FEATURE-BASED IMAGE REGISTRATION

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology
In
Telematics and Signal Processing

By

SOMARAJU.BODA
Roll No: 207EC108

Department of Electronics & Communication Engineering
National Institute of Technology
Rourkela
2009

FEATURE-BASED IMAGE REGISTRATION

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology
In
Telematics and Signal Processing

By
SOMARAJU.BODA
Roll No: 207EC108

Under the guidance of
Prof.U.C.PATI

Department of Electronics & Communication Engineering
National Institute of Technology
Rourkela
2009

National Institute Of Technology

Rourkela

CERTIFICATE

This is to certify that the thesis entitled "**FEATURE-BASED IMAGE REGISTRATION"** submitted by **Mr. SOMARAJU BODA** in partial fulfillment of the requirements for the award of Master of Technology Degree in **Electronics and Communication Engineering** with specialization in **"Telematics and Signal Processing"** at National Institute of Technology, Rourkela (Deemed University) is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Date:

Prof. U.C Pati
Department of E.C.E
National Institute of Technology
Rourkela

ACKNOWLEDGEMENTS

First of all, I would like to express my deep sense of respect and gratitude towards my advisor and guide **Prof. U.C. Pati**, who has been the guiding force behind this work. I am greatly indebted to him for his constant encouragement, invaluable advice and for propelling me further in every aspect of my academic life. His presence and optimism have provided an invaluable influence on my career and outlook for the future. I consider it my good fortune to have got an opportunity to work with such a wonderful person.

Next, I want to express my respects to Prof.S.K. Patra, Prof.G.S. Rath, Prof.G. Panda, Prof.K.K. Mahapatra, Dr.S. Meher, and Dr.S.K. Behera, for teaching me and also helping me how to learn. They have been great sources of inspiration to me and I thank them from the bottom of my heart.

I would like to thank all faculty members and staff of the Department of Electronics and Communication Engineering, N.I.T. Rourkela for their generous help in various ways for the completion of this thesis.

I would like to thank all my friends and especially my classmates for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious. I've enjoyed their companionship so much during my stay at NIT, Rourkela.

I am especially indebted to my parents for their love, sacrifice, and support. They are my first teachers after I came to this world and have set great examples for me about how to live, study, and work.

Somaraju.Boda
Roll No: 207EC108
Department of ECE
National Institute of Technology
Rourkela

CONTENTS

ABSTRACT

Image registration is the fundamental task used to match two or more partially overlapping images taken, for example, at different times, from different sensors, or from different viewpoints and stitch these images into one panoramic image comprising the whole scene. It is a fundamental image processing technique and is very useful in integrating information from different sensors, finding changes in images taken at different times, inferring three-dimensional information from stereo images, and recognizing model-based objects. Some techniques are proposed to find a geometrical transformation that relates the points of an image to their corresponding points of another image. To register two images, the coordinate transformation between a pair of images must be found. In this thesis, a feature-based method is developed to efficiently estimate an eight-parametric projective transformation model between pairs of images.

The proposed approach applies wavelet transform to extract a number of feature points as the basis for registration. Each selected feature point is an edge point whose edge response is the maximum within a neighborhood. During the real matching process, we check each candidate pair in advance to see if it can possibly become a correct matching pair. Due to this checking, many unnecessary calculations involving cross-correlations can be screened in advance. Therefore, the search time for obtaining correct matching pairs is reduced significantly. Finally, based on the set of correctly matched feature point pairs, the transformation between two partially overlapping images can be decided.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

CPs	Control Points
CT	Computed Tomography
CWT	Continuous Wavelet Transform
DTWT	Discrete Time Wavelet Transform
DWT	Discrete Wavelet Transform
FIR	Finite Impulse Response filters
fMRI	Functional Magnetic Resonance Imaging
JPEG	Joint Photographic Experts Group
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
QMF	Quadrature Mirror Filter
SPECT	Single Photon Emission Computed Tomography
US	Ultrasound
WT	Wavelet Transform

CHAPTER 1

INTRODUCTION

1.1 Introduction to Image Registration

Registration is a fundamental task in image processing used to match two or more pictures taken, for example, at different times, from different sensors, or from different viewpoints. Virtually all large systems which evaluate images require the registration of images, or a closely related operation, as an intermediate step. Specific examples of systems where image registration is a significant component include matching a target with a real-time image of a scene for target recognition, monitoring global land usage using satellite images, matching stereo images to recover shape for autonomous navigation, and aligning images from different medical modalities for diagnosis.

Image registration, sometimes called image alignment, is an important step for a great variety of applications such as remote sensing, medical imaging and multi-sensor fusion based target recognition. It is a prerequisite step prior to image fusion or image mosaic. Its purpose is to overlay two or more images of the same scene taken at different times, from different viewpoints and/or by different sensors. It is a fundamental image processing technique and is very useful in integrating information from different sensors, finding changes in images taken at different times, inferring three-dimensional information from stereo images, and recognizing model-based objects.

Registration can be performed either manually or automatically. The former refers to human operators manually selecting corresponding features in the images to be registered. In order to get reasonably good registration results, an operator has to choose a considerably large number of feature pairs across the whole images, which is not only tedious and wearing but also subject to inconsistency and limited accuracy. Thus, there is a natural need to develop automated techniques that require little or no operator supervision.

Over the years, a broad range of techniques has been developed for various types of data and problems. These techniques have been independently studied for several different applications, resulting in a large body of research. This thesis organizes this material by establishing the relationship between the variations in the images and the type of registration techniques which can most appropriately be applied. Three major types of variations are distinguished. The first type is the variations due to the differences in acquisition which cause

the images to be misaligned. To register images, a spatial transformation is found which will remove these variations. The class of transformations which must be searched to find the optimal transformation is determined by knowledge about the variations of this type. The transformation class in turn influences the general technique that should be taken. The second type of variations is those which are also due to differences in acquisition, but cannot be modeled easily such as lighting and atmospheric conditions. This type usually affects intensity values, but they may also be spatial, such as perspective distortions. The third type of variations is differences in the images that are of interest such as object movements, growths, or other scene changes. Variations of the second and third type are not directly removed by registration, but they make registration more difficult since an exact match is no longer possible. In particular, it is critical that variations of the third type are not removed. Knowledge about the characteristics of each type of variation effect the choice of feature space, similarity measure, search space, and search strategy which will make up the final technique. All registration techniques can be viewed as different combinations of these choices. This framework is useful for understanding the merits and relationships between the wide variety of existing techniques and for assisting in the selection of the most suitable technique for a specific problem.

A frequent problem arises when images taken, at different times, by different sensors or from different viewpoints need to be compared. The images need to be aligned with one another so that differences can be detected. A similar problem occurs when searching for a. prototype or template in another image. To find the optimal match for the template in the image, the proper alignment between the image and template must be found. All of these problems, and many related variations, are solved by methods that perform image registration. A transformation must be found so that the points in one image can be related to their corresponding points in the other. The determination of the optimal transformation for registration depends on the types of variations between the images. The objective of this paper is to provide a framework for solving image registration tasks and to survey the classical approaches.

1.2 Motivation

Image registration shows up in a rich range of application domains, such as medical image analysis (e.g. diagnosis), neuroscience (e.g. brain mapping), computer vision (e.g. stereo image matching for shape recovery), astrophysics (e.g. the alignment of images from different frequencies), military applications (e.g. target recognition), etc. Image registration can serve as a powerful tool to investigate how regional anatomy is altered in disease, with age, gender, handedness, and other clinical or genetic factors. One of the most obvious clinical applications of registration is the area of serial imaging. Comparison of scans from a given patient acquired over various time intervals can be routinely performed to follow disease progression, response to treatment and even dynamic patterns of structure change during organ development. Unfortunately, diagnostic imaging scans are not routinely registered in most radiology department; in contrast, the common practice for an examiner is to do one's best to look at film montages of slices that do not match and try to access disease changes. For gross changes this method of comparison may be adequate. For subtle changes, visual comparison of unmatched images is not enough.

Image registration can also fuse information from multiple imaging devices to correlate different measures of brain structures and function. Integration of functional and anatomical imaging provides complementary information not available from independent analysis of each modality. Registered high-resolution anatomy in magnetic resonance imaging (MRI), ultrasound (US), or computed tomography (CT) images provides a much more precise anatomical basis for the interpretation of functional and pathologic image data, like single photon emission computed tomography (SPECT), positron emission tomography (PET), and functional MRI (fMRI). Without registration, the observed activity would be less accurately mapped to the corresponding anatomical structures. Most importantly, intermodality registration can aid interpretation of clinical imaging for critically important treatment decisions.

Registration algorithms can encode patterns of anatomic variability in large human populations and can use the information to create a disease-specific population-based brain atlas. This is a standard brain which is constructed from a number of brains and given in a fixed coordinate system to contain information about physical properties of neuroanatomies.

When we have mapped an object and an atlas to each other, we can remove individual anatomical variation and pass segmentation or labeling information from the atlas to the subject image.

Many surgical procedures require highly precise localization, often of deeply buried structures, in order for the surgeons to extract targeted tissue with minimal damage to nearby structures. The plan is constructed in the coordinate system relative to preoperative image data, while the surgical procedure is performed in the coordinate system relative to the patient. Although methods such as MRI and CT are invaluable in imaging and displaying the internal structure of the body, the surgeons still face a key problem in applying that information to the actual procedure. With the help of registration, the relationship between preoperative data and patient coordinate system can be established, and thus provides the surgeon information about position of his instruments relative to the planned trajectory, and allows the surgeon to directly visualize important structures, and act accordingly.

1.3 Literature Review

Li *et al.* [1] proposed a contour-based approach to register images 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. On the other hand, the area-based method usually adopts a window of points to determine a matched location using the correlation technique [2, 3]. The most commonly used measure is normalized cross-correlation. This method is more robust than the feature-based method in some situations. However, if the orientation difference between the two images is large, the value of cross-correlation will be greatly influenced and the correspondences between feature points, thus, hard to derive. Therefore, De Castro and Morandi [4] proposed an elegant method, called phase correlation, to overcome this problem. However, when the overlapping area between images is small, their method becomes unreliable. In order to solve the problem, it is necessary to develop a method to estimate the rotation parameter in advance. In [3], Zheng and Chellappa proposed a 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.* [5], a number of feature points are extracted from the image pair. Then, these 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 works well for most cases. However, if a scene includes many buildings and objects, the method will fail due to the fact that the illumination conditions in one image may not be equivalent to those in the other. In general, the estimation of a rotation angle in their approach is rough. Further, their approach requires aGabor function decomposition in the feature extraction process. This decomposition is computationally intensive. Another drawback of their approach is that when false matches emerge, their method cannot handle them.

1.4 Thesis Organization

This thesis contains 5 chapters. Following the introduction, the rest of the thesis is organized as follows:

Chapter 2 gives an overview of image registration. This includes image registration, classification of image registration and applications of image registration.

Chapter 3 gives proposed feature-based image registration algorithm. This includes proposed registration algorithm, wavelet decomposition, feature points extraction using wavelets, finding matching pairs, deriving the transformation parameters and refine matching.

Chapter 4 gives experimental results of proposed registration algorithm.

Chapter 5 gives conclusion and future works.

CHAPTER 2

AN OVERVIEW OF IMAGE REGISTRATION

2.1 Image Registration

Image registration is used to match two or more partially overlapping images and stitch them into one panoramic image of the scene. To register two images, the coordinate transformation between a pair of images must be found from class of transformations. The optimal transformation depends on the types of relation between the overlapping images.

To find the relationship between two images we rely on the estimation of the parameters of the transformation model. The number of parameters depends on the chosen transformation model. A common assumption is that the coordinate transformations between two images are rigid planar models. Rigid planar transformation is composed of scaling, rotation, and translation changes, which map the pixel (x_1, y_1) of image f_1 to the pixel (x_2, y_2) of another image f_2:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = sR\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + T = s\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

The rigid transformation is sufficient to match two images of a scene taken from the same viewing angle but from different position. That is, the camera can rotate about its optical axis. In the case of remote sensing, where the distance approaches infinity, the transformation between the captured images behaves like a planar rigid transformation.

2.2 Classification of Image Registration

The classification of image registration algorithms [6] are described in this section.

2.2.1. Area-based vs Feature-based

Image registration algorithms fall within two realms of classification: area based methods and feature based methods [7]. The original image is often referred to as the reference image and the image to be mapped onto the reference image is referred to as the target image. For area based image registration methods, the algorithm looks at the structure

of the image via correlation metrics, Fourier properties and other means of structural analysis. Alternatively, most feature based methods, instead of looking at the overall structure of images, fine tune their mappings to the correlation of image features: lines, curves, points, line intersections, boundaries, etc.

2.2.2. Transformation model

Image registration algorithms can also be classified according to the transformation model used to relate the reference image space with the target image space. The first broad category of transformation models includes linear transformations, which are a combination of translation, rotation, global scaling, shear and perspective components. Linear transformations are global in nature, thus not being able to model local deformations. Usually, perspective components are not needed for registration, so that in this case the linear transformation is an affine one.

