NOVEL PIXEL-LEVEL AND SUBPIXEL-LEVEL REGISTRATION ALGORITHMS FOR MULTI-MODAL IMAGERY DATA

By

Mohamed Ibrahim Elbakary

A Dissertation submitted to the faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In partial fulfillment of the requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2005

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Mohamed Ibrahim Ahmed Elbakary

entitled Novel Pixel-Level and Subpixel-Level Registration Algorithms for Multi-Modal Imagery Data

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

_____ Date: September 2, 2005
Dr. Malur K. Sundareshan

_____ Date: September 2, 2005
Dr. Robert Schowengerdt

_____ Date: September 2, 2005
Dr. Hal S. Tharp

_____ Date: September 2, 2005

_____ Date: September 2, 2005

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: September 2, 2005
Dissertation Director: Dr. Malur K. Sundareshan

STATEMENT BY THE AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at University of Arizona and is deposited in the University library to be made available to borrowers under the rules of library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by her head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Mohamed Ibrahim Elbakary

ACKNOWLEDGEMENTS

It is strange odd that the person who demanded most from me is also the person who gave me the most encouragement. It is with his guidance and his encouragement that I'm able to complete this dissertation. As such, I wish to express my utmost gratitude to my advisor, Dr. Malur Sundareshan, without him this work not be completed.

I also wish to acknowledge all other excellent professors who had taught me various fields of electrical engineering.

I am indebted to Tami Whelan, whose incredible efficiency does make a difference at ECE department at the University of Arizona.

I want to thank my parent for dedicating their life for us and giving me the chance to complete my graduate education.

Finally, I am grateful to my wife, Dr. Aza Shetewy, for her patience in taking care of our children during pursuing this research.

To my mother and my father,

For their lasting parenthood…

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

LIST OF FIGURES

Figure

LIST OF FIGURES - Continued

LIST OF FIGURES - Continued

LIST OF TABLES

ABSTRACT

Image registration is an important pre-processing operation to be performed before many image exploitation and processing functions such as data fusion, super-resolution frame reconstruction, change detection, image mosaicing, etc. Given two image frames, obtained from the same sensor or from different sensors, the registration problem involves determining the transformation that most nearly maps (or aligns) one image frame into the other. Due to the tremendous practical importance of this problem in several applications, both military (surveillance, tracking, missile guidance, etc.) and non-military (medical imaging, HDTV, homeland security, etc.), many sophisticated image processing algorithms are presently being developed which provide varying degrees of registration accuracy and robustness to scene characteristics. Typically, image registration requires intensive computational effort and the developed techniques are scene dependent. Furthermore, the problems of multimodal image registration (i.e. problem of registering images acquired from dissimilar sensors) and sub-pixel image registration (i.e. registering two images at sub-pixel accuracy) are highly challenging and no satisfactory solutions exist.

This dissertation introduces novel techniques to solve the image registration problem both at the pixel-level and at the sub-pixel level. For pixel-level registration, a procedure is offered that enjoys the advantages that it is not scene dependent and provides the same level of accuracy for registering images acquired from different types of sensors. The new technique is based on obtaining the local frequency content of an image and using this local frequency representation to extract control points for establishing

correspondence. To extract the local frequency representation of an image, a computationally efficient scheme based on minimizing the latency of a Gabor filter bank by exploiting certain biological considerations is presented. The dissertation also introduces an extension of using local frequency to solve the sub-pixel image registration problem. The new algorithm is based on using the scaled local frequency representation of the images to be registered, with computationally inexpensive scaling of the local frequency of the images prior to correlation matching. Finally, this dissertation provides a novel approach to solve the problem of multi-modal image registration. The principal idea behind this approach is to employ Computer Aided Design (CAD) models of man-made objects in the scene to permit extraction of regions-of-interest (ROI) whose local frequency representations are computed for extraction of stable matching points. Detailed performance evaluation results from an extensive set of experiments using diverse types of images are presented to highlight the strong points of the proposed registration algorithms.

CHAPTER 1

Introduction

1.1 Sensor-based Systems and The Need for Image Registration

The past decade has witnessed a revolution in the development of remote sensors for generating images. Recent developments in infrared technology and sensors operating at microwave and optical wavelengths are providing detailed and accurate representations of the scenes previously not possible. Using more than one sensor can result in significant improvements. Consider the following example of a moving object, such as an aircraft, that is observed by a pulsed radar and a forward-looking infrared (FLIR) imaging sensor. The radar can accurately determine the aircraft's range but has a limited ability to determine the angular direction of the aircraft. By contrast, the infrared imaging sensor can determine the angular direction but can't measure the range. The combination of the two sensors can provide a more accurate determination of the location than could be obtained by either of the two independent sensors. However, before fusion of the data from the sensors, the data should be transferred to the same reference coordinates. In the general case, each sensor provides data in its local reference frame, which could be different from the reference frame of other sensors. Hence, aligning all of the data from the different sensors to the same reference coordinate is obligatory before any processing is performed on the data. The process of aligning the images to one reference coordinate system is called *image registration*. Registration is necessary for many other problems besides integrating information captured from different sensors. It is necessary for finding

changes in images taken at different times or under different conditions, inferring three-dimensional information from 2-D images in computer vision field, and for model-based object recognition, as we will see in Section 1.3.

1.2 Sensors and Images

This section presents a short introduction to sensors and explains how different kinds of sensors produce images with different characteristics and different types of information. A sensor consists of front-end hardware, called a transducer, and a data processor, respectively. The transducer converts the energy entering the aperture into another form of energy from which target and background discrimination information is extracted in the data processor. When the operating frequencies of the sensors are selected from a wide electromagnetic spectrum, the generated data from those sensors are more likely to be independent. As an example, data acquired by a forward looking infrared sensor is independent from the data for the same scene which is generated by a charged couple device camera (CCD), because the operating frequency of the former sensor is in the invisible region from the electromagnetic spectrum while the operating frequency of the CCD camera is in the visible region from the electromagnetic spectrum. In general, according to the application and the task of the system, specific types of sensors are selected and deployed in the scene. A good reference for the different types of sensors and their operating frequencies and functions can be found in [1].

A broad classification of sensors can be made as *active sensors*, *passive sensors*, and *active/passive sensors*. This classification depends on whether the sensor illuminates the

scene or not. If the sensor illuminates the scene and then detects the reflected energy, then it is called an active sensor. If it detects the scene depending on natural ambient conditions, it is called a passive sensor. Some sensors work in both active and passive modes. Some examples of sensors are millimeter-wave camera (MMW) (active sensor), electro-optical sensor (EO) (passive sensor), infrared sensor (IR) (passive sensor), RADAR (active sensor), laser radar (LADAR) (active sensor), forward looking infrared (FLIR) (active sensor), and synthetic aperture radar (SAR) (active sensor).

In addition to the previously discussed sensors, there are other kinds of sensors, which are used in the medical field. Some of them are:

Computed Tomography imaging (CT), CT images are used primarily for structural (anatomy) imaging. With functional imaging of how an organ does its function, it is possible by using intravenous injection of contrast materials to image dye distribution in vessels and organs.

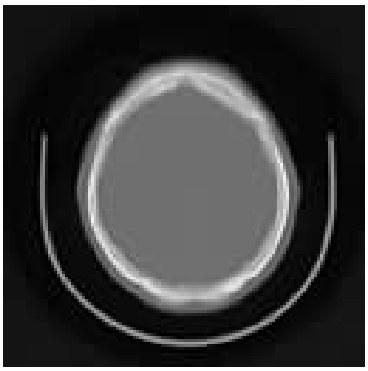
Magnetic Resonance Imaging (MRI). This is a soft tissue imaging modality, providing highly detailed structural images. It can also be used to study the biochemical composition of tissues.

Positron Emission Tomography (PET). PET is a nuclear medicine imaging modality that yields transverse tomographic images of the distribution of positron-emitting radionuclides systematically administered to the subject under study.

Most of the sensors provide 2-D images of the scanned scene and a few provide the range image of the scene. However, it is possible to encode the range information to establish 2-D intensity image for the scene. Fig. 1.1 shows examples of images captured by

(a) EO image

(b) IR image



(c) Computed Tomography (CT) image

Fig. 1.1 Example for images from different types of sensors.

different sensors like EO, IR, and CT. Fig. 1.2 shows an example of mis-alignment between two images, which are captured by the same kind of sensor and for the same scene.

1.3 Applications of Image Registration

Most sensor-based systems encounter the problem of image registration. The applications include military and civilian applications. Image registration can also be used in image fusion applications, change detection applications, or super-resolution imaging applications. The applications can be summarized as follows:

(1) Image Fusion. Image fusion is a fast developing field of research used in a growing number of applications. It is an essential operation for multi-sensor based systems. Image fusion deals with procedures for the combination of data sets from different sources. Multi-sensor integration and fusion can be described as an organized scheme for the use of different information from multiple sensors to assist in the overall understanding of a phenomenon, measure events or combine decisions. Image fusion provides a complete picture of the environment and the events, increases the system robustness, improves system reliability, and increases the accuracy of systems. Applications of image fusion could be in a medical system that uses MRI and CT imaging to integrate structural information about a patient. Another application is a remote sensing system that uses multi-spectral, or different electromagnetic bands like visual band, infrared band, and microwave, to improve its function in classification. In addition, target recognition and surveillance systems are included in this class of applications.

