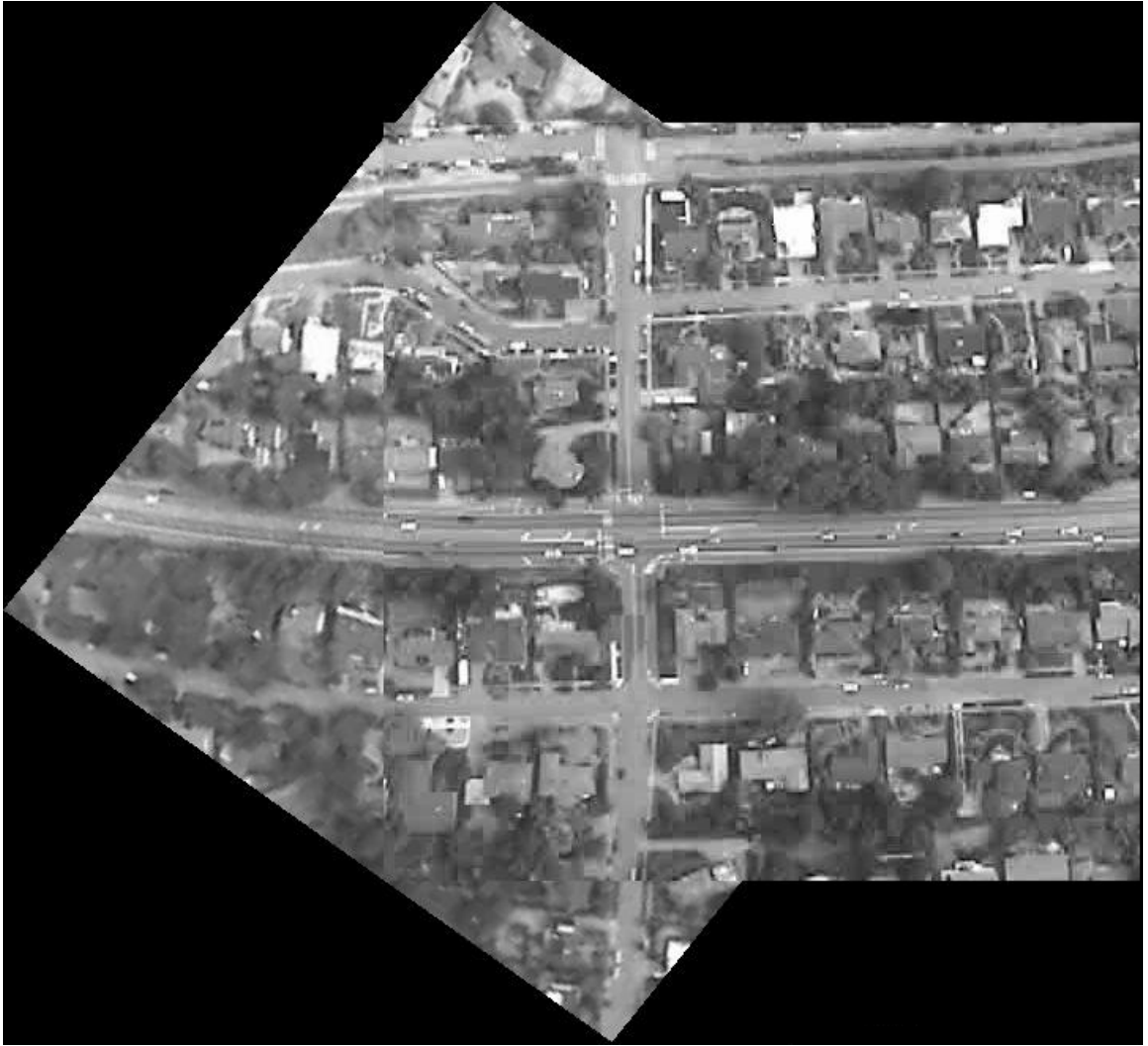


(a)