The second category includes 'elastic' or 'nonrigid' transformations. These transformations allow local warping of image features, thus providing support for local deformations. Nonrigid transformation approaches include polynomial wrapping, interpolation of smooth basis functions (thin-plate splines and wavelets), and physical continuum models.

2.2.3. Search-based vs direct methods

Image registration methods can also be classified in terms of the type of search that is needed to compute the transformation between the two image domains. In search-based methods the effect of different image deformations is evaluated and compared. In direct methods, such as the Lucas Kanade method and phase-based methods, an estimate of the image deformation is computed from local image statistics and is then used for updating the estimated image deformation between the two domains.

2.2.4. Spatial-domain methods

Many image registration methods operate in the spatial domain, using features, structures, and textures as matching criteria. In the spatial domain, images look 'normal' as the human eye might perceive them. Some of the feature matching algorithms are outgrowths of traditional techniques for performing manual image registration, in which operators choose matching sets of control points (CPs) between images. When the number of control points exceeds the minimum required to define the appropriate transformation model, iterative algorithms like RANSAC are used to robustly estimate the best solution.

2.2.5. Frequency-domain methods

Other algorithms use the properties of the frequency-domain to directly determine shifts between two images. Applying the phase correlation method to a pair of overlapping images produces a third image which contains a single peak. The location of this peak corresponds to the relative translation between the two images. Unlike many spatial-domain algorithms, the phase correlation method is resilient to noise, occlusions, and other defects typical of medical or satellite images. Additionally, the phase correlation uses the fast fourier transform to compute the cross-correlation between the two images, generally resulting in large performance gains. The method can be extended to determine affine rotation and scaling between two images by first converting the images to log-polar coordinates. Due to properties of the fourier transform, the rotation and scaling parameters can be determined in a manner invariant to translation. This single feature makes phase-correlation methods highly attractive vs. typical spatial methods, which must determine rotation, scaling, and translation simultaneously, often at the cost of reduced precision in all three.

2.2.6. Single-modality vs Multi-modality

Another useful classification is between single-modality and multi-modality registration algorithms. Single-modality registration algorithms are those intended to register images of the same modality (i.e. acquired using the same kind of imaging device), while multi-modality registration algorithms are those intended to register images acquired using different imaging devices. There are several examples of multi-modality registration

algorithms in the medical imaging field. Examples include registration of brain CT/MRI images or whole body PET/CT images for tumor localization, registration of contrast-enhanced CT images against non-contrast-enhanced CT images for segmentation of specific parts of the anatomy and registration of ultrasound and CT images for prostate localization in radiotherapy.

2.2.7. Image similarity-based methods

Image similarity-based methods are broadly used in medical imaging. A basic image similarity-based method consists of a transformation model, which is applied to reference image coordinates to locate their corresponding coordinates in the target image space, an image similarity metric, which quantifies the degree of correspondence between features in both image spaces achieved by a given transformation, and an optimization algorithm, which tries to maximize image similarity by changing the transformation parameters.

The choice of an image similarity measure depends on the nature of the images to be registered. Common examples of image similarity measures include cross-correlation, mutual information, sum of square differences and ratio image uniformity. Mutual information and its variant, normalized mutual information are the most popular image similarity measures for registration of multimodality images. Cross-correlation, sum of square differences and ratio image uniformity are commonly used for registration of images of the same modality.

2.3 Feature Space

Many techniques that have been proposed to solve the registration problem in many different forms can be broadly classified into three types, namely feature-based matching, intensity-based matching and hybrid approaches. Each of these approaches has its advantages and disadvantages.

2.3.1 Feature-based registration

Feature-based approaches attempt to find the correspondence and transformation using distinct anatomical features that are extracted from images. These features include

points [8]; [9]; [10], curves [11]; [12]; [13], or a surface model [14]; [15]; [16] of anatomical structures. Feature-based methods are typically applied when the local structure information is more significant than the information carried by the image intensity. They can handle complex between-image distortions and can be faster, since they don't evaluate a matching criterion on every single voxel in the image, but rather rely on a relatively small number of features. The simplest set of anatomical features is a set of landmarks. However, the selection of landmarks is recognized to be a difficult problem, whether done automatically or manually. For many images, this is a serious drawback because registration accuracy can be no better than what is achieved by the initial selection of landmarks. For practical reasons, the number and precision of landmark locations is usually limited. Hence, spatial coordinates and geometric primitives often oversimplify the data by being too sparse and imprecise.

2.3.2 Intensity-based registration

The intensity-based registration methods operate directly on the image gray values, without reducing the gray-level image to relatively sparse extracted information. The basic principle of intensity-based techniques is to search, in a certain space of transformations, the one that maximizes (or minimizes) a criterion measuring the intensity similarity of corresponding voxels. Some measures of similarity are sum of squared differences in pixel intensities ([18]), regional correction [18], or mutual information ([19]). Mutual information has proved to be an excellent similarity measure for cross-modality registrations, since it assumes only that the statistical dependence of the voxel intensities is maximal when the images are geometrically aligned. The intensity similarity measure, combined with a measure of the structural integrity of the deforming scan, is optimized by adjusting parameters of the deformation field. Such an approach is typically more computationally demanding, but avoids the difficulties of a feature extraction stage.

2.3.3 Hybrid registration

Further hybrid approaches, based on a combination of feature-based and intensity-based criteria, are likely to benefit from both the advantages of each strategy. Christensen et al. [20] introduced a hierarchical approach to image registration combining a landmark-based scheme with a intensity-based approach using a fluid model. The approach has been applied

to the registration of 3D cryosection data of a macaque monkey brain as well to MR images of the human brain.

2.4 Warp Space

One of the important factors to categorize registration techniques is the warp space used, because it contains warping functions. Warping functions are candidate solutions of the registration problem. These exhibits the wide range of methods developed over the last twenty years and includes the basis-function expansions, physical models, and optical flow methods.

2.4.1 Basis function expansions

Nonrigid registration produces a deformation field, giving a mapping for each pixel (voxel) in the deforming image. The deformation may be expressed in terms of a truncated set of basis functions. Specific examples include sinusoidal or wavelet basis [21] and radial basis functions such as thin-plate splines [22]. Compactly supported basis function such as B-splines [23] or other radially symmetric compactly supported basis functions have also been used.

2.4.2 Physical models

Nonrigid image registration can also be performed by modeling the deformation of the source image into the target image as an elastic physical process. A wide range of these methods [24] are based on Navier-Stokes linear elastic partial differential equation:

$$\mu\Delta u(x) + (\mu + \lambda)\Delta(\Delta.u(x)) = k(x, u(x)),$$

Where μ and λ , is the Lame constant, which refer to the elastic properties of the medium. u = (u1, u2, u3) is the displacement field, and k is the force acting on the elastic body. The term $\Delta = \Delta^T\Delta$ is the Laplacian operator and $\Delta(\Delta.u)$ is gradient of the divergence of u. Generally λ is set to zero to ensure that deformations are only effect in the directions of applied forces; μ takes a value between 0 and 1. The external forces can be derived from the

local optimization of the similarity function which is defined by the intensity values or by the correspondence of boundary structures.

Navier-Stokes equation has a disadvantage of disallowing extensive localized deformations due to the modeling of stresses that increase proportionally to deformation strength. This can be handled by viscous fluid model [25]. Instead of using the displacement directly in Navier-Stokes equation, the deformation velocity v is used.

$$\mu\Delta v(x) + (\mu + \lambda)\Delta(\Delta . v(x)) = k(x, u(x)).$$

The advantage of using the velocity is that we can allow large deformations which cannot be performed by the Navier-Stokes equation alone. A resulting problem is that the likelihood of misregistrations is increased.

2.4.3 Optical flow methods

The optical flow constraint equation was derived to estimate the motion between two successive frames in an image sequence. It is based on the assumption that the intensity value of a given point in the image does not change over small time increments. The constraint can be expressed as:

$$(\Delta I)^T . \vec{v} + I_t = 0$$

in which ΔI, $\vec{v}$ and I_t are the intensity gradient of the image, the unknown motion vector between the images, and the temporal derivative of the image, respectively. This equation is underconstrained (a problem known as the aperture problem) and a number of regularization scheme have been proposed to address this problem [26].

2.5. Applications of Image Registration

Image registration is widely used in remote sensing, medical imaging, computer vision, video processing, and many others. In general, its applications can be divided into three main groups according to the manner of the image acquisition:

1. Different times: Images of the same scene are acquired at different times. The aim is to find and evaluate changes in the scene. Examples of applications are landscape planning in remote sensing, automatic change detection in video surveillance, motion tracking in computer vision, monitoring of healing therapy and tumor evolution in medical imaging, and motion estimation and superresolution reconstruction in video processing.

2. Different viewpoints: Images of the same scene are acquired from different viewpoints. The aim is to gain a larger 2-D view or a 3-D representation of the scene being imaged. Examples of applications are image mosaicing of the surveyed area in remote sensing, shape recovery and structure from motion in computer vision, and sprite generation and coding in video compression.

3. Different sensors: Image of the same scene are acquired from different sensors. The aim is to integrate the information obtained from different sensors to gain more complex and detailed representation of the scene. Examples of applications are multisensor image fusion in remote sensing and medical imaging, monitoring activities in multisensor surveillance, and image fusion in vehicular navigation.

CHAPTER 3

A REGISTRATION ALGORITHM

The overall flow chart of the proposed algorithm is illustrated in Figure 3.1.

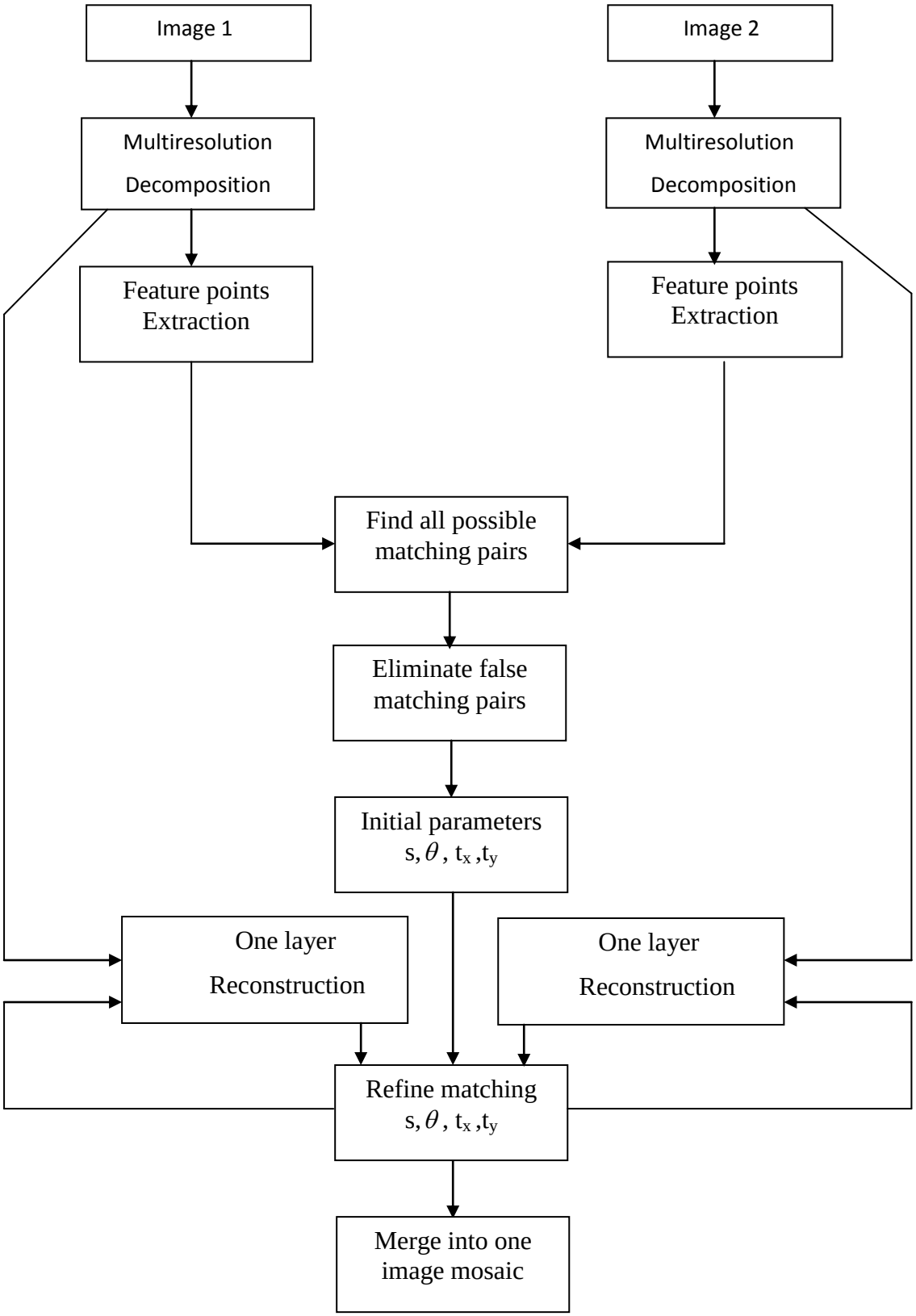
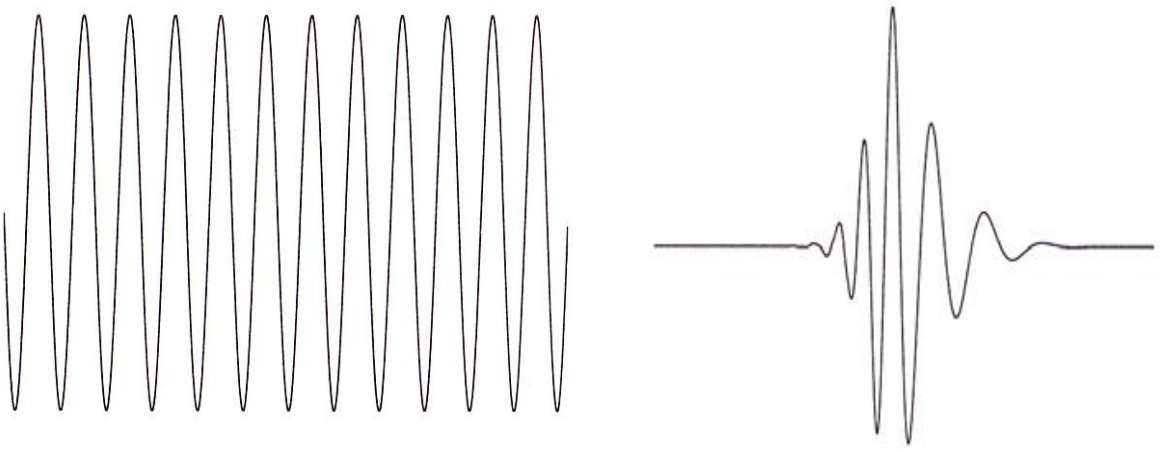