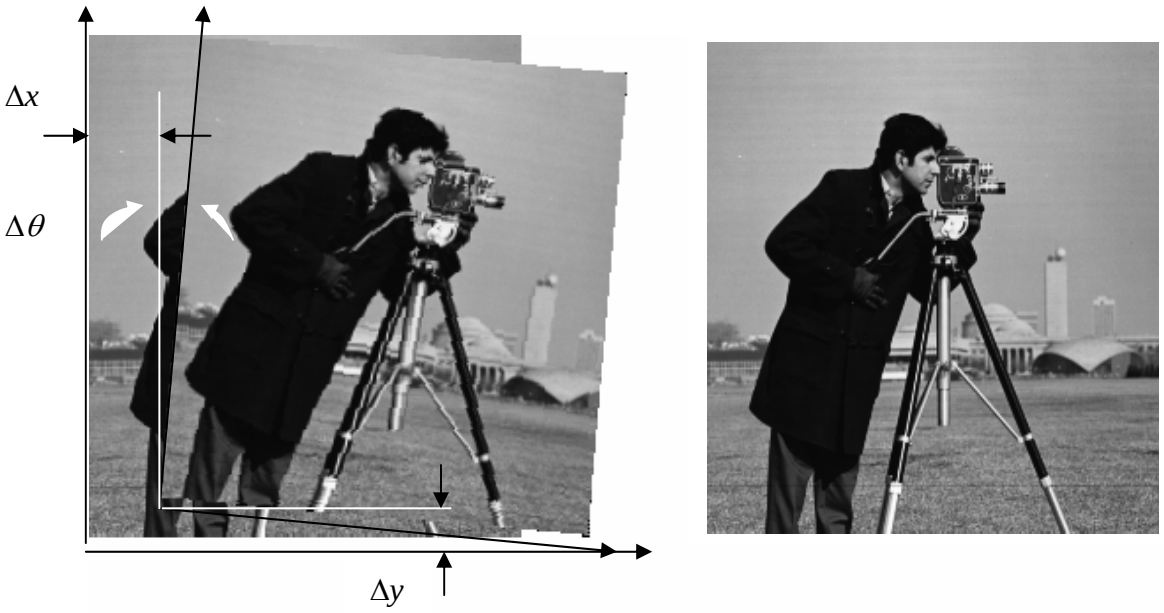


Fig. 1.2 An example of mis-alignment between two images of the same scene

(2) Change Detection. Change detection is an important problem in many applications. In this process, a change in a scene might be detected by registering two images of the same scene taken at different times. Change detection systems can be used to detect the change in the scenes at different times or detect the change in the scene under different conditions. These systems include remote sensing systems for detecting manmade changes, natural resources monitoring, urban growth monitoring and surveillance of nuclear plants.

(3) Super-Resolution Imaging. In all image-based systems, the demand for high-resolution images is gradually increasing. However, in many imaging systems, the quality of image resolution is limited by physical constraints. A digital imaging system yields aliased and under-sampled images in nearly all cases. Digital image processing approaches have been investigated to reconstruct a high-resolution image from aliased low-resolution images. Super-resolution is defined as the processing of an image so as to recover object information from beyond the spatial frequency bandwidth of the optical system that formed the image. Super-resolution is an essential operation for performing better diagnosis in medical applications and is necessary for surveillance and observation systems. Super-resolution is accomplished by using multiple frames for the same scene with sub-pixel shifts between them. The task of registration techniques in super-resolution is to estimate the sub-pixel shifts between the frames of the scene.

1.4 Approaches to Image Registration

Recent research has provided numerous algorithms to solve the problem of image

registration. Any image registration algorithm can be allocated within one of the following classes: *Feature-based methods*, *Similarity measure methods*, or *Search strategy methods*. The feature-based methods use features extracted from the image, such as edges and contours, as control points to solve the correspondence problem and to estimate the registration parameters. In the similarity measure methods, the algorithms are based on using a common information measure, such as the phase of the image, to estimate the registration parameters. The search strategy methods are based on optimizing a search method to minimize a predefined cost function, for instance, the computational cost of the registration.

However, using any one of the algorithms in the literature depends on the specific application. In other words, since there are a huge variety of scene complexities, the algorithm that is designed for one specific scene is highly unlikely to work well for another scene. In addition, registration of multimodal images is still a challenging problem, especially when the operating frequencies of the sensors are different, such as the LADAR and FLIR.

This dissertation tackles the image registration problem for two purposes. First, to find an efficient algorithm which can be applied in different applications. Second, to introduce an algorithm for solving the problem of image registration between images acquired from multiple modalities.

1.5 Contributions of the Dissertation

Recently, local frequency representation of an image has been used to solve the problem of image registration for multimodal medical images. Local frequency is relatively invariant to image illumination variation and has a good localization in the spatial domain. Local Frequency for an image is defined as the spatial derivative of the phase of the analytic form of that image. The analytic form for an image is obtained by convolving the image with a bank of Gabor filters. The size of the filter bank is selected by ad hoc methods to cover the frequency content of the image. The Gabor filter is a complex filter and is parameterized with the following parameters: the spatial frequency of the filter, the spatial orientation of the filter, and the bandwidth of the filter.

To achieve the first goal of this dissertation, we adopted local frequency representation (LFR) for the image to solve the problem of image registration. For this purpose, we have introduced an efficient way to determine the Gabor filter bank to obtain the LFR for an image and developed a simple scheme to estimate the registration parameters.

Registration of images from different modalities is a hard problem because the sensors could be working according to different principles. For example, LADAR is a sensor for measuring the distance between the target and the sensor and gives range image for the scene. A FLIR sensor measures the difference in temperature between the target and its background. On the other hand, a CCD camera generates an image of the scene reflectance in the visible region electromagnetic spectrum. So, finding a common measure, or common features between the images captured from these sensors, is quite

difficult, especially between images from LADAR and other sensors. This dissertation presents a novel algorithm to solve this problem. We used Computer Aided Design (CAD) models to substitute for man-made objects in the image incorporated with LFR. CAD provides 3D models for man-made objects and has been used to recognize objects and to recover their poses by using different sensors such as FLIR, LADAR, and MMW.

The major contributions of this dissertation can be summarized as follows:

1) This dissertation presents a computationally efficient technique to obtain the local frequency representation for an image. The developed technique is based on computing the local frequency representation of an image in an efficient way by using an optimal Gabor filter bank. The optimality of the filter bank is based on some theoretical and practical results. Computing the local frequency representation by the proposed technique is more efficient than the existing techniques in the literature.

2) The dissertation introduces a new algorithm based on the local frequency representation for an image to solve the image registration problem. Local frequency is relatively invariant to illumination changes and has good localization in the spatial domain. These advantages of local frequency representation make it a promising candidate to solve the image registration problem. The new algorithm is designed such that it does not depend on specific features in the images, but depends on the local frequency representation of the image regardless of the complexity of the scene in the image. The proposed algorithm is based on using the local frequency representation for

the image to extract control points from the image. Then, the control points are employed to solve the correspondence problem between the images and to estimate the transformation parameters. The proposed scheme is simple and can be applied with the same accuracy in different applications, which may have diverse images that vary from complex highly varying gray-scale images, to simple low variable gray-scale images.

3) This dissertation provides a novel technique to solve the problem of registration between images from different modalities. The proposed technique is based on avoiding the inconsistent information in the features between the images and is based on using the consistent information only. The algorithm uses man-made objects in the scenes as consistent information and as clues for the registration. Then, man-made objects are replaced by their image representation of Computer aided design (CAD) models after recovering their poses. Then, we incorporate the local frequency representation with image representation of CAD to solve the registration problem in a region-of-interest-to-region-of-interest (ROI-to-ROI) mapping fashion. The conducted experiments in this dissertation show the importance of multimodal image registration for image fusion applications.

4) Finally, the dissertation presents an extension for using the local frequency representation to solve the problem of sub-pixel image registration. Many experiments are conducted to compare the accuracy of the proposed technique to that of existing techniques to solve the problem of sub-pixel image registration. Also, the conducted

experiments include an implementation for a super-resolution imaging technique to show the importance of sub-pixel estimation.

1.6 Organization of The Dissertation

Chapter Two presents some background information for the problem of image registration and provides a literature survey. The third chapter presents the concepts of local frequency and the necessary background. In addition, it also presents the new scheme to obtain the local frequency for an image and the new algorithm to solve the image registration problem using the local frequency representations of images to be registered. Chapter Four gives the results of extension of the local frequency representation to solve the problem of sub-pixel registration and the conducted experiments for super-resolution imaging. Chapter Five presents the new technique for solving the problem of image registration between different modalities using CAD models together with local frequency representations. Chapter Six presents the conclusions and offers some future directions.

CHAPTER 2

Background on Image Registration and Literature Survey

2.1 An Overview of Image Registration

The importance of the problem of image registration has grown recently, when the demand on using sensor-based systems has increased [2-5]. The demand on sensors has increased due to the technological advancements in the manufacturing of sensors and the development in other fields like communications and VLSI technology. Image registration problem has many applications. The requirements of the application determine the type of the problem. Examples include the following.

Larger view. The registration problem arises here between images of the same scene that are acquired from different viewpoints to gain a larger 2-D view for that scene. Examples include mosaicing of images of a surveyed area and shape recovery from stereo images in computer vision.

Scene changes evaluation. The problem arises between images of the same scene with the same viewpoint but the images are acquired at different times or under different conditions. The aim in this kind of application is determining the changes in the scene. Examples include landscape planning, motion tracking, security monitoring, and monitoring tumor evolution.

Information integration. This kind of application includes images which are acquired by different sensors with the same viewpoint to a scene. The goal is gathering different information about the scene by fusing the images acquired by different sensors. Fusing

the images gives one complete detailed representation of the scene. The registration is a mandatory step before implementing the fusion. Examples include fusion of remotely sensed images of different spatial resolution, fusion of RADAR and EO images, and fusion of MRI for anatomical body structure and PET for functional and metabolic body activities.

Image localization. This application includes allocating an image within a bigger image of a scene or registering an image with a computer-generated model for the scene. The aim is comparing the images. Examples include registration of images into maps, matching target templates with real-time images, and comparison of the patient's image with digital anatomical atlases.

The first step in solving the problem of image registration, is determining the type of distortion between the images and the corresponding spatial transformation model of that distortion [6-14]. Then, the second step is determining the method of solving for the parameters of the transformation model using only the information in the two images. So, to solve the problem of image registration, one has to address two sub-problems:

First: Identify the types of distortion between the images and the mapping models, which are known as the *spatial transformations*.

Second: Extract relevant information in the images, which is useful for solving the problem of image registration.

The following two sections present the details of the two sub-problems.

2.1.1 Spatial Transformations

The core characteristic of any image registration technique is the type of the mapping or the spatial transformation used to properly align the two images. Many types of variations might exist between two images; however, the employed class of spatial transformation will remove only the modeled distortion between those images. If we define the images as two 2-D arrays denoted by I_1 and I_2, then the mapping between the two images can be expressed as

$$I_1(m,n) = \lambda(I_2(f(l,k))), \tag{2.1}$$

where f is the transformation that maps the 2D spatial-coordinates l and k to new spatial coordinates m and n, $(m,n) = f(l,k)$. λ is a 1D intensity transformation function [6]. The most common transformations are affine, projective, perspective, and global polynomial transformations.