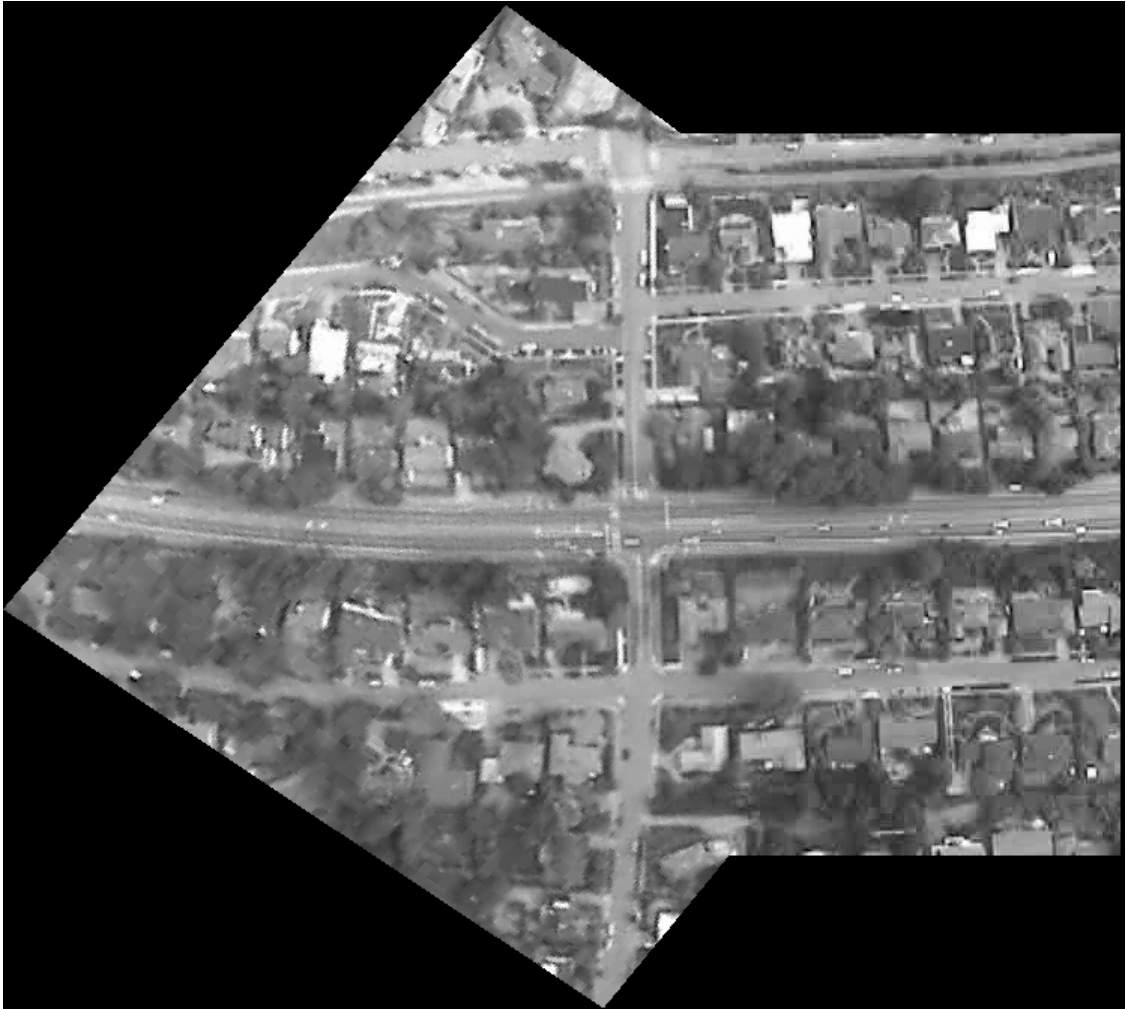
(b)

**Figure 35:** An aerial video frame pairs.



**Figure 36.** Image registration based on initial matching.





**Figure 37.** Image registration using the projective model

**Table 6.** NCC comparison based on the proposed method.

Experiment	Initial matching	Final matching
1	0.8519	0.9729

The registration result after initial matching is shown in Figure 36. A close examination of Figure 36 indicates the matching errors. Based on the proposed method, the final registration result is shown in Figure 37. It is noticed by visual inspection that the registration reduced is greatly reduced, which is also indicated in the Table 6 of NCC comparison between the initial and the final matching.

## CHAPTER NINE: CONCLUSIONS

This research presented in this thesis examined a hybrid and hierarchical approach to image registration. The research presented in this dissertation contains the following key conclusions:

1. In Chapter three, a hybrid and hierarchical image registration framework is proposed, which consists of two stages: *initial matching* and *final matching*. The purpose of the initial matching is to provide a good initial estimate to the second stage of final matching. The first-stage algorithm is applied to the coarsest level of both images. In the second stage, gradient-based algorithms are incorporated to precisely register images using a multi-resolution method.
2. In Chapter four, we have presented a new approach for edge contour extraction based on a three-step procedure. In the first step we obtain greatly improved estimated gradient information from *Shigeru's* operators. The second step applies RCBS edge detection, which is an effective way to control the balance between the two conflicting performance requirements, namely noise immunity and accurate localization. The third step post-processes the resulted edge map using some strategies, which generate qualified edge contours for higher visual processing tasks. For some applications with time constraints, digital filtering

techniques have been applied for solving the problem of regularized cubic B-spline fitting instead of the matrix approaches. The results of this work have been reported in [65].