Figure 3.1 Flow chart of feature-based image registration

3.1 Multiresolution Decomposition

Initially, input images are decomposed into their multiresolution representation by a discrete wavelet transform.

Mathematically a “wave” is expressed as a sinusoidal (or oscillating) function of time or space. Fourier analysis expands an arbitrary signal in terms of infinite number of sinusoidal functions of its harmonics. Fourier representation of signals is known to be very effective in analysis of time-invariant (stationary) periodic signals. In contrast to a sinusoidal function, a wavelet is a small wave whose energy is concentrated in time. Properties of wavelets allow both time and frequency analysis of signals simultaneously because of the fact that the energy of wavelets is concentrated in time and still possesses the wave-like (periodic) characteristics. Wavelet representation thus provides a versatile mathematical tool to analyse transient, time-variant (nonstationary) signals that may not be statistically predictable especially at the region of discontinuities - a special feature that is typical of images having discontinuities at the edges.



(a) (b)

Figure 3.2 Demonstrations of (a) a Wave and (b) a Wavelet

3.1.1 Wavelet Transforms

Wavelets are functions generated from one single function (basis function) called the prototype or mother wavelet by dilations (scalings) and translations (shifts) in time (frequency) domain. If the mother wavelet is denoted by $\psi(t)$, the other wavelets $\psi_{a,b}(t)$ can be represented as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) - - - - - - - - (3.1)$$

Where a and b are two arbitrary real numbers. The variables a and b represent the parameters for dilations and translations respectively in the time axis. From Eq. 3.1, it is obvious that the mother wavelet can be essentially represented as

$$\psi(t) = \psi_{1,0}(t) - - - - - - - - - - - - - (3.2)$$

For any arbitrary $a \neq 1$ and $b = 0$, it is possible to derive that

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) - - - - - - - - (3.3)$$

As shown in Eq. 3.3, $\psi_{a,0}(t)$ is nothing but a time-scaled (by a) and amplitude-scaled (by $\sqrt{|a|}$) version of the mother wavelet function $\psi(t)$ in Eq. 3.2. The parameter a causes contraction of $\psi(t)$ in the time axis when a < 1 and expression or stretching when a > 1. That's why the parameter a is called the dilation (scaling) parameter. For a < 0, the function $\psi_{a,b}(t)$ results in time reversal with dilation. Mathematically, substituting t in Eq. 3.3 by t-b to cause a translation or shift in the time axis resulting in the wavelet function $\psi_{a,b}(t)$ as shown in Eq. 3.1. The function $\psi_{a,b}(t)$ is a shift of $\psi_{a,0}(t)$ in right along the time axis by an amount b when b > 0 whereas it is a shift in left along the time axis by an amount b when b < 0. That's why the variable b represents the translation in time (shift in frequency) domain.

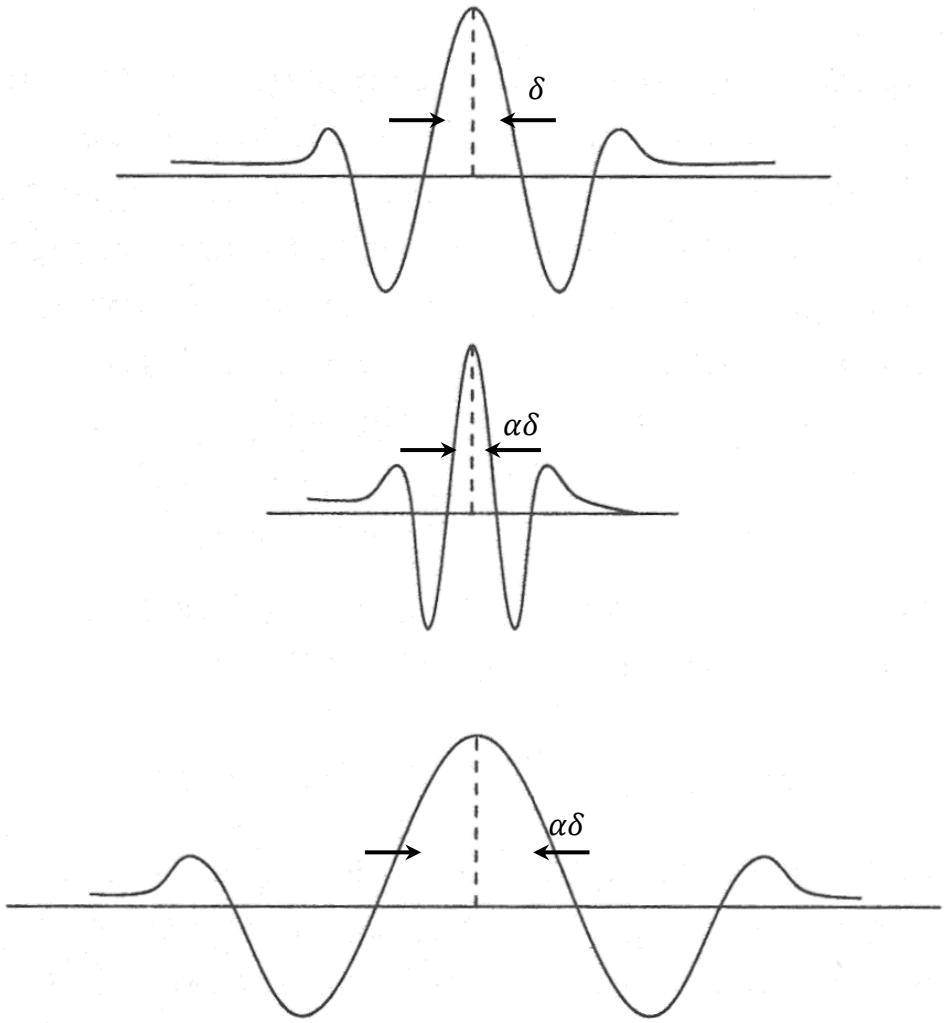


Figure 3.3 (a) A mother wavelet, (b) $\boldsymbol{\psi(t/\alpha)}$: $\mathbf{0} < \boldsymbol{\alpha} < 1$, (c) $\boldsymbol{\psi(t/\alpha)}$: $\boldsymbol{\alpha} > \mathbf{1}$.

Figure 3.3 shows an illustration of a mother wavelet and its dilations in the time domain with the dilation parameter a = α. For the mother wavelet ψ(t) shown in Figure 3.3(a), a contraction of the signal in the time axis when α < 1 is shown in Figure 3.3(b) and expansion of the signal in the time axis when α > 1 is shown in Figure 3.3(c). Based on this definition of wavelets, the wavelet transform (WT) of a function (signal) *f (t)* is mathematically represented by

$$W(a,b) = \int_{-\infty}^{+\infty} \psi_{a,b}(\mathrm{t})\, f(\mathrm{t})\mathrm{dt} \quad ----------- (3.4)$$

The inverse transform to reconstruct f (t) from W (a, b) is mathematically represented by

$$f(\mathrm{t}) = \frac{1}{C}\int_{a=-\infty}^{+\infty}\int_{b=-\infty}^{+\infty}\frac{1}{|a|^2}\mathrm{W}(a,b)\psi_{a,b}(\mathrm{t})dadb \text{---} (3.5)$$

Where

$$C = \int_{-\infty}^{\infty}\frac{|\Psi(\omega)|^2}{|\omega|}d\omega \text{----------} (3.6)$$

and $\Psi(\omega)$ is the Fourier transform of the mother wavelet $\psi(\mathrm{t})$.

If a and b are two continuous (nondiscrete) variables and *f (t)* is also a continuous function, W (a, b) is called the continuous wavelet transform (CWT). Hence the CWT maps a one-dimensional function *f (t)* to a function W(a, b) of two continuous real variables a (dilation) and b (translation).

3.1.2 Discrete Wavelet Transforms

Since the input signal (e.g., a digital image) is processed by a digital computing machine, it is prudent to define the discrete version of the wavelet transform. To define the wavelet in terms of discrete values of the dilation and translation parameters a and b instead of being continuous, make a and b discrete using Eq. 3.6,

$$a = a_0^m, \qquad b = nb_0a_0^m$$

where *m* and *n* are integers. Substituted *a* and *b* in Eq. 3.1 by Eq. 3.6, the discrete wavelets can be represented by Eq. 3.7.

$$\psi_{m,n}(\mathrm{t}) = a_0^{-\mathrm{m}/2}\psi(a_0^{-m}t - nb_0) \text{----------} (3.7)$$

There are many choices to select the values of a_0 and b_0. By selecting $a_0 = 2$ and $b_0 = 1$, $a = 2^m$ and $b = n2^m$. This corresponds to sampling (discretization) of a and b in such a way that the consecutive dicrete values of a and b as well as the sampling intervals differ by a factor of two. This way of sampling is popularly known as dyadic decomposition. Using these values, it is possible to represent the discrete wavelets as in eq. 3.8, which constitutes a family of orthonormal basis functions.

$$\psi_{m,n}(t) = 2^{-\frac{m}{2}}\psi(a_0^{-m}t - n) \quad ----------(3.8)$$

In general, the wavelet coefficients for function $f(t)$ are given by

$$C_{m,n}(f) = a_0^{-m/2}\int f(t)\ \psi(a_0^{-m}t - nb_0)dt \quad ----------(3.9)$$

and hence for dyadic decomposition, the wavelet coefficients can be derived accordingly as

$$C_{m,n}(f) = 2^{-m/2}\int f(t)\ \psi(a_0^{-m}t - nb_0)dt \quad ----------(3.10)$$

This allows us to reconstruct the signal $f(t)$ in form the discrete wavelet coefficients as

$$f(t) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} c_{m,n}(f)\psi_{m,n}(t) \quad ----------(3.11)$$

The transform shown in Eq. 3.9 is called the wavelet series, which is analogous to the Fourier series because the input function *f(t)* is still a continuous function whereas the transform coefficients are discrete. This is often called the discrete time wavelet transform (DTWT). For digital signal or image processing applications executed by a digital computer, the input signal *f(t)* needs to be discrete in nature because of the digital sampling of the original data, which is represented by a finite number bits. When the input function *f(t)* as well as the wavelet parameters a and b are represented in discrete form, the transformation is commonly referred to as the discrete wavelet transform (DWT) of the signal f(t).

The discrete wavelet transform (DWT) became a very versatile signal processing tool after Mallat [27] proposed the multiresolution representation of signals based on wavelet decomposition. The method of multiresolution is to represent a function (signal) with a collection of coefficients, each of which provides information about the position as well as the frequency of the signal (function). The advantage of DWT over Fourier transformation is that it performs multiresolution analysis of signals with localization. As a result, the DWT decomposes a digital signal into different subbands so that the lower frequency subbands will have finer frequency resolution and coarser time resolution compared to the higher frequency subbands. The DWT is being increasingly used for image compression due to the fact that the DWT supports features like progressive image transmission (by quality, by resolution), ease of compressed image manipulation, region of interest coding, etc. Because of these characteristics, the DWT is the basis of the new JPEG2000 image compression standard.