2.1.1.1 Image Variations

Before exploring the methods of spatial transformation, we have to introduce the possible distortions and variations between images. In general, the distortions between images can be classified into two main classes depending on the ability of image registration techniques to remove the distortion. The first class is called corrected distortion and the second class is called uncorrected distortion.

Corrected distortions. These include all the variations which cause mis-alignment between images and can be removed by using image registration techniques. These types of distortions between images may be due to a change in the sensor viewpoint, changes in the subjects' positions, and other undesirable changes in the scene or sensor geometries. In other words, these types of distortions arise from differences in the way or the circumstances under which the images are acquired and are usually spatial and can be modeled by appropriate transformations.

Uncorrected distortions. These types of distortions arise from lighting and atmospheric variations and cannot be corrected by using image registration techniques alone.

There exist other types of variations between images, which we are interested in after solving the registration problem. These variations include, but are not limited to, changes in the scene, such as object movements, growths or deformations, and differences in sensor measurements when using sensors with varying sensitivities, or using sensors that measure different properties of the scene.

All the variations can be classified in different ways. They can be either static / dynamic, internal / external, or geometric / photometric.

Static variations do not change for all images and can be corrected by calibration techniques for the sensors such as radial distortions.

Dynamic variations include changes that arise from the change in the sensor, or from varied lighting conditions and atmospheric conditions.

Internal variations arise from the sensors themselves. Typically, internal geometric distortions are size, skew, and centering. Other types of internal distortions are

photometric. These types of distortions affect intensity values in the images. Typically, photometric distortions are caused by detector gain variations, lens distortions, sensor imperfections and sensor-introduced filtering. The effect on intensity values from either internal or external distortions is either of interest or is difficult to remove and does not cause mis-alignment between images.

External variations arise from changing scene characteristics and changing of the operating conditions of the sensors. These changes include either the viewing angles of the sensors or scene changes due to movement or atmospheric conditions.

Geometric distortions play the most critical role in the registration. They include the geometric distortions such as scaling, rotation, and translation.

Photometric distortions do not cause mis-alignment between images but only affect the intensity values in the images.

In general, sufficient information about the source of distortion is important, and it helps to model the distortion. For example, if the scene is a flat plane and if the two images of the scene differ only in their viewing geometries, and the relative difference is known, then an appropriate sequence of Cartesian transformations such as translation, rotation, and scaling can be found to align the two images.

2.1.1.2 Affine Transformation

The most commonly used registration transformation is the affine transformation, which is sufficient to match images of a planar scene taken from the same viewing angle, but there exists a movement of the camera and there may be a rotation of the camera

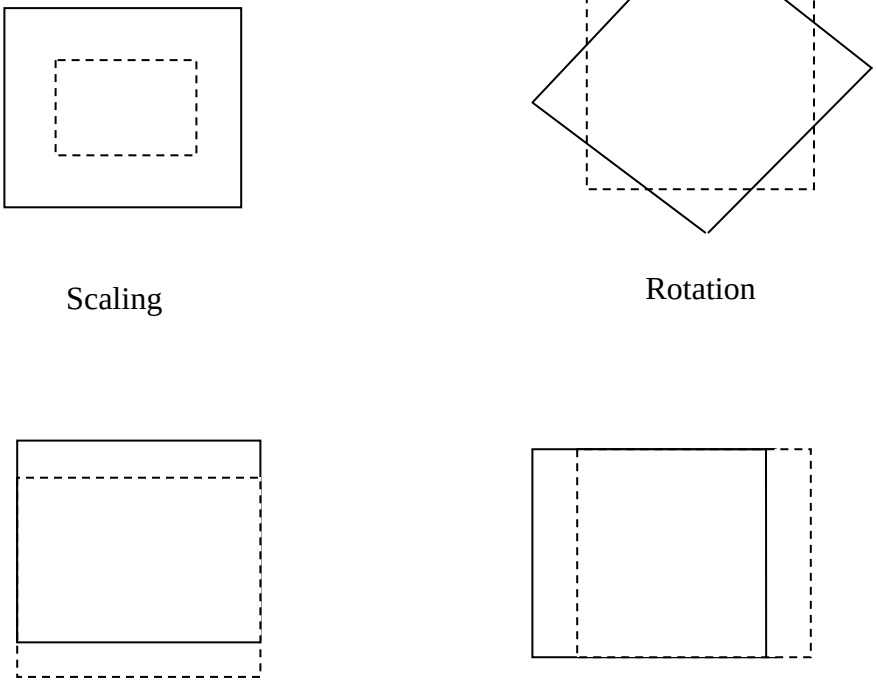


Fig. 2.1: The common elementary digital image transformations.

around its optical axis. This kind of transformation is composed of Cartesian operations of translation, scaling, and rotation. This type of transformation is rigid since the overall geometric relations between points remain unchanged. For example, a triangle in one of the images will be mapped into a similar triangle in the second image. An affine transformation typically has four parameters t_x, t_y, s, θ, where s is a scale factor, θ is the rotation angle, and t_x, t_y are translations in x and y directions, respectively [6,11]. Fig. 2.1 shows the common elementary digital image transformations. Assume a point P_1 with coordinates (x_1, y_1) is in one image and another point P_2 with coordinates (x_2, y_2) is in the second image; then the relation between the two points by using the affine transformation is expressed by the following equation,

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = s \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}. \tag{2.2}$$

The affine transformation preserves the angles and the lengths after the registration. The scaling factor s allows changes in the lengths, but these lengths have the same directions in the original image. In addition, one can easily recognize from Eq. 2.2 that an affine transformation is linear without the addition of the translation vector. A more computationally efficient way of describing the affine transformation matrix is the homogeneous coordinate representation, which is

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/s_x \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & 1/s_y \cos(\theta) & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} . \tag{2.3}$$

If the distortion is pure translation, Eq. 2.3 takes the following form

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}, \tag{2.4}$$

if the distortion is scaling only, then Eq. 2.3 takes the following form

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/s_x & 0 & 0 \\ 0 & 1/s_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} , \tag{2.5}$$

but if the distortion is rotation only, then Eq. 2.3 takes the following form

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}. \tag{2.6}$$

There exists a general 2D affine transformation and it has the following form,

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \quad . \tag{2.7}$$

The general affine transformation does not preserve the angles and the lengths, but parallel lines still remain parallel. The general affine transformation can account for spatial distortion such as shearing and aspect ratio changes. By applying a sequence of rigid-body transformations, the affine transformation describes the cumulative distortions. In this dissertation, we have emphasized the affine transformation because it is the most common spatial transformation and has many applications. However, there exist other spatial transformations, as we will see in the next section.

2.1.1.3 Perspective Transformation

A perspective transformation accounts for the distortion that occurs when a 3-D scene is projected through an optical system to form a 2-D image. In other words, the perspective distortions occur in registration of a 3-D scene to its corresponding 2-D image (3-D/2-D registration). The perspective distortions cause the imagery to appear smaller the farther it is from the camera, and the more it is inclined away from the camera. The perspective transformation is defined as follows. Assume a point (x_o, y_o, z_o) in a 3-D scene, then the corresponding point in the 2-D image of that scene (x, y) is defined by

$$x = \frac{-fx_o}{z_o - f} \ , \ y = \frac{-fy_o}{z_o - f} \ , \tag{2.8}$$

where f is the focal length of the lens of the optical system which is used in generating the image for the scene.

2.1.1.4 Polynomial Transformation

In many registration problems, the precise form of mapping function is unknown, and a general transformation is needed. In such cases, a bi-variate polynomial transformation is applied. This kind of transformation is expressed by

$$x_2 = \sum_{i=0}^{n} \sum_{j=0}^{i} a_{ij} x_1^i y_1^{j-i} \ , \quad y_2 = \sum_{i=0}^{n} \sum_{j=0}^{i} b_{ij} x_1^i y_1^{j-i} \tag{2.9}$$

where (x_1, y_1) are the coordinates in the reference image, (x_2, y_2) are the coordinates in the image to be transformed, and (a_{ij}, b_{ij}) are the coefficients to be determined. n is the order of the polynomial and its value depends on the trade-off between the accuracy and the needed speed for the registration problem under investigation. In many applications, a second order or a third order polynomial is sufficient and the coefficients are determined by employing a system of N equations determined by the mapping of N control points. Also, we have to say here that polynomial transformation is only useful to account for low-frequency distortions because of their unpredictable behavior when the degree of the polynomial is high.

2.1.2 Registration Methods

The second sub-problem to be addressed in solving the problem of image registration is selecting the registration technique. Image registration methods can be classified mainly into similarity measure methods, and feature-based methods. However, there exist other methods such as relaxation methods, pyramid methods, and optimization methods are considered as search strategy methods.

Similarity measure methods avoid extraction of the features, but rather they go to the matching step directly. These methods do not need a feature extractor or looking for the suitable feature for the application. The Similarity based methods are called direct methods because they use the raw images without any preprocessing. So, the first step in solving image registration is omitted when compared with the feature-based methods. The methods of similarity measure include cross-correlation methods, normalized cross-correlation methods, Fourier-based methods and mutual information methods. The cross-correlation methods search for the best correlation between the images to determine the transformation parameters. It gives a measure of the degree of similarity between the image and a template. Fourier-based methods use the Fourier Shift theorem to determine the difference in the translation, the rotation, and the scaling. Mutual information methods are based on information theory and they depend on using the information in both images to estimate the registration parameters.

Feature-based methods depend on extracting features from the images and then using these features to estimate the registration parameters. The majority of feature-based methods consist of the following steps:

Feature detection. In this step, distinguished regions (fields, lakes, etc.), distinguished edges (rivers, bridges, roads, building boundary, etc.), and distinguished points (roads intersections, corners, salient features in building and objects, etc.) are automatically or manually detected for further processing. The detected features are represented by distinctive points such as end points of lines and center points of gravity of regions. In this step, we have to decide what kind of features is suitable for the application. The features should be easily detectable, spread over the image, common between the reference image and the distorted image, stable in the time, locally accurate, and insensitive to the assumed distortion between the two images.