3. In Chapter five, we proposed *four* initial matching algorithms based on the proposed hybrid and hierarchical registration framework. The effectiveness of these algorithms has been experimentally demonstrated. The results of this work have been reported in [66-68, 71].
4. In Chapter six, we present a hierarchical scheme using both intensity-based and FFT-based methods. We apply the idea of optical flow estimation and augment it with a coarse-to-fine multi-resolution approach, in order to overcome some of the limitations of the intensity-based schemes discussed above. The results of this work have been reported in [68].
5. In Chapter seven, we propose a novel robust approach for registration of images under spatially non-uniform brightness variation. The image registration is formulated as a two-stage hybrid approach combining both a new point-based algorithm and robust estimation using M-estimators in a coarse-to-fine manner. With the point-based algorithm applied at the highest level of decomposition, the initial affine parametric model could be first estimated. Subsequently, the robust estimation using M-estimator is incorporated into the proposed hybrid approach using a multi-resolution scheme. It has two main contributions: First, it identifies

an appropriate image representation that emphasizes the common intensity information, suppresses the non-common information between the two images under brightness variation, and is suitable for coarse-to-fine hierarchical iterative processing. Second, the model-based robust estimation mechanism is incorporated into the proposed framework to reduce its sensitivity to violations of the underlying assumptions. The results of this work have been reported in [69,70,72].

6. In Chapter eight, we describe a hierarchical image registration algorithm for projective motion estimation. The parameters are computed iteratively in a coarse-to-fine hierarchical framework using a variation of the Levenberg-Marquadt nonlinear least squares optimization method. This approach yields a robust solution that precisely registers images with sub-pixel accuracy [72].

## LIST OF REFERENCES

- [1] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, 24 (1992) pp. 325-376.
- [2] B. Zitova, J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, 21 (2003) pp. 977-1000.
- [3] D.L.G. Hill, P.G. Batchelor, M. Holden, D.J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, 46 (2001) R1–R45.
- [4] H. Lester, S.R. Arridge, "A survey of hierarchical non-linear medical image registration", *Pattern Recognition*, 32 (1999) pp. 129–149.
- [5] J.B.A. Maintz, M.A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, 2 (1998) pp. 1–36.
- [6] P.A. van den Elsen, E.-J.D. Pol, M.A. Viergever, "Medical image matching-a review with classification," *IEEE Engineering in Medicine and Biology*, 12 (1993) pp. 26–39.
- [7] M.A. Audette, F.P. Ferrie, T.M. Peters, "An algorithmic overview of surface registration techniques for medical imaging," *Medical image Analysis*, 4 (2000) pp. 201–217.
- [8] L. Ding, A. Goshtasby, M. Satter, "Volume image registration by template matching," *Image and Vision Computing*, 19 (2001) pp. 821–832.
- [9] L.M.G. Fonseca, B.S. Manjunath, "Registration techniques for multisensor remotely

- sensed imagery,” *Photogrammetric Engineering and Remote Sensing*, 62 (1996) pp. 1049–1056.
- [10] E. Gulch, “Results of test on image matching of ISPRS WG,” *ISPRS Journal of Photogrammetry and Remote Sensing*, 46 (1991) pp. 1–18.
- [11] J. le Moigne, “First evaluation of automatic image registration methods,” *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS’98*, Seattle, Washington, 1998, pp. 315–317.
- [12] A. Rosenfeld, A. C. Kak, *Digital picture processing*, vol. I and II, Academic Press, Orlando, FL.
- [13] V. Govindu, C. Shekhar, “Alignment using distributions of local geometric properties,” *IEEE Trans. Pattern Anal. Machine Intell.*, 21 (1999) pp. 1031-1043.
- [14] H. Li, B. S. Manjunath, S. K. Mitra, “A contour-based approach to multisensor image registration,” *IEEE Trans. Image Processing*, 4 (1995) pp. 320-334.
- [15] Q. Zheng, R. Chellappa, “A computational vision approach to image registration,” *IEEE Trans. Image Processing*, 2 (1993) pp. 311-326.
- [16] B. S. Manjunath, R. Chellappa, and C. Malsburg, “A feature based approach to face recognition,” in *Proceedings, IEEE Conf. Comput. Vision Pattern Recognition*, Champaign, Illinois, 1992, pp. 373–378.
- [17] P. Anandan, “A computational framework and an algorithm for the measurement of visual motion,” *International Journal of Computer Vision*, 2 (1989) pp. 283-311.
- [18] J. R. Bergen, P. Anandan, K. Hanna, R. Hingorani, “Hierarchical model-based motion estimation,” *Proc. 2<sup>nd</sup> European Conference on Computer Vision*, 1992, pp.