3.1.3 Concept of Multiresolution Analysis

There are a number of orthogonal wavelet basis-functions of the form $\psi_{m,n}(t) = 2^{-m/2}\psi(a_0^{-m}t - n)$. The theory of multiresolution analysis presented a systematic approach to generate the wavelets. The idea of multiresolution analysis is to approximate a function f*(t)* at different levels of resolution.

In multiresolution analysis, two functions are considered: the mother wavelet $\psi(t)$ and the scaling function $\phi(t)$. The dilated (scaled) and translated (shifted) version of the scaling function is given by $\phi_{m,n}(t) = 2^{-m/2}\phi(a_0^{-m}t - n)$. For fixed m, the set of scaling functions $\phi_{m,n}(t)$ are orthonormal. By the linear combinations of the scaling function and its translations, a set of functions can be generated

$$f(t) = \sum_n \alpha_n \ \phi_{m,n}(t) - - - - - - - - - - - - - (3.12)$$

The set of all such functions generated by linear combination of the set $\{\phi_{m,n}(t)\}$ is called the span of the set $\{\phi_{m,n}(t)\}$, denoted by Span $\{\phi_{m,n}(t)\}$. Now consider Vm to be a vector space corresponding to Span $\{\phi_{m,n}(t)\}$. Assuming that the resolution increases with decreasing m, these vector spaces describe successive approximation vector spaces, ...

$\subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset$ ..., each with resolution 2m (i.e., each space Vj+1 is contained in the next resolution space Vj). In multiresolution analysis, the set of subspaces satisfies the following properties:

1. $V_{m+1} \subset V_{m1}$, for all m: This property state that each subspace is contained in the next resolution subspace.

2. $\overline{\cup V_m} = \mathcal{L}^2(\mathcal{R})$: This property indicates that the union of subspaces is dense in the space of square integrable functions $\mathcal{L}^2(\mathcal{R})$; $\mathcal{R}$ indicates a set of real numbers (upward completeness property).

3. $\cap V_m = 0$ (an empty set): This property is called downward completeness property.

4. $f(t) \in V_0 \leftrightarrow f(2^{-m}t) \in V_m$: Dilating a function from resolution space V_0 by a factor of 2^m results in the lower resolution space V_m (scale or dilation invariance property).

5. $f(t) \in V_0 \leftrightarrow f(t-n) \in V_0$: Combining this with the scale invariance property above, this property states that translating a function in a resolution space does not change the resolution (translation invariance property).

6. There exists a set $\{ \phi(t-n) \in V_0$: n is an integer$\}$ that forms an orthogonal basis of V_0.

The basic tenet of multiresolution analysis is that whenever the above properties are satisfied, there exists an orthonormal wavelet basis $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)$ such that

$$P_{m-1}(f) = P_m(f) + \sum c_{m,n}(f)\psi_{m,n}(t) - - - - - - - - (3.13)$$

Where P_j is the orthonormal projection of ψ onto V_j. For each m, consider the wavelet functions $\psi_{m,n}(t)$ span a vector space Wm. It is clear from Eq. 2.13 that the wavelet that generates the space Wm and the scaling function that generates the space Vm are not independent. Wm is exactly the orthogonal complement of Vm in Vm-1. Thus, any function in Vm-1 can be expressed as the sum of a function in Vm and a function in the wavelet space Wm. Symbolically, it is possible to express this as

$$V_{m-1} = V_m \oplus W_m \text{ - - - - - - - -- - - - - - - -- - - - - - - -- - - - - - -} \quad (3.14)$$

Since, *m* is arbitrary,

$$V_m = V_{m+1} \oplus W_{m+1} \text{ -} \quad (3.15)$$

Thus,

$$V_{m-1} = V_{m+1} \oplus W_{m+1} \oplus W_m \text{ - - - -- - - - - - - - -} \quad (3.16)$$

Continuing in this fashion, it is possible to establish that

$$V_{m-1} = V_k \oplus W_k \oplus W_{k-1} \oplus W_{k-2} \ldots \oplus W_m \text{ - - - - - - - -} \quad (3.17)$$

for any $k \geq m$.

Thus, a function belonging to the space Vm-1 (i.e., the function can be exactly represented by the scaling function at resolution m – 1), can be decomposed to a sum of functions starting with lower-resolution approximation followed by a sequence of functions generated by dilations of wavelet that represent the loss of information in terms of details. The successive levels of approximations can be considered as the representation of an image with fewer and fewer pixels. The wavelet coefficients can then be considered as the additional detail information needed to go from a coarser to a finer approximation. Hence, in each level of decomposition the signal can be decomposed into two parts, one is the coarse approximation of the signal in the lower resolution and the other is the detail information that was lost because of approximation. The wavelet coefficients derived by Eq. 3.9 or 3.10, therefore, describe the information (detail) lost when going from an approximation of the signal at resolution 2m-1 to the coarser approximation at resolution 2m.

3.1.4 Implementation by Filters and the Pyramid Algorithm

Multiresolution analysis decomposes signal into two parts – one approximation of the original signal from finer to coarser resolution and the other detail information that was lost sue to the approximation. This can be represented as

$$f(\mathrm{t}) = \sum_{\mathrm{n}} a_{m+1,n}\, \phi_{m+1,n} + \sum_{\mathrm{n}} c_{m+1,n}\, \psi_{m+1,n} \text{ - - - - - - - -} \quad (3.18)$$

Where *f (t)* denotes the value of input function *f (t)* at resolution 2m, cm+1, n is the detail information, and am+1,n is the coarser approximation of the signal at resolution 2m+1. The functions, $\phi_{m+1,n}$ and $\psi_{m+1,n}$ are the dilation and wavelet basis functions (orthonormal).

In 1989, Mallat [27] proposed the multiresolution approach for wavelet decomposition of signals using a pyramidal filter structure of quadrature mirror filter (QMF) pairs. Wavelets developed by Daubechies [28, 29], in terms of discrete-time perfect reconstruction filter banks, correspond to IFR filters. In multiresolution analysis, it can be proven that decomposition of signals using the discrete wavelet transform can be expressed in terms of FIR filters and the algorithm for computation of the wavelet coefficients for the signal *f (t)* can be represented as

$$\left.\begin{aligned} c_{m,n}(f) &= \textstyle\sum_k g_{2n-k}\, a_{m-1,k}(f) \\ a_{m,n}(f) &= \textstyle\sum_k h_{2n-k}\, a_{m-1,k}(f) \end{aligned}\right\} \qquad \text{- - - - - - (3.19)}$$

Where g and h are the high-pass and low-pass filters, $g_i = (-1)^i h_{-i+1}$ and $h_i = 2^{1/2} \int \phi(x - i)\, \phi(2x)dx$. Actually, $a_{m,n}(f)$ are the coefficients characteristics characterizing the projection of the function f(t) in the vector subspace Vm (i.e. approximation of the function in resolution 2m), whereas $c_{m,n}(f) \in W_m$ are the wavelet coefficients (detail information) at resolution 2m. If the input signal f(t) is in discrete sampled form, then it is possible to consider these samples as the highest order resolution approximation coefficients $a_{0,n}(f) \in V_0$ and Eq. 3.19 describes the multiresolution subband decomposition algorithm to construct $a_{m,n}(f)$ and $c_{m,n}(f)$ at level m with a low-pass filter h and high-pass filter g from $c_{m-1,n}(f)$, which were generated at level m-1. These filters are called the analysis filters. The recursive algorithm to compute DWT in different levels using Eq 3.19 is popularly called Mallat's Pyramid Algorithm. Since the synthesis filters h and g have been derived from the orthonormal basis functions ϕ and ψ, these filters give exact reconstruction

$$a_{m-1,i}(f) = \sum_n h_{2n-i}\, a_{m,n}(f) + \sum_n g_{2n-i}\, c_{m,n}(f) \quad \text{- - - - - - - - - (3.20)}$$

Most of the orthogonal wavelet basis functions have infinitely supported ψ and accordingly the filters h and g could be with infinitely many taps. However, for practical and computationally efficient implementation of the DWT for image processing applications, it is desirable to have finite impulse response filters (FIR) with a small number of taps. It is possible to construct such filters by relaxing the orthonormality requirements and using biorthogonal basis functions. It should be noted that the wavelet filters are orthogonal when (h', g') = (h, g), otherwise it is biorthogonal. In such a case the filters (h' and g', called the synthesis filters) for reconstruction of the signal can be different than the analysis filters (h and g) for decomposition of the signals. In order to achieve exact reconstruction, construct the filters such that it satisfies the relationship of the synthesis filter with the analysis filter as shown in Eq. 3.21:

$$\left.\begin{aligned} g'_n &= (-1)^n h_{-n+1} \\ g_n &= (-1)^n h'_{-n+1} \\ \textstyle\sum_n h_n\, h'_{n+2k} &= \delta_{k,0} \end{aligned}\right\} \quad \text{- - - - - (3.21)}$$

If (h', g') = (h, g), the wavelet filters are called orthogonal, otherwise they are called biorthogonal. The popular (9, 7) wavelet filter adopted in JPEG2000 is one example of such a biorthogonal filter. The signal is still decomposed using Eq. 3.19, but the reconstruction equation is now done using the synthesis filters h' and g' as shown in Eq. 3.22:

$$a_{m-1,i}(f) = \sum_n a_{m,n}(f)\, h'_{2n-i} + \sum_n c_{m,n}(f)\, g_{2n-i} \quad \text{- - - - - - - - - - - - - (3.22)}$$

Let's summarize the DWT computation here in terms of simple digital FIR filtering. Given the input discrete signal x(n) (shown as a(0,n) in Figure 3.4), it is filtered parallelly by a low-pass (h) and a high-pass (g) at each transform level. The two output streams are then subsampled by simply dropping the alternate output samples in each stream to produce the low-pass subband yL. The above arithmetic computation can be expressed as follows:

$$y_L(n) = \textstyle\sum_{i=0}^{\tau_L - 1} h(i)\ x(2n-i)\,, \qquad y_H(n) = \sum_{i=0}^{\tau_H - 1} g(i)\ x(2n-i) \text{ – (3.23)}$$

Where τ_L and τ_H are the lengths of the low-pass (h) and high-pass (g) filters respectively. Since the low-pass subband a(1, n) is an approximation of the input, applying the above computation again on a(1,n) to produce the subbands a(2,n) and c(2,n) and so on. During the inverse transform to reconstruct the signal, both a(3,n) and c(3,n) are first upsampled by inserting zeros between two samples, and then they are filtered by low-pass (h') and high-pass (g') filters respectively. These two filtered output streams are added together to reconstruct a(2,n). The same continues until the reconstruction of the original signal a(0,n).

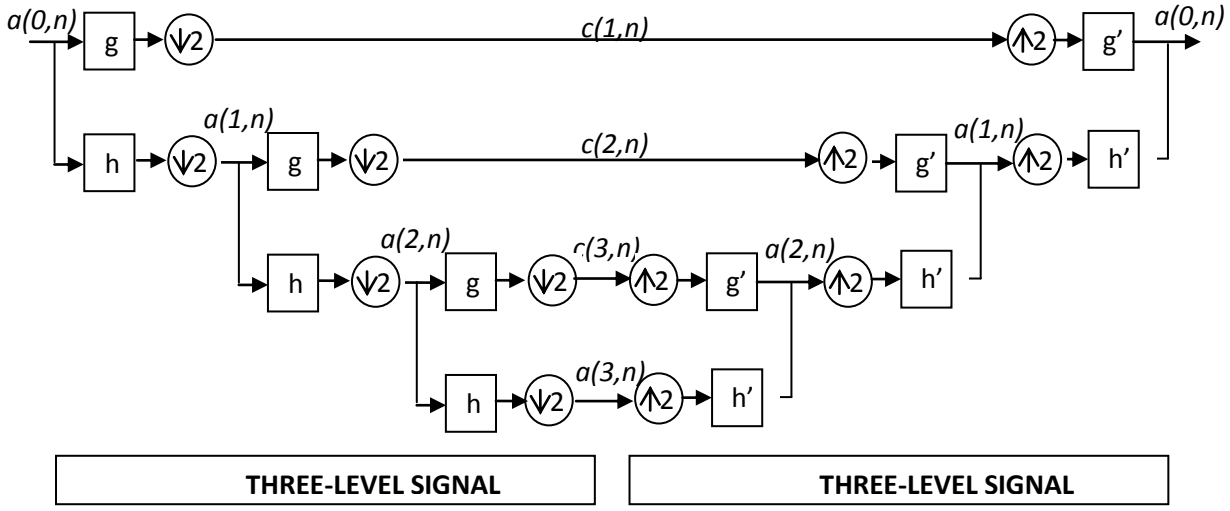


Figure 3.4 Three-level multiresolution wavelet decomposition and reconstruction of signals using pyramidal filter structure

3.1.5 Extension to Two-Dimensional Signals

The two-dimensional extension of DWT is essential for transformation of two-dimensional signals, such as a digital image. A two-dimensional digital signal can be represented by a two-dimensional array X[M, N] with M rows and N columns, where M and N are nonnegative integers. The simple approach for two-dimensional implementation of the DWT is to perform the one-dimensional DWT row-wise to produce an intermediate result and then perform the same one-dimensional DWT column-wise on this intermediate result to produce the final result. This is shown in Figure 3.5(a). This is possible because the two-dimensional scaling functions can be expressed as separable functions which is the product of

two-dimensional scaling function such as $\emptyset_2(x,y) = \emptyset_1(x)\emptyset_1(y)$. The same is true for the wavelet function $\psi(x,y)$ as well. Applying the one-dimensional transform in each row, two subbands are produced in each row. When the low-frequency subbands of all the rows (L) are put together, it looks like a thin version (of size M x $\frac{N}{2}$) of the input signal as shown in Figure 3.5(a). Similarly put together the high-frequency subbands of all the rows to produce the H subband of size M x $\frac{N}{2}$, which contains mainly the high-frequency information around discontinuities (edges in an image) in the input signal. Then applying a one-dimensional DWT column-wise on these L and H subbands (intermediate result), four subbands LL, LH, HL, and HH of size $\frac{M}{2} \times \frac{N}{2}$ are generated as shown in Figure 3.5(a). LL is a coarser version of the original input signal. LH, HL, and HH are the high frequency subband containing the detail information. It is also possible to apply one-dimensional DWT column-wise first and then row-wise to achieve the same result. Figure 3.5 comprehends the idea describe above.