Correspondence establishing. Once the features are detected and are represented in the reference image and in the distorted image, the correspondence problem between these features needs to be solved. The spatial relationship between the features in the reference image and in the distorted image can be used for this purpose. In this step, incorrect feature detection due to a poor feature detector can cause problems. The physical similarity between the images is weakened in some cases due to illumination condition changes or the sensitivity of the sensors. So, choosing a feature detector and feature representation are important aspects and these have significant effect on the image registration problem. The matching algorithm in this step should be feature-invariant and should be able to detect the outliers. The features without correspondence in the two images should not influence the performance of the algorithm.

Transform model estimation. The mapping function and its parameters are determined in this step. Information about the source of the distortion can help in determining the

transformation model. If prior information is not available about the source of the degradation, then the mapping function should be flexible and general enough to handle all the expected distortions which might appear between the images. Once the transformation model is determined and the correspondence problem is solved, the extracted features are used to estimate the transformation parameters. The transformation model should not remove all the differences between the images because there may exist types of differences we are looking for in the specific application.

Image transforming. The estimated parameters in the last step are used as parameters for the mapping functions for transforming the image.

Beside the feature based methods and the similarity measure methods, there exist other methods that depend on using optimization methods, relaxation methods, wavelet transform methods and elastic registration methods. Search strategy methods are also used to decide the next transformation in the search space to find the optimal transformation. They speed up many approaches by guiding the search through progressively finer resolutions. These approaches include Linear and Nonlinear Programming, Hierarchical Techniques, Tree and Graph Matching, Decision Sequencing, generalized Hough transform, and Dynamic Programming [6,8,9].

2.1.2.1Similarity Measure Methods

The methods of similarity measure use the intrinsic structure, i.e., the invariance properties of the image, to establish matching between the images and to estimate the

registration parameters. Typical similarity measures are cross-correlation, normalized cross-correlation [15-23], phase correlations [24-30], and mutual information [31-40].

These methods estimate the transformation parameters without attempting to detect the salient features or to analyze the structure in the images. In these methods, the entire image or a predefined size window is used to estimate the correspondence between the images. The window-based methods, such as cross-correlation method, have limitations. One such limitation is that they are suitable only if the mis-registration between the considered images is a translation only, but if the mis-registration is more than translation, the window in the reference image will not be able to correlate and to cover the same part from the scene in the distorted image. Some authors changed the shape of the windows, but the proposed shapes do not work in all the transformations [9]. Another limitation for window-based methods is the so-called “remarkableness” of the window content, where windows with smooth areas and without salient features are likely to match incorrectly other smooth windows. These methods are sensitive to intensity changes due to varying illumination, noise or different sensors.

Similar to the cross-correlation methods, the methods of sequential similarity detection algorithm are also based on windows. The methods of sequential similarity detection are less accurate than other window-based methods, but they are faster. The sequential similarity detection algorithm sequentially searches for the transformation between the reference image and the distorted image. It uses the accumulation of the absolute difference, as measuring distance between the windows in the reference and the distorted images. The sum of squared differences is another measure, which is proposed

by different authors [41]. The correlation ratio based method is used to register multimodal images [42]. This approach can handle the difference in the intensities between images due to the usage of different types of sensors. There exist two main disadvantages for cross-correlation methods. These disadvantages are the flatness of the similarity measure maxima and the computation cost. Different authors tried to improve the cross-correlation method by sharpening the maxima by introducing preprocessing on the image, such as image filtering or using the edges in the image [43].

The phase correlation method is based on The Fourier Shift Theorem. First, it is proposed for solving the problem of image registration if the distortion is translation only, but then other authors introduced an extension to the distortion of rotation and scaling by using the Fourier-Mellin transform. These algorithms are used in remote sensing and in the medical applications, as well as, in 3D applications. The method is applied to multimodal images by using edge representation instead of using the original images. Recently, Foroosh *et al*. [125] extended the phase correlation method to sub-pixel estimation. Chapter 4 will present this method in detail as the sub-pixel registration algorithm proposed in this dissertation employs a variation of this technique.

The mutual information (MI) method has been introduced recently by Viola *et al*[31]. It is the leading technique to solve the problem of multimodal medical image registration. Later, authors combined MI with other approaches such as hierarchical search strategy and Parzen's window to find the maximum of MI [44]. Other authors introduced MI to different kinds of images such as PET and registering 3D objects to their images.

Gradient descent optimization and Gauss-Newton numerical minimization are used in the searching for the maxima for MI.

Choosing the type of similarity measure is important because it determines how the image registration transformation is estimated. For example, estimated transformation parameters, using cross-correlation, are found at the peak value of the correlation values. The direct methods or the similarity measure methods are computationally expensive and typically they need a good initial guess to ensure correct convergence. Therefore, these approaches do not always guarantee proper convergence and may converge to local minima for large values of the transformation parameters.

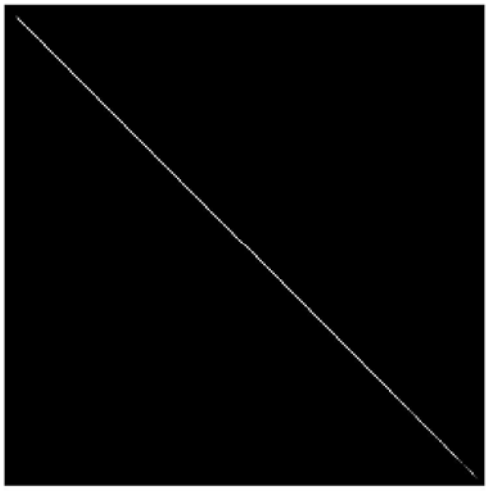
Mutual Information

Mutual information is an example of similarity measure methods to solve the image registration problem between two images. The mutual information method is based on the information theory, and it is a measure of the statistical correlation between the two images. The method uses directly the raw data and does not require an invariant image representation such as features. This method is based on the assumption that the statistical correlation between the two images is global, which is often violated [45]. Moreover, the statistical correlation between raw multi-sensor images tends to decrease with the reduction in the spatial resolution. In general, the methods of similarity measure suffer from computational complexity and, in addition, may employ an optimization algorithm, which can be trapped in local minimum. When two images are perfectly aligned, their

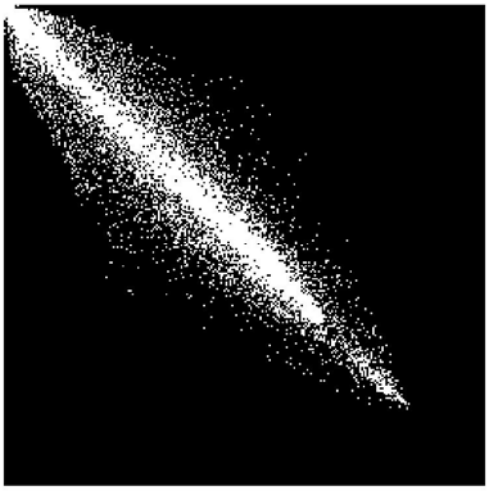
mutual information attains a maximum. Given two images A and B, their mutual information, $I(A,B)$, is defined by

$$I(A,B) = \sum_{a,b} P_{A,B}(a,b) \log \frac{P_{A,B}(a,b)}{P_A(a) \bullet P_B(b)} \quad , \tag{2.10}$$

where $P_{A,B}(a,b)$ is the joint probability density and is approximated by the normalized joint histogram of the two images, while the true joint probability density is often unknown. $P_A(a)$ is the probability density associated with image A and it is approximated by the normalized histogram of that image. Similarly, $P_B(b)$ is the probability density associated with the image B and is approximated by the normalized histogram of that image. If the two images are perfectly aligned and have the same pixel intensity mapping, then the joint histogram should be a straight diagonal line. Otherwise, the straight line is replaced by a figure corresponding to the relationship between pixel intensities. On the other hand, if the images are not registered, the joint histogram will have entries distributed over most of the area. Fig. 2.2 shows the normalized joint histogram between two “Cameraman” images.



(a)



(b)

Fig.2.2 The normalized joint histogram between two standard “Cameraman” images. (a) The two images are perfectly registered. (b) There exists a mis-registration between the two images.

In Fig. 2.2(a) the normalized joint histogram is a straight diagonal line because the two images are registered perfectly. Fig. 2.2(b) shows the normalized joint histogram for the same two images, but with a mis-registration between them. Obviously, the normalized joint histogram is no longer a straight line.

2.1.2.2 Feature-Based Methods

Feature-based methods do not use the raw images directly, but require detecting and extracting features from the images [46-56]. These are based on extraction of salient features in the images, such as regions (lakes, fields), lines (roads, coastlines, region boundaries) and/or points (line intersection, points of high curvatures) [57-74]. The features should be efficiently detectable in both images and spread over the images. Also, they should have fixed positions in the images. The number of detected features should be high enough, regardless of the captured scene, radiometric conditions, and presence of additive noise. Due to the nature of edge and line features, feature-based methods are suitable for situations when illumination changes exist and multi-sensor image registration is required.

The first step in registering two images is to decide what kind of features in the feature space is suitable for matching. For example, remotely sensed images usually contain roads, rivers, coastlines, fields, and lakes, but landscape images contain objects such as cars, buildings, furniture, and people.

Region features such as lakes, fields, and urban areas are segmented and then they are represented by their centers of gravity, which are invariant to rotation and scaling.

Accuracy of segmentation methods for these features significantly influences the resulting accuracy of the registration. For this reason, some authors have tried to estimate the registration parameters during the segmentation process and they used the estimated transformation parameters to refine the segmentation process [62].

Line features such as roads, coastal lines, object contours, and rivers are represented by end points or middle points. The edge detection technique significantly influences the results of image registration. Methods such as Canny edge detector and Laplacian of Gaussian are used for edge detection, however ridge detection needs different types of detectors, the so-called ridge detection operators [75]

Point features such as road crossings, lines intersections, points of high curvature, and corners are used as control points to estimate the registration parameters. Among these point features, corners took a lot of attention and many authors proposed different corner detectors [69,70,73,76], because corners are well perceived by a human observer.

The feature space is a fundamental aspect of image registration as it is for computer vision tasks, and by choosing suitable features, it is possible to improve significantly the registration performance. However, feature detectors should be image invariant and accurate.

Deciding on the features for image registration depends on:

- Which properties of the image will most likely establish matches between the images? For image registration, structures like edges, corners, contours, and crossing points are more useful than texture.