237-252.

- [19] W. Enkelmann, "Investigations of multi-grid algorithms for estimation of optical flow fields in image sequences," *Computer Vision, Graphics, and Image Processing*, 43 (1988) pp. 150-177.
- [20] J. R. Bergen, E. H. Adelson, "Hierarchical, computationally efficient motion estimation algorithm," *Journal of the Optical Society of America A*, 4 (1987) pp. 4-35.
- [21] S. Negahdaripour, C. Yu, "A generalized brightness change model for computing optical flow," *Proc. Fourth International Conference on Computer Vision*, 1993, pp. 11-14.
- [22] G.D. Hager, P.N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Trans. Pattern Anal. Machine Intell.*, 20 (10) (1998) pp. 1025-1039.
- [23] Lai. S-H, "Robust image matching under partial occlusion and spatially varying illumination change," *Computer Vision and Image Understanding*, 78 (2000) pp. 84-98.
- [24] Y. Altunbasak, R.M. Merserau, A.J. Patti, "A fast parametric motion estimation algorithm with illumination and lens distortion correction," *IEEE Trans. Image Processing*, 12 (4) (2003) pp. 395-408.
- [25] H.W. Haussecker, D.J. Fleet, "Computing optical flow with physical models of brightness variation," *IEEE Trans. Pattern Anal. Machine Intell.*, 23 (6) (2001) pp. 661-673.



- [26] M.D. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 1338-1359, 1997.
- [27] J. Koplowitz and V. Greco, "On the edge localization error for local maximum and zero-crossing edge detectors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 1207-1212, 1994
- [28] D.J. Williams and M. Shah, "Edge contours using multiple scales," *CVGIP*, vol. 51, pp. 256-274, 1990.
- [29] R.W. Cannata, et. al., "Autonomous video registration using model parameter adjustments," *Proc. 29th Applied Imagery Pattern Recognition Workshop*, pp. 215-222, 2000.
- [30] R. Wildes, et. al., "Video georegistration: algorithm and quantitative evaluation," *Proc. IEEE International Conference on Computer Vision*, pp. 343-350, 2001.
- [31] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-8, pp. 147-163, 1986.
- [32] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond.*, vol. B 207, pp.187-217, 1980.
- [33] R.J. Qian and T.S. Huang, "Optimal Edge Detection," *IEEE Image Trans. Image Processing*, vol. 5, pp. 1215-1220, 1996.
- [34] A. Yulle and T.A. Poggio, "Scaling theorems for zero crossings," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 15-25, 1986.
- [35] R. M. Haralick and L. Watson, "A facet model for image data," *Computer Graphics*

- and Image Processing*, vol. 15, pp. 113-129, 1981.
- [36] R.M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 58-68, 1984.
- [37] G. Chen and Y. H. Yang, "Edge detection by regularized cubic B-spline fitting," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 25, pp. 636-643, 1995.
- [38] V.S. Nalwa and T.O. Binford, "On Detecting Edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 699-714, 1986.
- [39] S. Ando, "Consistent Gradient Operators," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 252-265, 2000.
- [40] M. Unser, A. Aldroubi, and M. Eden, "Fast B-spline transforms for continuous image representation and interpolation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 277-285, 1991.
- [41] C.H. Reinsh, "Smoothing by spline functions," *Numer. Math.*, vol. 10, pp. 177-183, 1967.
- [42] I.J. Schoenberg, "Spline functions and the problem of graduation," *Proc. Nat. Acad. Sci.*, vol. 52, pp. 947-950, 1964.
- [43] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: part I - theory," *IEEE Trans. Signal Processing*, vol. 41, pp. 821-833, 1993.
- [44] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: part II - efficient design and applications," *IEEE Trans. Signal Processing*, vol. 41, pp. 834-848, 1993.
- [45] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal.*