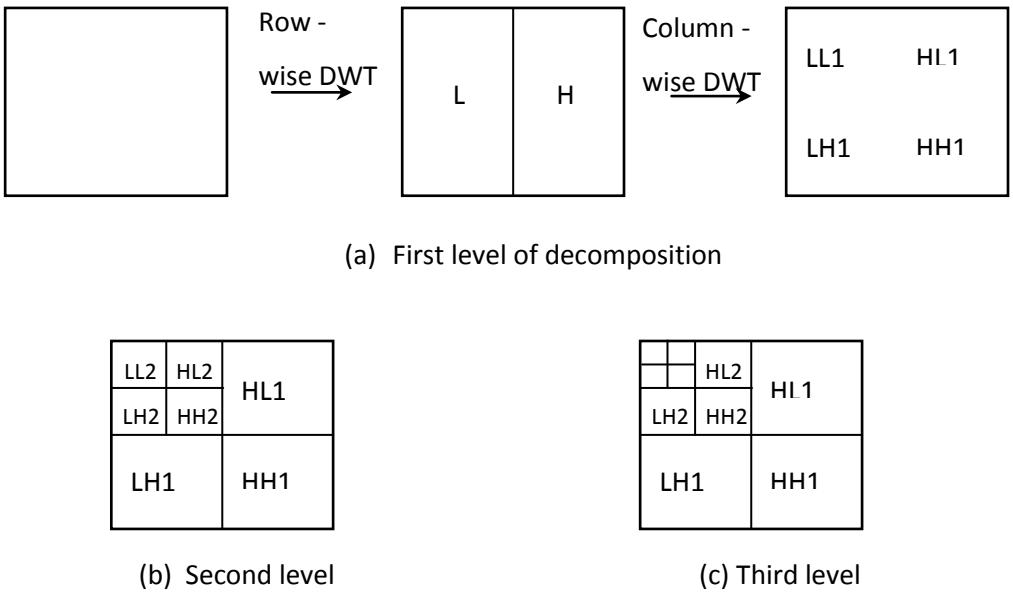


Figure 3.5 Row - Column computation of two-dimensional DWT

The multiresolution decomposition approach in the two-dimensional signal is demonstrated in Figures 3.5 (b) and 3.5 (c). After the first level of decomposition, it generates four subbands LL1, HL1, LH1, and HH1 as shown in Figure 3.5 (a). Considering the input signal is an image, the LL1 subband can be considered as a 2:1 subsampled (both horizontally

and vertically) version of image. The other three subbands HL1, LH1, and HH1 contain higher frequency detail information. These spatially oriented (horizontal, vertical or diagonal) subbands mostly contain information of local discontinuities in the image and the bulk of the energy in each of these three subbands is concentrated in the vicinity of areas corresponding to edge activities in the original image. Since LL1 is acoarser approximation of the input, it has similar spatial and statistical characteristics to the original image. As a result, it can be further decomposed into four subbands LL2, LH2, HL2 and HH2 as shown in Figure 3.5(b) based on the principle of multiresolution analysis. Accordingly the image is decomposed into 10 subbands LL3, LH3, HL3, HH3, HL2, LH2, HH2, LH1, HL1 and HH1 after three levels of pyramidal multiresolution subband decomposition, as shown in Figure 3.5 (c). The same computation can continue to further decompose LL3 into higher levels.

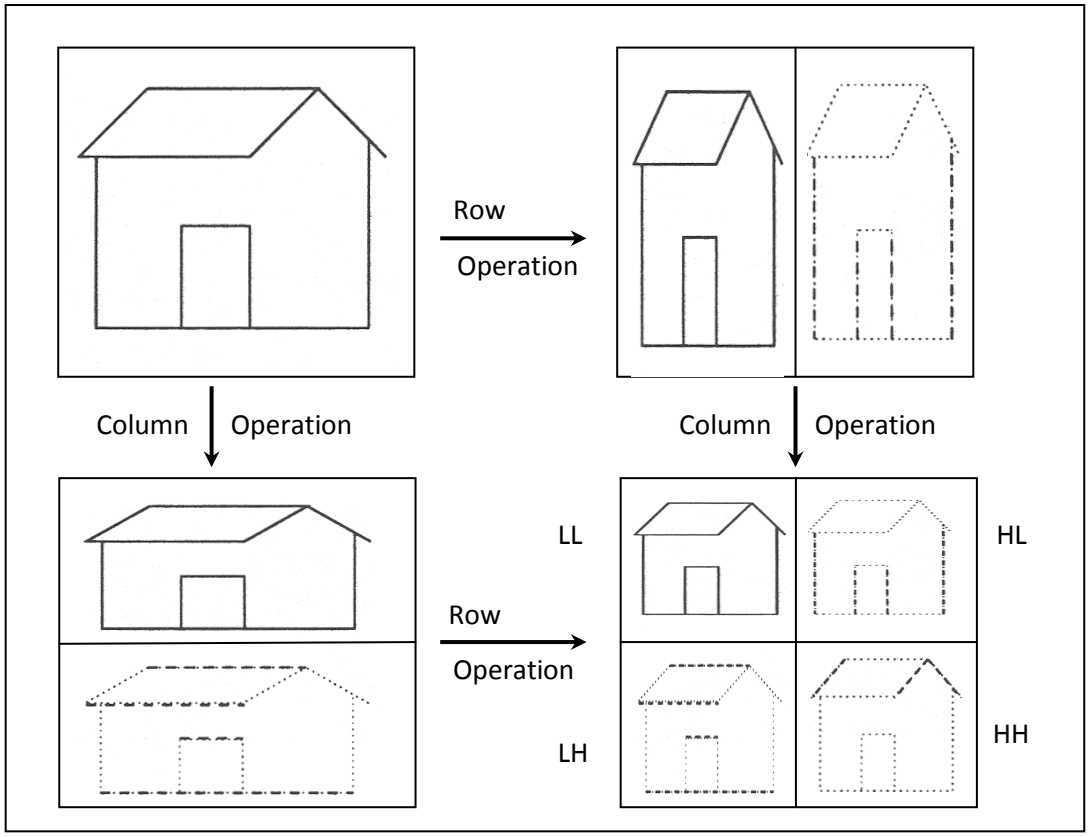


Figure 3.6 Extension of DWT in two - dimensional signals.

3.2. Feature Points Extraction

In this section, we shall describe in detail the process for applying wavelet transforms in the detection of feature points. Wavelet transform (WT) [30] for multiresolution local analysis on signals has been proved to be very effective. It has been successfully applied to many image analysis tasks such as edge detection [31] [32], corner detection [33], texture classification, object recognition, image segmentation, and shape recovery. In this section, we shall introduce how this technique is applied to the problem of scene registration.

3.2.1 Multiscale edge detection

Most of the multiscale edge detectors smooth the signal at various scales and detect sharp variation points from their first- or second-order derivative. The extrema of the first derivative correspond to the zero crossings of the second derivative and to the inflection points of the smoothed signal. This section explains how these multiscale edge detection algorithms are related to the wavelet transform.

We call a smoothing function any function $\theta(x)$ whose integral is equal to 1 and that converges to 0 at infinity. For example, one can choose $\theta(x)$ equal to a Gaussian. We suppose that $\theta(x)$ is twice differentiable and define, respectively, $\psi^a(x)$ and $\psi^b(x)$ as the first- and second-order derivative of $\theta(x)$.

$$\psi^a(x) = \frac{d\theta(x)}{dx} \quad \text{and} \quad \psi^b(x) = \frac{d^2\theta(x)}{dx^2} \quad \text{----------} \quad (3.24)$$

By definition, the functions $\psi^a(x)$ and $\psi^b(x)$ can be considered to be wavelets because their integral is equal to 0

$$\int_{-\infty}^{+\infty} \psi^a(x)dx = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} \psi^b(x)dx = 0$$

In this one, we denote

$$\xi_s(x) = \frac{1}{s}\xi(\frac{x}{s})$$

which is the dilation by a scaling factor s of any function $\xi(x)$.

A wavelet transform is computed by convolving the signal with a dilated wavelet. The wavelet transform of *f(x)* at the scale s and position x, computed with respect to the wavelet $\psi^a(x)$, is defined by

$$W_s^a f(x) = f * \psi_s^a(x) \text{ ---------- (3.25)}$$

The wavelet transform of *f(x)* with respect to $\psi^b(x)$ is

$$W_s^b f(x) = f * \psi_s^b(x) \text{ ---------- (3.26)}$$

We derive that

$$W_s^a f(x) = f * (s\frac{d\theta_s}{dx})(x) = s\frac{d}{dx}(f * \theta_s)(x) \text{ ---------- (3.27)}$$

And

$$W_s^b f(x) = f * (s^2\frac{d^2\theta_s}{dx})(x) = s^2\frac{d^2}{dx^2}(f * \theta_s)(x) \text{ ---------- (3.28)}$$

The wavelet transforms $W_s^a f(x)$ and $W_s^b f(x)$ are, respectively, the first and second derivative of the signal smoothed at the scale s. The local extrema of $W_s^a f(x)$ thus correspond to the zero crossings of $W_s^b f(x)$ and to the inflection points of $f * \theta_s(x)$. In the particular case where $\theta(x)$ is a Gaussian, the zero-crossing detection is equivalent to a Marr-Hildreth edge detection, whereas the extrema detection corresponds to a Canny [34] edge detection. When the scale s is large, the convolution with $\theta_s(x)$ removes small signal fluctuations; we therefore only detect the sharp variations of large structures.

Detecting zero crossings or local extrema are similar procedures, but the local extrema approach has some important advantages. An inflection point of $f * \theta_s(x)$ can either be a maximum or a minimum of the absolute value of its first derivative. The maxima of the

absolute value of the first derivative are sharp variation points of $f * \theta_s(x)$, whereas the minima correspond to slow variations. With a second derivative operator, it is difficult to distinguish these two types of zero crossings. On the contrary, with a first-order derivative, we easily select the sharp variation points by detecting only the local maxima of $\left|W_s^a f(x)\right|$. In addition, zero crossings give position information but do not differentiate small amplitude fluctuations from important discontinuities. When detecting local maxima, we can also record the values of $\left|W_s^a f(x)\right|$ at the maxima locations, which measure the derivative at the inflection points.

The Canny edge detector is easily extended in two dimensions. We denote by

$$\xi_s(x, y) = \frac{1}{s^2}\xi(\frac{x}{s}, \frac{y}{s})$$

the dilation by s of any 2-D function $\xi(x, y)$. We use the term 2-D smoothing function to describe any function $\theta(x, y)$ whose integral over x and y is equal to 1 and converges to 0 at infinity. The image *f(x, y)* is smoothed at different scales s by a convolution with $\theta_s(x, y)$. We then compute the gradient vector $\vec{\nabla}(f * \theta_s)(x, y)$. The direction of the gradient vector at a point (x_0, y_0) indicates the direction in the image plane *(x, y)* along which the directional derivative of *f(x, y)* has the largest absolute value. Edges are defined as points (x_0, y_0) where the modulus of the gradient vector is maximum in the direction towards which the gradient vector points in the image plane. Edge points are inflection points of the surface $f * \theta_s(x, y)$.

Let us relate this edge detection to a 2-D wavelet transform. First of all, let *S*(*x*, *y*) be a 2D smoothing function. Two wavelets, $\psi^1(x, y)$ and $\psi^2(x, y)$ are the partial derivatives of the smoothing function *S*(*x*, *y*) in the *x* and *y* directions, respectively, where

$$\psi^1(x, y) = \frac{\partial S(x, y)}{\partial x} \text{ ---------- (3.29)}$$

and

$$\psi^2(x, y) = \frac{\partial S(x, y)}{\partial y} \text{ ---------- (3.30)}$$

Let $\psi^1_{2^j}(x,y) = \frac{1}{2^j}\psi^1(\frac{x}{2^j},\frac{y}{2^j})$ and $\psi^2_{2^j}(x,y) = \frac{1}{2^j}\psi^2(\frac{x}{2^j},\frac{y}{2^j})$ be two wavelets which are dilated of $\psi^1(x,y)$ and $\psi^2(x,y)$ by a scaling factor of 2^j, respectively.