- The computational cost including the preprocessing steps for feature extraction process.

Features can be found in each image independently in a preprocessing step and it is often possible to choose a feature space, which will eliminate uncorrected variations, which might otherwise make matches unreliable.

Feature based methods give accurate solutions, but obtaining correct matches between control points in images acquired by different types of sensors is a hard problem (for example, a Forward looking infrared sensor and a CCD camera).

Registration by Point Matching

This technique is an example of a method to solve image registration problems by using features in the image. The algorithm proposed by Ton and Jain [46] decomposes the images into homogenous regions with specific qualities. Then, they used the features as candidates for matching. They used different methods for region segmentation by utilizing similarity of pixel gray levels to define a region. Once the regions are identified, the center of gravity (centroid) of each region is used as a control point. The center of gravity of a region R is defined by (x_c, y_c) , where $x_c = m_{10} / m_{00}$, $y_c = m_{01} / m_{00}$, and

$$m_{pq} = \sum_i \sum_j i^p j^q R(i, j), \tag{2.11}$$

where i, j are the indices of region R. After the control points are found, they are input

to a correspondence matching algorithm used to estimate the registration parameters. Ton and Jain conducted many experiments to test the proposed scheme.

Discussion

The image registration community has provided numerous techniques to solve the problem [77-80]. Due to the huge diversity of images to be registered and due to the different types of deformations between images, it is not possible to design a universal method that can solve every registration problem. Every method has to take into account the geometric transformation, the type of the data, the corruption noise, and the required accuracy. Some algorithm developers have used feature spaces such as edge maps, corners, contours and points. However, each method is designed to handle a specific type of feature that exists in the structure and the complexity of the scene. So, the algorithm that is designed to register images using contours in the structure of the image cannot register images whose structures includes point intersection only. In addition to the problem of the dependency on the structure of the image to solve the problem, extraction of features requires tuning and optimality for feature extraction operators.

Other algorithms have been introduced; they do not depend on the feature space but they depend on the similarity measure. The similarity measure methods avoid the step of feature detection and extraction and go directly to the matching step. However, these methods suffer from computational complexity and require good guessing for the initial values of transformation parameters, otherwise they can be trapped into a local maximum

or minimum. In addition, they need optimization algorithms to reduce the searching space. These methods might not work if the distortion between the images is large.

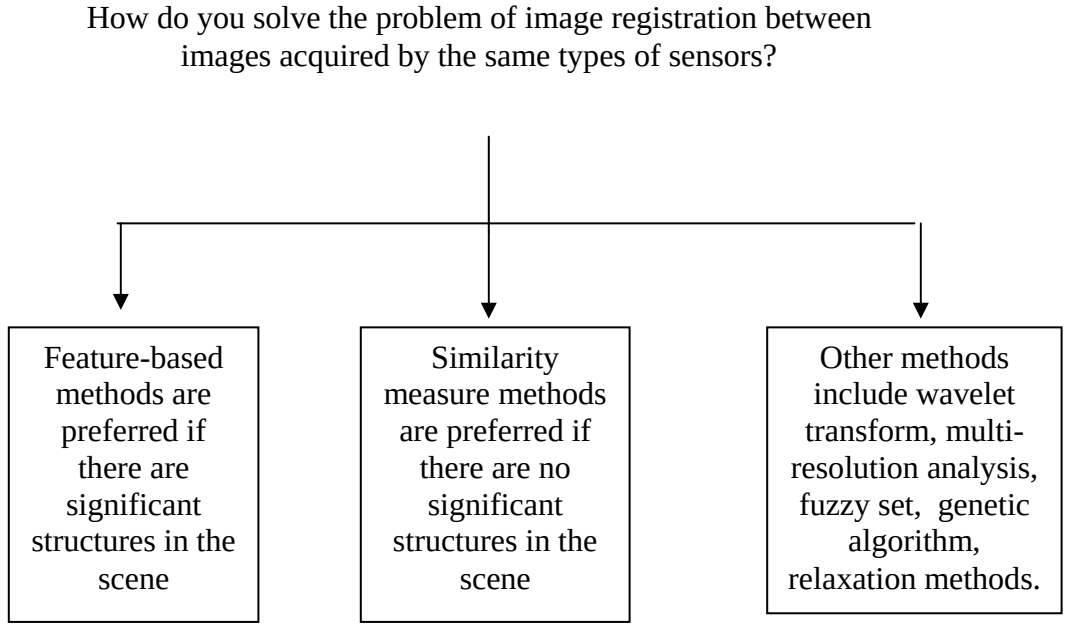
A few researchers have used a wavelet transform to solve the problem of image registration [81-83] and others have used multi-resolution analysis to align the images [84]. Other authors have combined the wavelet decomposition with the pyramid approach [85]. The methods differ in the used wavelet and the set of coefficients for achieving the correspondence between the images. The relaxation method is used by many authors; however the first work was done by Rosenfeld *et. al*.[86]. Fuzzy sets and genetic algorithms are also employed in [87].

Regarding the problem of multimodal image registration, it is still a challenging problem. Few authors have addressed this problem and their results are dependent on the application and on the employed type of sensors [88-92]. Multimodal image registration presents more difficulties than registration of images acquired by the same type of sensor. Due to the lack of consistency of information between multimodal images, finding common features is a hard problem. The authors in [90] used the edges presented in the images as a source of consistent information to accomplish the registration process, which is not the common case for IR images and EO images. In addition to the edges, contours are also employed [92]. In the medical field, the mutual information method is commonly used to register multimodal images. The multimodal image registration based on segmentation is also described by [89], where the authors fused IR and EO images based on segmented parts from the image but their algorithm required common features

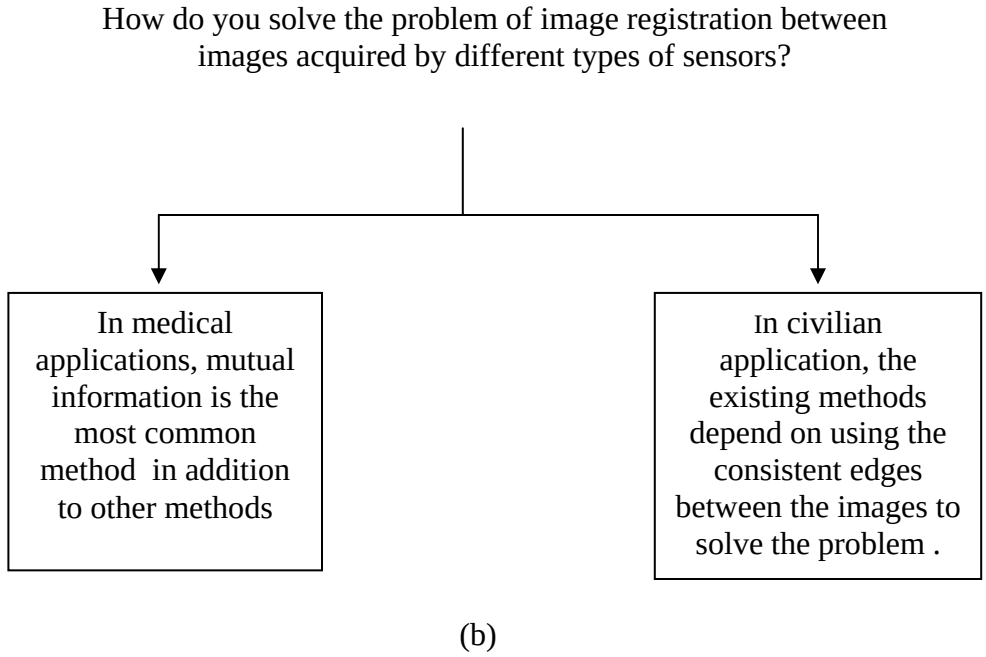
between the images, which is not always possible. The same constraint of common edges between images is required in [91]. The authors registered IR image with EO image, but they required consistent edges between the two images.

Apparently, image registration for images acquired by the same sensor is still an open problem and most of the existing methods suffer from computational complexity, and are dependent on the structure and the complexity of the scene (the application). Also, for registration of images acquired by different types of sensors, i.e. multimodal image registration, all the methods in the literature register IR image with EO image or SAR image with EO image and the methods require common edges between the images which is not the general case. So, while these methods are suitable for specific scenes, which have common edges or segments, this is not the general case. Figure 3 shows a schematic overview for literature survey of image registration.

Therefore, improved registration performance in terms of less computational complexity, better accuracy, and consistency in broad applications are the goals of this dissertation. In addition, the goal of the dissertation is to introduce an algorithm to solve the problem of multimodal image registration. The dissertation will also report many experiments to evaluate the proposed algorithms for applications of image fusion and super-resolution imaging.



(a)



(b)

Fig.. 2.3. The schematic diagram for the flow of the literature on the image registration problem. A) Images acquired by the same type of sensor. B) Mutimodal images.

2.2 Evaluation of Image Registration Methods

The results of image registration algorithms usually need evaluation. The purpose of this evaluation is to estimate the accuracy of the results. Accuracy evaluation is highly desirable regardless of the registration technique, the application area, and the complexity of the scene. The sources of error in the image registration process vary from the sources of error in the detection of features and their representation to match the features and solve the correspondence problem. Also, distortion in the image can cause errors through distortion of the features in the images.

Sources of errors

The first source of error in image registration is called *localization error*. Localization error arises from the displacement of the control point coordinates due to a lack in accuracy of their detection. In controlled experiments, the error in the estimation of the coordinate of the control points can be detected by comparison using computer simulation and ground truth; however, in the general case the source of this error cannot be detected directly. Localization error can be reduced using an optimal feature detector for the given image and the scene; however using an optimal feature detector can reduce the number of control points, which is not desirable in some cases. The second source of error is *matching error*. It is caused by false matching. Matching error can lead to the failure of the image registration process and should be avoided. Robust matching algorithms can ensure correct matching, however a consistency check can detect false matches. Consistent matching can be employed by using two different algorithms to

match the points between the reference image and the distorted image, and the pairs that are found by both algorithms are considered to be valid corresponding pairs, the other control points are excluded from the registration process. The cross-validation approach can also be used to achieve the same aim. The registration parameters that are estimated by the control points are used to compute the translation, rotation and scaling mapping for each pair of control points. Any pair of control points that exceeds a threshold for the distortion values are excluded from the registration process, and they are considered to be false pairs and other control points are accepted as valid control points. The last source of error in the registration process is *alignment*. Alignment error arises due to errors in estimated registration parameters and/or due to the fact that the chosen mapping model may not correspond to the actual distortion between the images. The lack of prior information about the reason for distortion between the images causes the selection of an unsuitable mapping function.