- Machine Intell.*, vol. PAMI-8, pp. 679-698, 1986.
- [46] A. Huertas, G. Medioni, "Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 651-664, 1986.
- [47] K. Liang, T. Tjahjadi, and Y-H. Yang, "Bounded diffusion for multiscale edge detection using regularized cubic B-spline fitting," *IEEE Trans. Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 29, pp. 291-297, 1999.
- [48] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Processing*, 5 (1996) 1266-1271.
- [49] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-95 (1987) 700-703.
- [50] S. Ranade and A. Rosenfeld, "Point pattern matching by relaxation," *Pattern Recognition*, 1980, 12, pp. 269-275.
- [51] J. Ton and A.K. Jain, "Registering landsat images by point matching," *IEEE Trans. Geosci. Remote Sensing*, 1993, 27, pp. 642-651.
- [52] A. Goshtasby and G.C. Stockman, "Point pattern matching using convex hull edges," *IEEE Trans. Systems Man Cybernet.*, 1985, SMC-15(5), pp. 631-637.
- [53] W. Wang and Y. Chen, "Image registration by control point pairing using invariant properties of line segments," *Pattern Recognition Letters*, 1997, 18, pp. 269-281.
- [54] Y. Lamdan, J.T. Schwartz, and H.J. Wolfson, "Affine invariant model-based object

- recognition,” *IEEE Trans. Robotics Automation*, 1990, 6, pp. 578-589.
- [55] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge “Comparing images using the Hausdorff distance,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 850-863, 1993.
- [56] P. J. Burt, E.H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Communications*, COM-31(4) (1983) 532-540.
- [57] M. J. Black, P. Anandan, “The robust estimation of multiple motions: parametric and piecewise-smooth flow fields,” *Computer Vision and Image Understanding*, 63 (1996) 75-104.
- [58] J.M. Odobez, P. Bouthemy, “Robust multi-resolution estimation of parametric motion models in complex image sequences,” *Proc. Seventh European Conference on Signal Processing*, 1994.
- [59] G. Li, *Robust regression*, in: D.C. Hoaglin, F. Mosteller, and J.W. Tukey, (Eds.), Exploring data tables, trends and shapes. Chapter 8. pp. 281-343, J. Wiley & Sons, NY, 1986.
- [60] H.S. Sawhney, S. Ayer, M. Gorkani, “Model-based 2D&3D dominant motion estimation for mosaicing and video representation,” *Proc. ICCV*, 1995, pp. 583-590.
- [61] D. G. Luengerber, *Linear and Nonlinear Programming*, Addison-Wesley, MA, 1984.
- [62] G.A.F. Seber, C.J. Wild, *Nonlinear Regression*, Wiley, NY, 1989.
- [63] P.R. Beaudet, “Rotationally invariant image operators,” *Proc. ICPR*, 1978, pp. 579-583.
- [64] D. W. Marquardt, “An algorithm for least-squares estimation of non-linear

- parameters,” *J. Soc. Indust. Appl. Math.*, vol. 11, pp. 431-441, 1963.
- [65] D. Xu and T. Kasparis, “Detection and Localization of Edge Contours,” *Proc. SPIE*, vol. 5097, pp. 79-89, 2003.
- [66] D. Xu, T. Kasparis, “A Hierarchical Approach to Image Registration using Feature Consensus and Hausdorff Distance,” *Proc. SPIE*, vol.5428, pp. 561-568, 2004.
- [67] D. Xu, T. Kasparis, “Video Registration to Geo-referenced Imagery,” *Proc. SPIE*, vol.5428, pp. 569-576, 2004.
- [68] D. Xu, T. Kasparis, “A New Hierarchical Approach to Aerial Image Registration,” *International Journal Pattern Recognition and Artificial Intelligence*, under revision.
- [69] D. Xu, T. Kasparis, “Robust Image Registration under Spatially Non-uniform Brightness Changes,” accepted by *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2005*.
- [70] D. Xu, T. Kasparis, “Robust Image Registration under Spatially Non-uniform Brightness Variation,” submitted to *Pattern Recognition*.
- [71] D. Xu, T. Kasparis, “A New Point-based Algorithm for Image Registration,” submitted to *Pattern Recognition Letter*.
- [72] D. Xu, T. Kasparis, “A Hybrid Approach to Robust Image Registration for Perspective Deformation Recovery,” under preparation for submission to *International Journal of Computer Vision*.