The 2D wavelet transform of an image *f(x, y)* at scale 2^j (i.e. level *j*) can be decomposed into two independent directions as

$W^1_{2^j} f(x,y) = f * \psi^1_{2^j}(x,y)$ and $W^2_{2^j} f(x,y) = f * \psi^2_{2^j}(x,y)$

It is derived in [1] that

$$W^1_{2^j} f(x,y) = f * \psi^1_{2^j}(x,y)$$

$$= f * (2^j \frac{\partial S_j}{\partial x})(x,y)$$

$$= 2^j \frac{\partial}{\partial x}(f * S_j)(x,y) \quad \text{---------- (3.31)}$$

and

$$W^2_{2^j} f(x,y) = f * \psi^2_{2^j}(x,y)$$

$$= f * (2^j \frac{\partial S_j}{\partial y})(x,y)$$

$$= 2^j \frac{\partial}{\partial y}(f * S_j)(x,y) \quad \text{---------- (3.32)}$$

That is,

$$\begin{pmatrix} W^1_{2^j} f(x,y) \\ W^2_{2^j} f(x,y) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial x}(f * S_j)(x,y) \\ \frac{\partial}{\partial y}(f * S_j)(x,y) \end{pmatrix} = 2^j \vec{\nabla}(f * S_j)(x,y) \quad \text{------ (3.33)}$$

Hence, the two components of the wavelet transform are proportional to the two components of the gradient vector. At each level, the modulus of the gradient vector is proportional to

$$M_{2^j} f(x,y) = \sqrt{\left|W^1_{2^j} f(x,y)\right|^2 + \left|W^2_{2^j} f(x,y)\right|^2} \quad \text{---------- (3.34)}$$

If the local maxima of $M_{2^j} f(x, y)$ are located and thresholded with a preset value, then all the edge points of *f (x, y)* at scale 2^j can be detected.

3.3. FINDING CORRECT MATCHING PAIRS

In the previous section, we have presented a systematic way to extract important features from two partially overlapping images. In this section, we shall show how to find a set of correct matching pairs between the above images. In what follows, the procedure will be elaborated in detail step by step.

3.3.1 Defining a Matching Metric

In this subsection we shall define a metric and then use it to evaluate the similarity between any two feature points. Let $p = (p_x, p_y)^t$ and $q = (q_x, q_y)^t$ be two feature points located, respectively, in $f_1(x, y)$ and $f_2(x, y)$, where $f_1(x, y)$ and $f_2(x, y)$ are two partially overlapping images. A crosscorrelation which can be used to measure the similarity degree between *p* and *q* is defined as [2, 3].

$$C_{f_1,f_2}(p;q) = \frac{1}{\sigma_1\sigma_2(2M+1)^2}\sum_{x,y=-M}^{x,y=M}[f_1(x+p_x, y+p_y)-\mu_1][f_2(x+q_x, y+q_y)-\mu_2],$$

---------- (3.35)

Where μ_i and σ_i are the local mean and variance of f_i (x, y), respectively; $(2M+1)^2$ represents the area of matching window. In general, the format of the similarity measure defined in Eq. (3.35) is very sensitive to rotation. Therefore, if the rotation effect is important in an application, Eq. (3.35) should be updated as

$$\overline{C}_{f_1,f_2}(p;q;\theta) = \frac{1}{\sigma_1\sigma_2(2M+1)^2}\sum_{\hat{x},\hat{y}=-M}^{\hat{x},\hat{y}=M}[f_1(x+p_x, y+p_y)-\mu_1][f_2(\hat{x}+q_x, \hat{y}+q_y)-\mu_2],$$

---------- (3.36)

Where $x = \hat{x}\cos\theta - \hat{y}\sin\theta$ and $y = \hat{x}\sin\theta + \hat{y}\cos\theta$. If the angle θ in Eq. (3.3.2) can be estimated in advance, then no matter how f_1 (x, y) or f_2(x, y) are rotated, finding the correct matching between the two images is always easier.

3.3.2 Estimating the Orientation of a Feature Point

In the previous subsection, we have mentioned that the orientation of a feature point is important for deriving a correct metric. In this subsection we shall discuss how the orientation of a feature point is estimated. In Section 3.2.1, we have mentioned that two sets of feature points are extracted, respectively, from two partially overlapping images. In order to perform accurate image registration between these two images, the corresponding feature points between the two images have to be identified. In Eq. (3.36), the fit measure $\overline{C}_{f_1,f_2}$ contains a rotation angle θ, which represents the orientation difference between two selected feature points. In order to solve θ, we have to determine the orientation of each point in advance. Basically, the orientation of a feature point can be estimated by using the results of the wavelet transform described in Section 3.2.1, i.e., $W^1_{2^j} f$ an $W^2_{2^j} f$. A standard representation of the orientation of an edge-based feature point at scale 2^j can be expressed as follows [31, 35]:

$$Arg(W^1_{2^j} f(x,y) + iW^2_{2^j} f(x,y)).$$

3.3.3 Estimating the Orientation Difference between Two Overlapping Images

From the detected feature points, as well as their corresponding orientations, it is not difficult to determine the matching pairs between two overlapping images. In what follows, we shall describe how the above mentioned information is used to estimate the orientation difference between two overlapping images. Let $FP_{f_1} = \{p_i = (p^i_x, p^i_y)^t\}_{i=1,2,.....N_{f_1}}$ and $FP_{f_2} = \{q_j = (q^j_x, q^j_y)^t\}_{i=1,2,.....N_{f_2}}$ be two sets of feature points extracted from two partially overlapping images, f_1 and f_2, respectively. N_{f_1} and N_{f_2}, respectively, represent the number

of elements in FP_{f_1} and FP_{f_2} . Let A (u) be the angle of an edge (or feature) point u. For a feature point p_i in FP_{f_1} and a feature point q_j in FP_{f_2} , the orientation difference between p_i and q_j can be calculated as

$$\theta_{i,j} = A(q_j) - A(p_i). \quad \text{---------- (3.37)}$$

Here, $\theta_{i,j}$ ranges from 0^0 to 359^0, and its value has to be an integer. In order to estimate the orientation difference between f_1 and f_2, $\theta_{i,j}$ and the similarity measure $\overline{C}_{f_1,f_2}$ are used. A so-called ''angle histogram'' H (θ) reports the distribution of the number of $\{p_i \Leftrightarrow q_j\}$ pairs at angle θ that satisfy the conditions $\theta_{i,j} = \theta$ and $\overline{C}_{f_1,f_2}(p_i; q_j; \theta_{i,j}) \geq 0.8$. Basically, from the angle histogram, the rotation angle can be decided by seeking the angle that corresponds to the maximum peak in the histogram. If the orientation difference between f_1 and f_2 is $\overline{\theta}$, then the highest peak in H(θ) must be very close to $\overline{\theta}$. Further, it is possible to modify H (θ) so that a more accurate estimation can be obtained. The modification is.

$$\overline{H}(\theta) = \sum_{i=-2}^{2} H(\theta + i) \quad \text{---------- (3.38)}$$

Here, $\theta + i$ may possibly exceed the limitation of 360^0. Under these circumstances, it has to be adjusted by modulation. Using $\overline{H}(\theta)$ and finding its maximum peak, a very accurate estimation for the orientation difference between two partially overlapping images can be found. In comparison with some conventional methods which require derivation of the correlations of all feature points in a window, the number of feature points used in the proposed method is very few. Therefore, the calculation of the similarity measure $\overline{C}_{f_1,f_2}$ between any $\{p_i \Leftrightarrow q_j\}$ pair only requires very short computation time.

3.3.4 Finding the Matching Pairs

In the previous section, we have described how the orientation difference between $f_1(x, y)$ and $f_2(x, y)$ is derived. Now, we are ready to find the matching feature point pairs between $f_1(x, y)$ and $f_2(x, y)$.The first step of the process is to rotate all the points in $f_1(x, y)$ to

their new positions located in $\hat{f}_1(\hat{x},\hat{y})$, where $\hat{x} = x\cos\bar{\theta} + y\sin\bar{\theta}$, $\hat{y} = -x\sin\bar{\theta} + y\cos\bar{\theta}$. If an edge point in f_1 is rotated to a new position in $\hat{f}_1$, the angle θ is also adjusted to suit the coordinates of $\hat{f}_1$, i.e., $\hat{\theta} \to \theta + \bar{\theta}$. Here, $\bar{\theta}$ is the orientation difference between f_1 and f_2. Let E(u) denote the set of edge points within a $(2M_s+1)^2$ window of an image *f*, where u is the window's center. Given a feature point $\hat{p}_i$ in $\hat{f}_1(x,y)$ the matching problem is used to find a proper point $\tilde{q}$ in E (q_k) for every $q_k \in FP_{f_2}$ such that the pair $\{\hat{p}_i \leftrightarrow \tilde{q}\}$ becomes a matching pair. A pair $\{\hat{p}_i \leftrightarrow \tilde{q}\}$ is qualified as a matching pair if two conditions are satisfied:

Condition 1. $\overline{C}_{\hat{f}_1,\hat{f}_2}(\hat{p}_i;\tilde{q}) = \max_{q_k \in FP_{f_2}} \max_{\tilde{q}_n \in E(q_k)} \overline{C}_{\hat{f}_1,f_2}(\hat{p}_i;\tilde{q}_n)$,

Condition 2. $\overline{C}_{\hat{f}_1,f_2}(\hat{p}_i;\tilde{q}) > T_c$, Where T_c =0.75

Condition 1 enforces finding an edge point $\tilde{q} \in E(q_k)$ and $q_k \in FP_{f_2}$ such that the measure $\overline{C}_{\hat{f}_1,f_2}$ is maximized. As for Condition 2, it forces the value of $\overline{C}_{\hat{f}_1,f_2}$ of a matching pair to be larger than a threshold (0.75 in this case). Further, by introducing another constraint, the orientation criterion, the searching speed can be even faster. As we know, $\hat{f}_1(x,y)$ is obtained by rotating $f_1(x,y)$ with an angle $\bar{\theta}$; therefore, the orientation difference between f $\hat{f}_1(x,y)$ and another image $f_2(x, y)$ is very small. Hence, it is reasonable to introduce another constraint which forces the orientation difference between $\hat{f}_1(x,y)$ and $f_2(x, y)$ to be less than 5^0, i.e., $|A(\hat{p}_i) - A(\tilde{q})| < 5^0$. Adding this criterion and using it, together with the previous two conditions, will significantly speed up the search time. In real implementation, the orientation constraint will be tested first. If the constraint is not satisfied, it is not necessary to test Condition 1 and Condition 2. In this way, only a few pairs are needed to calculate the cross-correlation measure $\overline{C}_{\hat{f}_1,f_2}$ which is considered a time bottleneck of the whole process.

3.3.5 Eliminating the False Matching Pairs

A new method is proposed to eliminate the incorrectly matched pairs. Li *et al.* [1] proposed an iterative scheme that could not remove false pairs completely and efficiently.

Here, we present a noniterative scheme based on the idea that the distance between two points in the same image will be preserved when it undergoes a rigid transform. Let $MP = \{p_i \Leftrightarrow q_i\} i = 1,2,.....N_m$ be a set of matching pairs, where N_m represents the number of elements in *MP*, $p_i = (p^i{}_x, p^i{}_y)$ is a point in $\hat{f}_1(x, y)$ and $q_i = (q^i{}_x, q^i{}_y)$ is a point in $f_2(x, y)$. If all the matching pairs in *MP* are correct, then the following equation should hold, i.e,

$$p_i = sq_i + T \text{ for } i=1,2,\ldots\ldots N_m \text{ ---------- (3.39)}$$

Where *s* and *T* are, respectively, a scalar and a translation vector. Since the orientation difference between $\hat{f}_1(x, y)$ and $f_2(x, y)$ is very small, the rotation matrix thus is not part of Eq. (3.39). Let $\{p_i \Leftrightarrow q_i\}$ and $\{p_j \Leftrightarrow q_j\}$ be two correct matching pairs in MP. The scale *s* between $\hat{f}_1(x, y)$ and $f_2(x, y)$ can be estimated as

$$s = d_2 / d_1 \text{ ---------- (3.40)}$$

Where

$$d1 = \sqrt{(p^i{}_x - p^j{}_x)^2 + (p^i{}_y - p^j{}_y)^2} \text{ and } d2 = \sqrt{(q^i{}_x - q^j{}_x)^2 + (q^i{}_y - q^j{}_y)^2}$$

Once the scale s is known, the translation T^i between p_i and q_i can be calculated by using Eq.(3.39). Furthermore, the translation T^j between p_j and q_j can be obtained accordingly. Since the pairs $\{p_i \Leftrightarrow q_i\}$and $\{p_j \Leftrightarrow q_j\}$ are correct matching pairs, the difference between Ti and Tj should be very small. Therefore, by checking the distance between Ti and Tj, we can decide whether $\{p_i \Leftrightarrow q_i\}$and $\{p_j \Leftrightarrow q_j\}$ are consistent or not. For the set of all possible matching pairs, a list of counters records how many other pairs are consistent with current pair. For those matching pairs with the largest consistency count, select the matching pair with maximum matching similarity (i.e. normalized correlation value) as a reliable pair, and eliminate all the matching pairs not consistent with the selected pair.