The accuracy of registration can be evaluated in different ways. The simplest measure is computing the mean square error or the root mean square error between the mapped control points and the reference control points. Another similar method is called test point error. This method is based on excluding some control points from the estimation of parameters of the registration, and then these parameters are used to map the excluded points and to compute the mean square error for those points. This method can be applied if we have a sufficient number of test points. Another method to evaluate the registration algorithm is by comparison with another method based on a different principle. The algorithm we use in the comparison is typically called the ‘gold standard

method', which is a method commonly believed to be the best in the specific application or for the given image type. In fact, the gold standard method plays a similar role to the ground truth. Usually, the gold standard method is used in medical applications, but for other applications such as remote sensing and image fusion, where no gold standard method exists, any different method of different nature can be used; however, this method does not guarantee evaluation of registration accuracy. Finally, the oldest method of evaluation can be used, which is visual assessment by a domain expert.

In general, feature-based methods are preferred when the scene in the image has significant structures such as buildings, objects, rivers, and roads, which can be detected and employed to solve the problem of image registration. The features should be visible, stable in time, and invariant to the assumed distortion between the images. On the other hand, direct methods are employed when the scene in the image does not have salient features and the distinctive information between the images is given by gray levels rather than by structures. However, the two images should have similar intensity functions which are either identical or at least statistically dependent. To speed searching in the direct methods, some optimization algorithms are preferred.

2.3 Problem Statement

The scenario of interest in this dissertation is the following. We are given two landscape images (reference image and distorted image) with the same viewing angle for a scene and acquired by the same type of sensors or by different types of sensors.

Unavoidable distortions exist between the two images, such as horizontal translation, vertical translation, rotation in the plane, and scaling. The distortions of interest are modeled by affine transformations. The goals of the dissertation are:

- Developing more computationally efficient and accurate algorithms for solving the problem of image registration.
- Ensuring that the accuracy of the proposed algorithm is consistent in different applications and should not depend on the scene.
- Solving the problem of multimodal image registration without putting constraints on the used images.
- Evaluating the proposed techniques by conducting different experiments.
- Demonstrating experimentally the applicability of image registration algorithms in the image fusion application and the super-resolution imaging.
- Investigating the applicability of the developed technique to sub-pixel image registration.

CHAPTER 3

Local Frequency Estimation

The primary focus in this chapter is to introduce the principles of local frequency estimation and developing an algorithm for the local frequency estimation of a given image. Section 3.1 and Section 3.2 present an introduction to the concept of local frequency and the principles of local frequency estimation as developed in the literature. Section 3.3 introduces a novel computational scheme to estimate the local frequency based on minimizing the latency of a Gabor filter bank. Section 3.4 employs the developed algorithm to solve the image registration problem. Section 3.5 presents the results of the conducted experiments to illustrate the performance of the algorithm.

3.1 Introduction

The concept of frequency of a signal is mathematically well defined. Any stationary signal can be represented as a weighted sum of sine and cosine functions having particular amplitudes, phases and frequencies. However, this is not the case for non-stationary signals and a description as a sum of cosines and sines is not possible. This fact has led to the introduction of the notion of *local or instantaneous frequencies* to allow non-stationary signals to be analyzed in a frequency-like manner, since most physical signals are non-stationary. The instantaneous frequency (IF) of a signal is commonly defined as the rate of change in phase of the corresponding analytic signal. A historical review and introduction of the concept of IF can be found in [93-95].

Recently, local frequency (LF) has been used successfully for multi-modal image registration [96,97]. The advantages of local frequency are as follows [96-98]:

- local frequency is relatively invariant to illumination changes (in other words, local frequency estimate is relatively insensitive to the level of signal energy),
- local frequency can detect the edges and ridges in the image at the same time,
- local frequency has good localization in the spatial domain and can be extracted from a given image using systematic signal processing operations,
- local frequency is relatively insensitive to photometric deformations that result from changes in viewing direction, surface normal and lighting conditions.

These attractive properties make the local frequency a promising candidate for invariant image representation for matching image data sets for image registration purposes. Therefore, we have introduced a computationally efficient scheme to estimate the local frequency of an image and then we have employed this local frequency representation to solve the registration problem as we will show in Section 3.3 and Section 3.4.

3.2. An Overview of Local Frequency Estimation

The local frequency representation for an image is estimated in two steps as follows:

1- Filtering the image with analytic filters tuned to a certain spatial frequency, orientation, and bandwidth.

2- Computing the spatial gradient of the phase of the filtered image.

In order to understand the principles of estimating the local frequency representation, a brief introduction to the principles of the analytic signal/function is presented first.

3.2.1 Analytic Signal

An analytic signal in signal processing terms means that the signal has no negative-frequency component. For a brief introduction to the local frequency representation, we will consider a one-dimensional signal, which can be generalized to higher dimensions readily. Consider a signal s in 1-D. Its corresponding analytic signal is defined to be $s_A = s - is_{H_i}$ where s_{H_i} is the Hilbert transformation of s. The Hilbert transform H_i, $H_i : c_1 \to c_2$, is defined by the integral

$$s_{H_i}(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\xi)}{\xi - x} d\xi \quad . \tag{3.1}$$

The Hilbert transform could be denoted by $s_{H_i} = H_i\{s\}$. Also, it may be noted that the Hilbert transform can be computed quite easily by performing the convolution operation using

$$s_{H_i} = s \otimes \frac{-1}{\pi x}, \tag{3.2}$$

where $\otimes$ denotes convolution. In other words, the Hilbert transform is computed by convolving s with the function $\frac{-1}{\pi x}$.

Example:

Let $s(x) = A\cos(wx)$, A>0; the Hilbert transform of s is $s_{H_i}(x) = -A\sin(wx)$. Then,

$$s_A(x) = A[\cos(wx) + i\sin(wx)] = Ae^{iwx} \tag{3.3}$$

which means that the amplitude A is given by

$$A = |s| = |s_A| = \sqrt{|s(x)|^2 + |s_{H_i}(x)|^2} \quad . \tag{3.4}$$

The transformation of a real signal to its corresponding analytic one can be regarded as the result of convolving the real signal with a complex filter, such as a Gabor filter [99]. The argument of s_A is called the local phase of s , which is defined in the spatial domain. The spatial derivative of the local phase is called *instantaneous or local frequency* [93-97].

To summarize the above, given a real signal in 1-D, its corresponding analytic signal is complex with real part being the original signal itself, and the imaginary part being the Hilbert transform of the real signal. In other words, the corresponding analytical signal of a real signal is obtained by convolving the real signal with a complex Gabor filter. The argument of s_A refers to the local phase of s whose spatial derivative gives the instantaneous frequency or local frequency. For a 2-D image, a filter bank is used to recover all the spatial frequencies in the image. Therefore, a Gabor filter bank is used to obtain the analytic form of the image such that each filter has a unique spatial orientation and unique spatial frequency. Then, the outputs of the filters are fused to construct the local frequency representation of the image. The next section presents the details of a

Gabor filter.

3.2.2 Gabor Filter

Gabor filters are optimal filters that are constrained by a lower limit for joint localization properties in the spatial domain and the spatial frequency domain [99,100]. A Gabor filter is characterized by the orientation, spatial frequency, and spatial kernel of the filter. These characteristics help in extracting various types of information from the image, which may be different in the orientations, the scales and in corresponding spatial frequencies. Further, many authors have pointed out the close relationship between the neurophysiological processing of visual and tactile stimuli and families of Gabor filters [101]. Therefore, it seems to be a natural tool to examine the spatial information. A two-dimensional Gabor function $G(x, y, f_o, \theta, \sigma)$ centered at the origin in the spatial domain implements a bandpass filter centered at f_o in the spatial frequency domain. The Gabor function is defined by

$$G(x, y, f_o, \theta, \sigma) =$$

$$\frac{1}{2\pi\sigma^2}\exp(i2\pi(xf_o\cos(\theta)+yf_o\sin(\theta))-\frac{x^2+y^2}{2\sigma^2}), \tag{3.5}$$

where x and y are the spatial coordinates, the parameter f_o denotes the radial central spatial frequency of the band pass filter, θ is an arbitrary spatial orientation of the filter coordinates in the counterclock-wise direction. It is important to say here that θ in Eq. 3.5 is also the orientation of f_o in the spatial frequency domain. σ is the standard deviation of the Gaussian envelope of the filter. Fig. 3.1 shows a Gabor filter (the real

part) encoded as an intensity image. The size of the image is 128 x 128 pixels. The parameters of the filter are as follows: σ = 10 pixels, $f_o = 8\sqrt{2}$ cycles/image_width, and θ = 45 degree.

The 2-D Fourier Transform of the Gabor function defined in Eq. 3.5 is given by

$$\hat{G}(w_x, w_y, f_o, \theta, \sigma) = \exp(-2\pi^2\sigma^2((w_x - f_o\cos(\theta)) + (w_y - f_o\sin(\theta))), \quad (3.6)$$

which is a Gaussian bandpass filter with a radial central frequency f_o and orientation θ with respect to the w_x-axis in the counterclock-wise direction. Fig. 3.2 shows the fast Fourier transform of the Gabor function with an arbitrarily selected set of parameters, σ = 2 pixels, $f_o = 20\sqrt{2}$ cycles/image_width. (a) θ=45 degree, (b) θ=135 degree.