Next, we propose a noniterative method based on the aforementioned consistency test to eliminate those mistakenly matched pairs. Let S(i) denote a counter of the number of times the $\{p_i \Leftrightarrow q_i\}$pair is consistent with other matching pairs. Assume that $\{p_j \Leftrightarrow q_j\}$ is a matching pair to be checked. $\{p_j \Leftrightarrow q_j\}$ is considered to be consistent $\{p_i \Leftrightarrow q_i\}$ if and only

if the distance between their translation vectors, T^i and T^j, is less than a threshold. If the two pairs are consistent, S(i) is increased by 1. The process proceeds until all the matching pairs are compared. Since there are Nm elements in MP, the total number of consistency tests will be $N_m(N_m-1)/2$. After the consistency test, a counter value S(i) will be associated with every matching pair $\{p_i \Leftrightarrow q_i\}$ for i= 1, 2, . . . , N_m. Since two matching pairs can uniquely determine a set of registration parameters, the value 2 can be used as a threshold to determine whether the pair $\{p_i \Leftrightarrow q_i\}$ can be accepted. However, if $\{p_i \Leftrightarrow q_i\}$ is said to be a correct match, to compensate for inaccuracies, we require that the value of S(i) be larger than 2. Therefore, if the value of S(i) is less than or equal to 2, then its corresponding matching pair is considered mismatched and should be eliminated.

3.4 DERIVING THE TRANSFORMATION PARAMETERS

After the elimination of the false-matching pairs, a set of correct matching pairs are left. These correct matching pairs are used to derive the image transformation parameters.

3.4.1 Deriving four rigid planar parameters

Initially, four rigid planar parameters (i.e. scaling, rotation angle, and two translation shifts) are calculated by a singular value decomposition (SVD) method [36][37].

Let $M = \{p_i \Leftrightarrow q_i, i = 1,2,.....,N_c\}$ be the set of correct matching pairs, where N_c is the total number of correct matching pairs. In order to derive four rigid planar parameters (*s*, *R* and *T*) based on the relation:

$$q_i = sRp_i + T \quad \text{for i=1,2,3,……., } N_c \text{ , ----------(3.41)}$$

Where s is a scalar, $R = \begin{bmatrix} \cos\hat{\theta} & -\sin\hat{\theta} \\ \sin\hat{\theta} & \cos\theta \end{bmatrix}$ represents a rotation matrix, $T = (t_x, t_y)^t$ is a translation vector, $P_i = (p_x^i, p_y^i)^t$ is a point in $\hat{f}_1(x,y)$, $q_i = (q_x^i, q_y^i)^t$ represents a point in $f_2(x, y)$ and $\hat{\theta}$ is the orientation difference between $\hat{f}_1(x,y)$ and f_2 (x, y). In Section 3.3.3, the

initial orientation angle $\bar{\theta}$ between $f_1(x, y)$ and $f_2(x, y)$ has been estimated by using the histogram $\overline{H}(\theta)$. However, $\bar{\theta}$ can only be considered a rough guess. In what follows, we shall take advantage of the set of correct matching pairs, $\{p_i \Leftrightarrow q_i\}_{i=1,2,.....N_c}$ to fine-tune the previous result. Basically, the fine-tuning, $\bar{\theta}$ is conducted to derive the orientation difference between $\hat{f}_1$ and f_2. If $\hat{\theta}$ can be derived, a more accurate rotation angle, $\theta = \hat{\theta} + \bar{\theta}$, between f_1 and f_2 can be obtained. Next, a method which can be used to derive s, $\hat{\theta}$, and T, based on the correct matching pairs is presented.

In order to derive s, R, and T based on the set of correct matching pairs $\{p_i \Leftrightarrow q_i\}_{i=1,2,.....N_c}$.

The error function is defined as

$$\phi = \sum_{i=1}^{N_c} \left\| sRp_i + T - q_i \right\|^2 .....(3.42)$$

By minimizing ϕ, a set of optimal solutions can be derived. In [38], Umeyama proposed a good approach to solving the above problem. Here, we will simply follow his methodology.

From $\frac{\partial \phi}{\partial T} = 0$, one can obtain

$$T = \bar{q} - sR\bar{p} .....(3.43)$$

Where

$$\bar{p} = \frac{1}{N_c} \sum_{i=0}^{N_c} p_i \quad \text{and} \quad \bar{q} = \frac{1}{N_c} \sum_{i=0}^{N_c} q_i .$$

Substituting Eq. (3.43) into Eq. (3.41) and from $\frac{\partial \phi}{\partial s} = 0$, we have

$$s = \frac{\sum_{i=1}^{N_c} \tilde{q}_i^{\,t} R \tilde{p}_i}{\sum_{i=1}^{N_c} \tilde{p}_i^{\,t} \tilde{p}_i} .....(3.44)$$

Where $\tilde{p}_i = p_i - \bar{p}$ and $\tilde{q}_i = q_i - \bar{q}$.

Substituting Eqs. (3.43) and (3.44) into Eq. (3.41) and reorganizing the content of ϕ we obtain

$$\phi = \sum_{i=1}^{N_c} \tilde{q}_i{}^t \tilde{q}_i - \frac{\left[\sum_{i=1}^{N_C} \tilde{q}_i{}^t R \tilde{p}_i\right]^2}{\sum_{i=1}^{N_c} \tilde{p}_i{}^t \tilde{p}_i} \quad \text{.....(3.45)}$$

Therefore, minimizing ϕ can be converted into maximizing the term

$$\phi' = \left[\sum_{i=1}^{N_c} \tilde{q}_i{}^t R \tilde{p}_i\right] \quad \text{.....(3.46)}$$

Now, the problem at hand is to solve *R*. In [39], Arun *et al.* proposed a singular value decomposition (SVD) method to solve *R*. The procedure is illustrated as follows:

Step 1. Calculate the 2×2 matrix *H*, that is

$$H = \sum_{i=1}^{N_c} \tilde{p}_i \tilde{q}_i{}^t$$

Step 2. Find the SVD of *H*, i.e.

$$H = P \wedge Q^t$$

Step 3.
$$R = QP^t \quad \text{---------- (3.47)}$$

The whole procedure for estimating the registration parameters can be summarized as follows. First, the rotation matrix R is found by solving Eq. (3.47). Then, the fine-tuning, $\hat{\theta}$, between $\hat{f}_1(x, y)$ and f_2(x, y) can be estimated from R. Next, by solving Eq. (3.44), the scale *s* can be obtained. Furthermore, by solving Eq. (3.43), the translation vector T is obtained. Based on these parameters, registration between images can be achieved.

3.4.2 REFINE MATCHING

In the section, an iterative procedure is proposed to repeatedly refine the registration parameters. In the first step, with the assistance of the initial values of *s*, *R*, and *T*, the input image $f_1(x, y)$ is transformed to $\hat{f}_1(x, y)$, and the relation between $(x, y)^t$ and $(\hat{x}, \hat{y})^t$ is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{s}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \hat{x} - t_x \\ \hat{y} - t_y \end{pmatrix} \text{---------- (3.48)}$$

Where $\theta = \overline{\theta} + \hat{\theta}$, $\overline{\theta}$ is obtained by applying the method described in Section 3.3.3, and $\hat{\theta}$ is obtained from Eq. (3.47). Let $M_p = \{p_i \Leftrightarrow q_i\} i = 1,2,.....N_c$ denote the set of matching pairs obtained in the previous stage, where N_c is the number of elements in MP, p_i a point in f_1, and q_i is a point in f_2. For each point q_i in f_2, our goal is to find its exact corresponding point $\hat{p}_i$ in $\hat{f}_1$ such that the refined corrections for the parameters *s*, *R*, and *T* can be obtained. It is known that $\hat{f}_1$ is obtained by transforming f_1 with the initial values of parameters *s*, *R*, and *T*; therefore, the location of $\hat{p}_i$ in $\hat{f}_1$ should have a coordinate very similar to q_i in f_2. Hence, for each q_i in f_2, its corresponding point $\hat{p}_i$ in $\hat{f}_1$ can be found by searching the pixels within a neighborhood centered at the coordinate $\hat{p}_i$ in $\hat{f}_1$ such that the measure $C_{\hat{f}_1, f_2}$ is maximized and has a value larger than a threshold; that is,

$$C_{\hat{f}_1, f_2}(\hat{p}_i, q_i) = \max_{p \in N_{q_i}} C_{\hat{f}_1, f_2}(p, q_i) \quad , \; C_{\hat{f}_1, f_2}(\hat{p}_i, q_i) \geq 0.75 \quad \text{--------- (3.49)}$$

Where N_{q_i} is the neighborhood centered at the coordinate q_i. In most cases, *Nc* is small. If it is larger, then the set M_p can be resampled so that it will be controlled properly. In general, the use of only a few feature points is sufficient to derive *R*, *T*, and *s* with high accuracy. Therefore, the refining process can be performed very quickly. Let $\hat{M}_p = \{\hat{p}_i \Leftrightarrow q_i\}_{i=1,2,.....\hat{N}_c}$ denote the set of matching pairs obtained from the refining process, where $\hat{N}_c$ is the number of elements in $\hat{M}_p$. From the set $\hat{M}_p$, the refined corrections $\hat{s}$, $\hat{\theta}$, $\hat{t}_x$ and $\hat{t}_y$ for the parameters *s*, θ, t_x, and t_y can be obtained by Eqs. (43), (44), and (47), respectively. With $\hat{s}$, $\hat{\theta}$, $\hat{t}_x$ and $\hat{t}_y$, more accurate values of *s*, θ, t_x, and t_y can be derived. Since

$$\begin{aligned} q_i &= \hat{s}R(\hat{\theta})\hat{p}_i + \hat{T} \\ &= \hat{s}R(\hat{\theta})[sR(\theta)p_i + T] + \hat{T} \\ &= s\hat{s}R(\theta + \hat{\theta})p_i + [\hat{s}R(\hat{\theta})T + \hat{T}], \end{aligned}$$

the values of s, θ, t_x, and t_y can be corrected, based on the following equation [3]:

$$(s,\theta,t_x,t_y)^t \Leftarrow (s\hat{s},\theta+\hat{\theta},\hat{s}\cos\hat{\theta}t_x+\hat{s}\sin\hat{\theta}t_y+\hat{t}_x,-\hat{s}\sin\hat{\theta}t_x+\hat{s}\cos\hat{\theta}t_y+\hat{t}_y)^t$$

---------- (3.50)

Using the results of (3.50) as a new set of initial values in Eq. (3.48) and applying the refining process iteratively, all the registration parameters can be updated continuously until satisfactory accuracy is achieved.

CHAPTER 4

EXPERIMENTAL RESULTS

TABLE 4.1

Registration parameters after applying registration algorithm

Images	Scale: s	Traslation: t_x	Translation: t_y	Rotation: θ
Building	0.9986	-117.748	43.7386	22.7470^0
Aerial	1.0477	-271.135	-235.781	9.1615^0

Figure 4.1(a) and 4.1(b) are images of a building taken from different views are act as input image pair. Figure 4.1(c) and figure 4.1(d) show wavelet decomposition of figure 4.1(a) and figure 4.1(b) respectively. Figure 4.1(e) and figure 4.1(f) show edge detection of decomposed images of figure 4.1(c) and figure 4.1(d) respectively. The registered image of two input images is shown in figure 4.1(g).

Figure 4.2(a) and 4.2(b) are images of a building taken from different views are act as input image pair. Figure 4.2(c) and figure 4.2(d) show wavelet decomposition of figure 4.2(a) and figure 4.2(b) respectively. Figure 4.2(e) and figure 4.2(f) show edge detection of decomposed images of figure 4.2(c) and figure 4.2(d) respectively. The registered image of two input images is shown in figure 4.2(g).