Local frequency using a Gabor filter bank

Obtaining a local frequency representation for a given image consists of building a filter bank such that the filters are appropriately tuned with the parameters f_o and θ. The parameter σ can be varied to generate a scale space representation and is not considered as a design parameter in the construction of the filter bank. A filter bank is employed to estimate the local frequency representation of an image. Since the image of interest may contain significant details typically covering a wide range of frequencies and a single filter characterized by a specific center frequency f_o and a specific orientation θ is not sufficient to cover the entire frequency range, a bank of filters will be needed for

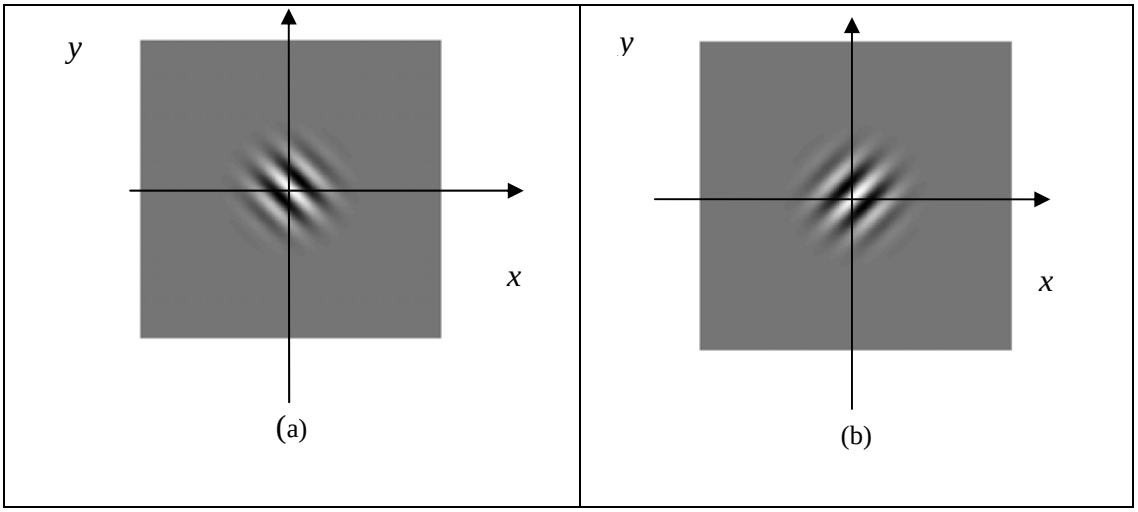


Fig. 3.1. Gabor function (real part) encoded as intensity map images. Light and dark gray colors correspond to positive and negative function values, respectively. σ = 10 pixels, $f_o = 8\sqrt{2}$ cycles/image_width for both images. (a) θ =45 degree , (b) θ =135 degree.

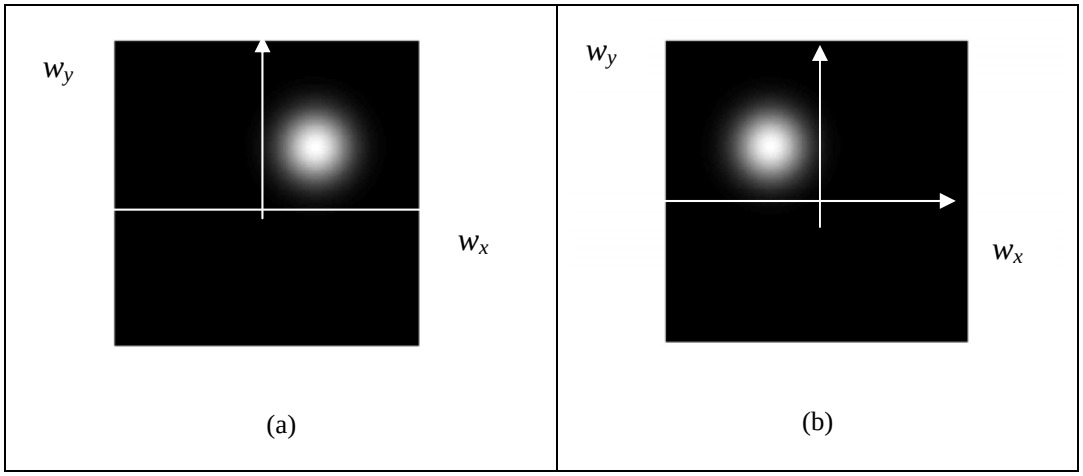


Fig. 3.2 Fourier Transform of Gabor filter function encoded as intensity map images.Higher gray level values correspond to higher function values. Both functions in the spatial domain have σ = 2 pixels, $f_o = 20\sqrt{2}$ cycles/image_width. (a) θ =45 degree , (b) θ =135 degree.

obtaining an accurate local frequency representation.

Typically, an input image $I(x, y)$, $(x, y) \in \Omega$ (Ω is the set of image points), is convolved with a set of a 2-D Gabor functions $G_k(x, y, f_k, \theta_k)$, $(x, y) \in \Omega$ and tuned to a preferred spatial frequency , f_k, and a preferred spatial orientation, θ_k, to obtain the analytic signal $u_k(x, y)$ as follows:

$$r_k(x, y) = \iint_{\Omega} I(\varepsilon, \eta) G_k(x - \varepsilon, y - \eta) d\varepsilon d\eta \tag{3.7}$$

for $k = 1, 2, ..., n$ where n is any integer number. Having obtained the analytic signal, the local frequency can be estimated using two methods. First, the local phase can be computed using the ATAN function and then the gradient of the local phase can be computed to estimate the local frequency. The second method involves deriving a mathematical expression for the local frequency using the analytic function. Let $u_{k+}(x, y)$ and $u_{k-}(x, y)$ be the result of the convolution of an image , I , with the real and the imaginary parts of G_k, respectively. Then, the local phase gradient or the local frequency estimate from each filter is computed using the following equation:

$$\Gamma_k(x, y) = [u_{k+}(x, y) * \nabla u_{k-}(x, y) - u_{k-}(x, y) * \nabla u_{k+}(x, y)] / [u_{k+}^2(x, y) + u_{k-}^2(x, y)] \tag{3.8}$$

where ∇ is the gradient operator. Then, the overall local frequency representation of the image is constructed by fusion of the results of the filters at every pixel in the image.

To implement the gradient operator, the forward finite difference or the backward finite difference can be employed. For a 1-D signal, the forward finite difference is defined in Eq. 3.9 and the backward finite difference is defined in Eq. 3.10.

$$u_{forward}(n) = u(n+1) - u(n) \tag{3.9}$$

$$u_{backword}(n) = (u(n) - u(n+1)) \tag{3.10}$$

Computing the central finite difference is another option and is defined by

$$u_{central}(n) = u(n+1) - u(n-1)\,. \tag{3.11}$$

3.3. An Approach to Minimize Latency of the Filter Bank

Fundamental questions in design of filter bank

From Section 3.2, the computation of the local frequency representation for a given image involves two steps:

1) Filtering the input image with a set of 2-D Gabor filters each tuned to a specific spatial frequency, orientation and bandwidth.
2) Computing the gradient of the phase of the filtered images and appropriately combining these into a single local frequency representation.

There are two questions of central importance that must be addressed for an efficient implementation of this approach. First, how many filters need to be included in the filter bank for a given input image? Second, what is the rule to be employed for combining the outputs of the various filters in order to form an accurate local frequency representation? These are questions that have not received significant attention in the literature. As

regards the first question of how many filters to use, researchers that have attempted employing local frequency representations to aid in image processing applications have left this question unanswered and typically propose using an ad hoc selection of the number of filters [95,96]. In practice, one usually errs on the conservative side by employing a larger than needed set of filters in order to cover the entire range of frequencies, thus sacrificing computational efficiency. For large format images, in applications where real-time performance is of interest, minimizing the latency of the filter bank, *i.e.* using the least number of filters and yet covering the entire frequency range, is important. On the second question of how to integrate the outputs of the individual filters to obtain a single accurate local frequency map, a detailed study does not appear to have been made. These are the two questions that are addressed in our research to estimate the local frequency representation of an image.

3.3.1 Local Frequency Representation by Fusion of Four Gabor Filter Outputs

The estimation of natural scales of shapes contained in a given image is a problem that has several applications ranging from pattern recognition to template matching for detection and automatic target classification. Consequently, this problem has received some attention in the literature [102,103]. One of the more popular approaches is the analysis of spatial information by multiple detectors that are sensitive to relevant shapes. An important contribution to this problem was recently made by Fdez-Valdivia et. al. [104] who proposed a multi-channel approach comprising four detectors, each with an orientation bandwidth of 45 degrees. This result was based on some biological evidence

observed in the receptive fields of mammalian visual cortex cells. Although not directly related to the image registration application that is of interest in our research, we draw inspiration from this work to design an optimized multi-channel Gabor filtering scheme for local frequency representation. As a matter of fact, biological evidence shows that the median orientation bandwidth is about 40 degrees, which means that the mammalian visual cortex cell responds better to all the spatial activity in the image within this spatial bandwidth. Therefore, four Gabor functions, with each Gabor function representing a mammalian visual cortex cell, will cover all the spatial activities in the image in all the orientations. This scheme of implementing a Gabor filter bank restricts the spatial orientation, θ , values to 22.5^{o}, 67.5^{o}, 112.5^{o}, and 157.5^{o}. To estimate the spatial frequency of each filter, the authors in [104] proposed a scheme based on splitting each orientation bandwidth into an active region and a passive region according to their spatial frequency contributions to the spatial frequency content of the whole image. Having allocated the active regions in each spatial orientation bandwidth, the center of the active region in the spatial frequency domain determines the spatial frequency, f , for the corresponding Gabor filter. A description of the various steps of the whole algorithm, together with some justifications, will now be given.

3.3.2 Estimation of Filter Bank Spatial Frequencies

Based on biological evidence and theoretical results that explain that the median orientation bandwidth of visual cortex cells is about 40 degrees, the spatial orientation bands of four filters are selected as 22.5^{o}, 67.5^{o}, 112.5^{o}, and 157.5^{o}, respectively

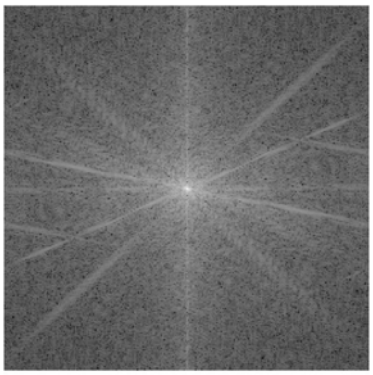


Fig. 3.3 Spectrum of "Cameraman" image

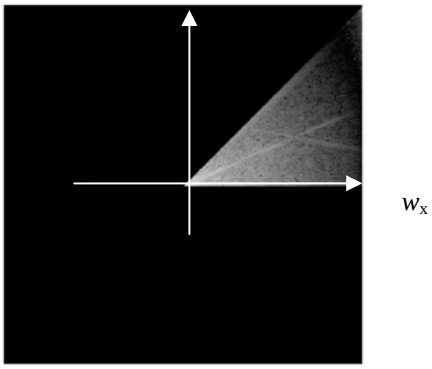


Fig. 3.4. Orientation band at 22.5^o

[104,105]. Let us denote these bands as B_1, B_2, B_3, B_4 and the corresponding set of filters as $G_k(x, y, f_k, \theta_k, \sigma)$, $k = 1,...,4$. Fig. 3.3 and Fig. 3.4 show the 2-D Fourier transform of a "Cameraman" image of size 64 x 64 and the filtered output orientation bandwidth at 22.5^o in the frequency domain of that image, respectively.