Figure 4.1(a): Building image 1

Figure 4.1(b): Building image 2

Figure 4.1(c): Wavelet decomposition of figure 4.1(a)

Figure 4.1(d): Wavelet decomposition of figure 4.1(b)

Figure 4.1(e): Feature point extraction of figure 4.1(c)

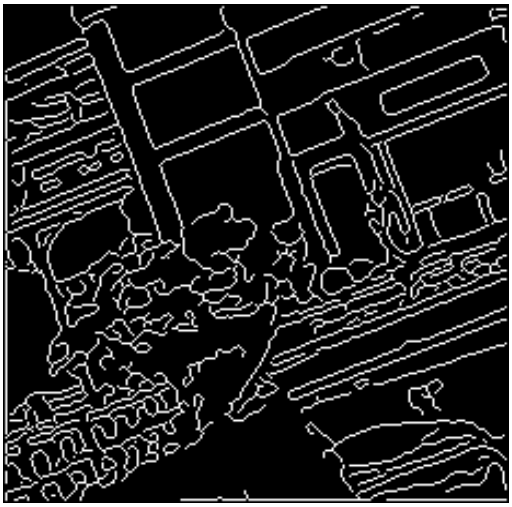


Figure 4.1(f): Feature point extraction of figure 4.1(d)

Figure 4.1(g): Registration result of figure 4.1(a) and figure 4.1(b)

Figure 4.2(a): Aerial image 1

Figure 4.2(b): Aerial image 2

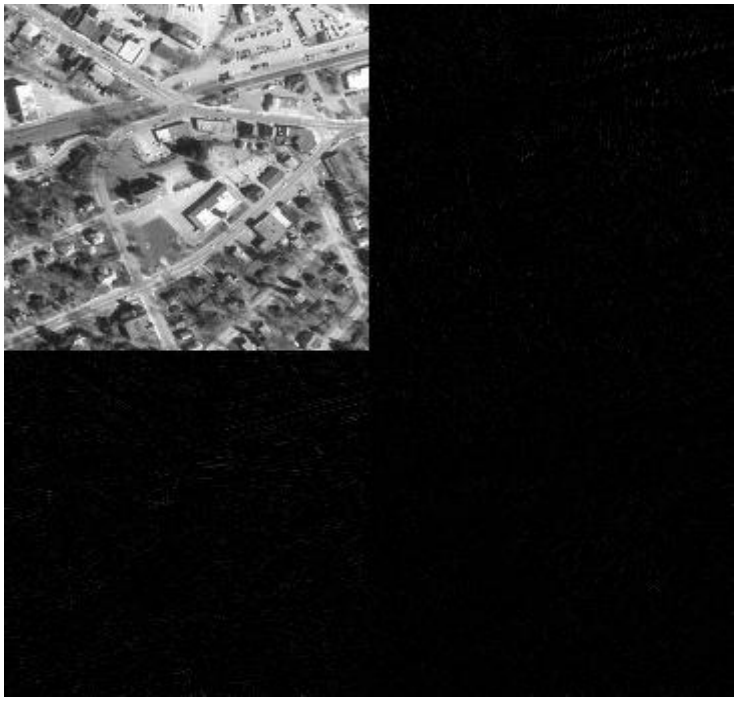


Figure 4.2(c): Wavelet decomposition of figure 4.2(a)

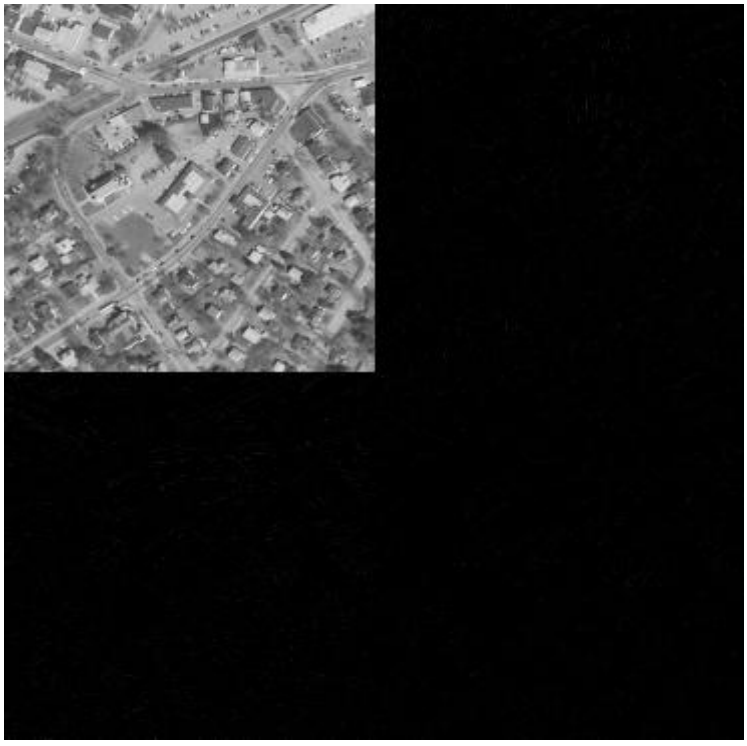


Figure 4.2(d): Wavelet decomposition of figure 4.2(b)

Figure 4.2(e): Feature point extraction of figure 4.2(c)

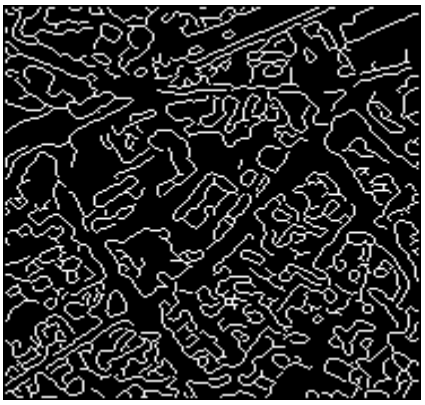


Figure 4.2(f): Feature point extraction of figure 4.2(d)

Figure 4.2(g): Registration result of figure 4.2(a) and figure 4.2(b)

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The proposed method applies the wavelet transform technique to extract feature points from a partially overlapping image pair. By defining a similarity measure metric, the two sets of feature points can be compared, and the correspondences between the feature points can be established. Once the set of correctly matched feature point pairs between two images are found, the registration parameters can be derived accordingly. Hence the registered image of two input images can be obtained.

5.2 Future work

In the future, the idea of an ultimate registration method, able to recognize the type of given task and to decide by itself about the most appropriate solution, can motivate the development of expert systems. They will be based on the combination of various approaches, looking for consensus of particular results.

REFERENCES

[1] H. Li, B. S. Manjunath, and S. K. Mitra, A contour-based approach to multisensor image registration, IEEE Trans. Image Processing, vol. 4, no. 3, pp. 320–334, March 1995.

[2] L. G. Brown, A survey of image registration techniques, ACM Comput. Surv. 24, No. 4, 1992, 325–376.

[3] Q. Zheng and R. Chellappa, A computational vision approaches to image registration, IEEE Trans. Image Process. 2, No. 3, 1993, 311–326.

[4] E. De Castro and C. Morandi, ‘‘Registration of translated and rotated image using finite Fourier transform,’’ IEEE Trans. Pattern Anal. Machine Intell., vol. 9, no. 5, pp. 700–703, Sept. 1987.

[5] 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.

[6] http://en.wikipedia.org/wiki/Image_registration.

[7] Gang Hong and Yun Zhang, “Combination of feature-based and area-based image registration technique for high resolution remote sensing image,” IEEE International conference July 2007.

[8] Bookstein, F.L., 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Pan. Anal. Mach. Intell. 11 (6), 567-585.

[9] Amit, Y., "Graphical shape templates for automatic anatomy detection with applications to MRI brain scan", IEEE Trans. Med. Imaging 16 (1), 28-40, 1997.

[10] Davis, M.N., Khotanzad, A., Flamig, D.P., Harms, S.E., 1997. "A physics based coordinate transformation for 3D image matching", IEEE Trans. Med. Imaging 16 (3), 317-328.

[11] Subsol, G., Thirion, J. P., and Ayache, N., "A general scheme for automatically building 3-D morphometric anatomical atlases: Application to a skull atlas", Med. Image Anal., vol. 2, no. 1, pp. 37-60, 1998.

[12] Ge, Y., Fitzpatrick, J.M., Kessler, R.M., Jeske-Janicka, M., "Inter- subject brain image registration using both cortical and subcortical landmarks". In: Proceedings Image Processing, SPIE 2434, pp. 81-95, 1995.

[13] Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L., "A Feature-Based, Robust, Hierarchical Algorithm for Registering Pairs of Images of the Curved Human Retina", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 3, March 2002.

[14] Szeliski, R. and Lavallee, S., "Matching 3-D anatomical surfaces with nonrigid deformations using octree-splines", SPIE Geometric Meth. Comput. Vis., vol. 2031, pp. 306-315, 1993.

[15] Thompson, P.M., Toga, A.W., "A surface-based technique for warping 3 dimensional images of the brain", IEEE Trans. Med. 15 (4), 402-417, 1996.

[16] Audette, M.A., Ferrie, F.P., Peters, T.M., " An algorithmic overview of surface registration techniques for medical imaging", Medical image Analysis 4 (2000) 201-217.

[17] Ashburmer J, Neelin P, Collins DL, Evans AC, Friston KJ, "Incorporating Prior Knowledge into Image Registration", Neuroimage 6(4):344-352, 1997.

[18] Collins, D.L., Holmes, C.J., Peters, T.M., and Evans, A.C., "Automatic 3D model-based neuroanatomical segmentation". Human Brain Mapping, 3:190-208, 1995.

[19] Kim, B., JL Boes, KA Frey and CR Meyer. "Mutual information for automated unwarping of rat brain autoradiographs". NeuroImage 5(1):31-40, 1997.

[20] Christensen, G.E., Joshi, S.C., and Miller, M.I., "Transformation of Brain Anatomy", IEEE Trans. on Medical Imaging 16:6; 864-877, 1997.

[21] Wu, Y., Kanade, T., Li, C., and Cohn, J., "Image registration using wavelet-based motion model", International Journal of Computer Vision, 38(2), 2000.

[22] Rohr, K., Stiehl, H.S., Buzug, T.M., Weese, J., Kuhn, M.H., "Landmark-based elastic registration using approximating thin-plate splines", IEEE Trans. on Medical Imaging 20; 526-534, 2001.

[23] Rodhe, G., Aldroubi, A., and Dawant, B., "The Adaptive Bases Algorithm for Intensity Based Non Rigid Image Registration", IEEE Trans. on Medical Imaging, Vol. 22, No. 11, November 2003.

[24] Davatzikos, Ch., "Spatial transformation and registration of brian images using elastically deformable models", Comput. Vis. Image Understanding, vol. 66, no. 2, pp. 207-222, 1997.

[25] Christensen, G., Rabbitt, R., and Miller, M., "Deformable templates using large deformation kinematics", IEEE Transactions on Image Processing, 5(10):1435V1447, Oct 1996.

[26] Hellier, P., Barillot, C., Memin, E., and Perez, P., "Hierarchical Estimation of a Dense Deformation Field for 3-D Robust Registration", IEEE Trans. on Medical Imaging, 20(5), pp 338-402, 2001.

[27] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 11, No.7, pp. 674-693, July 1989.

[28] I. Daubechies, "The Wavelet Transform, Time-Frequency localization and Signal Analysis," IEEE Trans. on Inform. Theory, Vol. 36, No. 5, pp. 961-1005, September 1990.

[29] I. Daubechies, Ten Lectures on Wavelets. SIAM, CBMS series) Philadelphia, 1992.

[30] S. Mallat, ‘‘Multifrequency channel decompositions of images and wavelet models,’’ IEEE Trans. Acoustics, Speech, Signal Processing, vol. ASSP-37, no. 12, pp. 2091–2110, Dec. 1989.

[31] S. Mallat and S. Zhong, ‘‘Characterization of signals from multi-scale edges,’’ IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-14, no. 7, pp. 710–732, July 1992.

[32] J. W. Hsieh, H. Y. Liao, and K. C. Fan, ‘‘A new approach for edge detection using wavelet transforms,’’ Asian Conference on Computer Vision, Osaka, Nov. 23–25, pp. 520–525, 1993.

[33] J. S. Lee, Y. N. Sun, C. H. Chen, and C. T. Tsai, Wavelet based corner detection, Pattern Recognit. 26, No. 6, 1993, 853–865.

[34] J. Canny, “A computational approach to edge detection,” IEEE Trans. Putt. Anal. Machine Intell., vol. PAMI-X, pp. 679-698, 1986.

[35] J. F. Canny, ‘‘A computational approach to edge detection,’’ IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-8, no. 6, pp. 679–697, Nov. 1986.

[36] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-square fitting of two 3-D point sets,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-9, no. 5. Pp. 698-700, Sept. 1987.

[37] J. W. Hsieh, H. Y. M. Liao, K. C. Fan, M. T. Ko and Y. P. Hung, “Image Registration Using a New Edge-Based Approach,” Computer Vision and Image Understanding, Vol. 67, No. 2, pp. 112-130, August 1997.

[38] S. Umeyama, ‘‘Least-squares estimation of transformation parameters between two point patterns,’’ IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-13, no. 4, pp. 376–380, April 1991.

[39] K. S. Arun, T. S. Huang, and S. D. Blostein, ‘‘Least-square fitting of two 3-D point sets,’’ IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-9, no. 5, pp. 698–700, Sept. 1987.

[40] W. H. Press et al., Numerical Recipes in C: the Art of Scientific Computing, 2nd ed., Cambridge University Press, Cambridge, 1992.

[41] S. Mann and R. W. Picard, “Virtual Bellows: Constructing High-Quality Images for Video,” Proc. First ICIP, vol. 1, IEEE CS Press, 1994, pp. 363-367.