Gabor filter spatial frequency adapting

To compute the spatial central frequency f for each filter according to the spatial information for the image, each band is split along the radius in the polar coordinate into regions (sensors), which are called active regions and passive regions in the frequency domain according to its contribution to the spatial frequency content of the image. To show that, an index $\zeta_{B_k}(\rho_{\sup})$ in Eq. 3.12 is constructed to show a natural splitting of each orientation band B_k, k=1,..,4, into a number of spatial frequency channels (ρ_{j-1}, ρ_j) which are not similar in the importance of relative frequency components in the frequency domain,

$$\zeta_{B_k}(\rho_{\sup}) = \frac{\iint_{B_k^{\rho_{\sup}}} |R_{\sigma=1}(\rho,\theta)| \rho d\rho d\theta}{\iint_{B_k} |R_{\sigma=1}(\rho,\theta)| \rho d\rho d\theta}, \tag{3.12}$$

where $|R_{\sigma=1}(\rho,\theta)|$ denotes the Fourier transform magnitude of the input image smoothed by a Gaussian filter with a scale of σ=1. The double integral $\iint_{B_k^{\rho_{\sup}}}$ covers the coordinates (ρ,θ) for the 2-D spectral sector corresponding to the spatial frequency channel (0, $\rho_{\sup}$)

upon the orientation band B_k, while the double integral $\iint_{B_k}$ covers the entire frequency band B_k. (ρ,θ) are polar coordinates. The reason that the index is written as a function of ρ_{sup} is because the index may be seen as a function of radial frequency ρ_{sup} and increases as ρ_{sup} increases. Fig. 3.5 shows the index $\zeta_{B_k}(\rho_{sup})$, for the orientation band at 22.5^o in Fig. 3.4. Visual investigation of Fig. 3.5 shows that there exists significant rise in the value of the index $\zeta_{B_k}(\rho_{sup})$ at specific locations ρ_{sup} . These locations represent the natural splitting for the band B_1 into spatial frequency channels. The rise at these locations is less or more abrupt depending on the relative importance of the spatial frequency components that are added. To detect these locations, the extrema of the second derivative with respect to ρ can be employed.

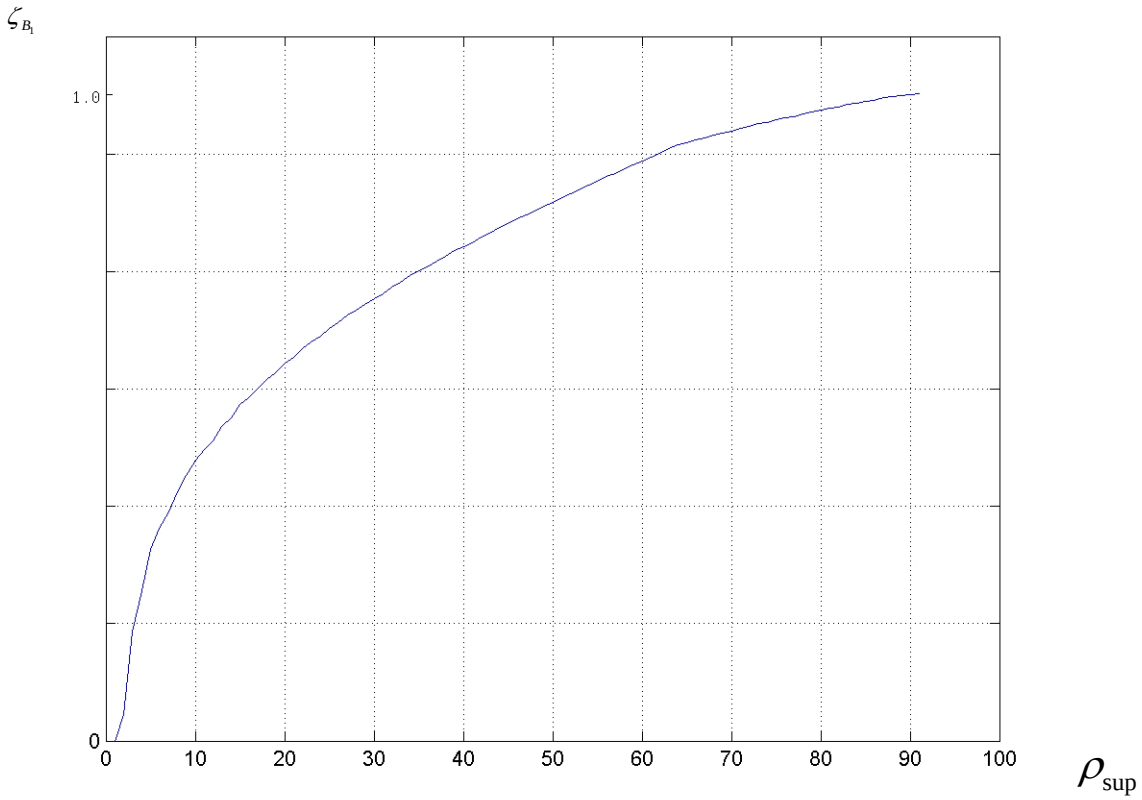


Fig. 3.5. Index $\zeta_{B_1}(\rho_{\text{sup}})$ for the orientation band at 22.5$^{\text{o}}$ in Fig. 3.4.

To distinguish between these channels in their importance to the frequency domain for the image, another measure is used, $w(ch_{k_i})$,

$$w(ch_{k_i}) = \frac{\iint_{ch_{k_i}} |R_{\sigma=1}(\rho,\theta)|^2 \, \rho d\rho d\theta}{\iint_{\Omega} |R_{\sigma=1}(\rho,\theta)|^2 \, \rho d\rho d\theta}. \tag{3.13}$$

Where $|R_{\sigma=1}(\rho,\theta)|^2$ denotes the power spectrum of the image smoothed at the scale $\sigma = 1$ and the double integral $\iint_{ch_{k_i}}$ covers the channel *i* in the band *k*. The double integral $\iint_{\Omega}$ covers the entire frequency domain of the image. This measure is used to distinguish between the important channels and not-important channels in the spectral domain for each orientation, B_k, k=1,…,4. Having obtained these features, they are employed to cluster the channels in each band into active cluster and passive cluster categories. *K-means algorithm* is employed to cluster all the channels in each band into only two main clusters: an active cluster (includes the active channels) and a passive cluster (includes passive channels). The seed of the active cluster in k-means algorithm is the maximum value of $w(ch_{k_i})$ in the orientation band *k* and the seed of the passive cluster in k-means algorithm is the minimum value of $w(ch_{k_i})$ in the orientation band *k*.

The location of the active cluster center in a specific orientation band is adapted as the spatial frequency f for the Gabor filter that is tuned to that band. Fig. 3.6 shows the feature measure, $w(ch_{k_i})$, against the channels generated from Fig. 3.5. Fig. 3.7 shows the flow chart of all the steps of the algorithm.

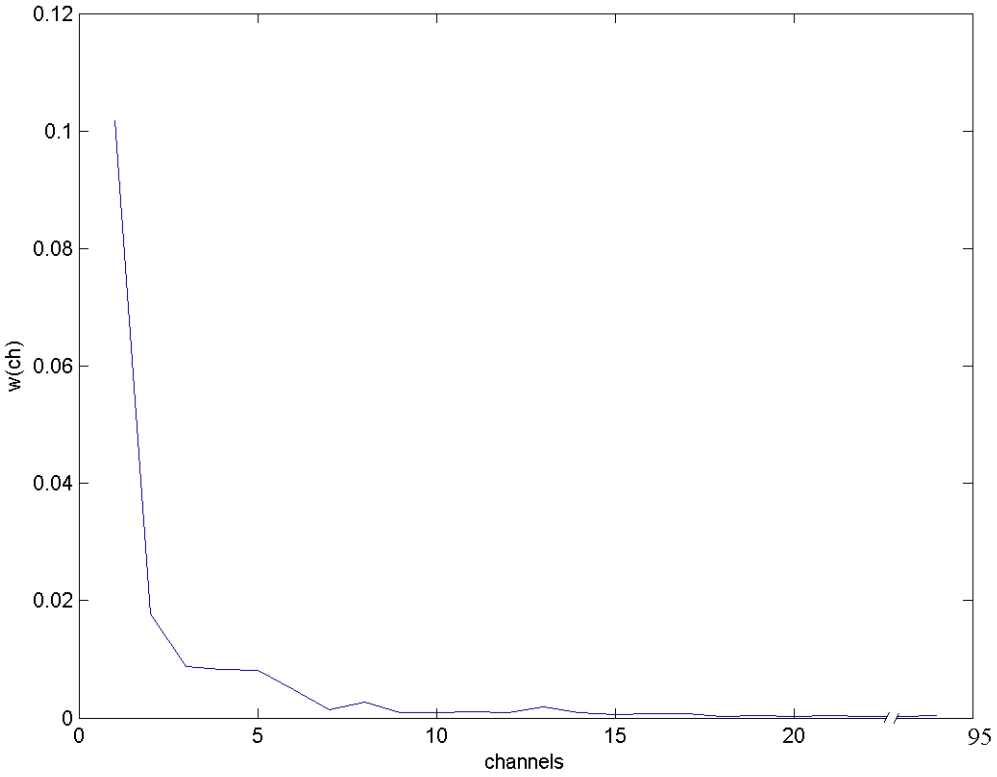


Fig. 3.6 Measure $w(ch_{k_i})$ against the channels generated from Fig. 3.5.

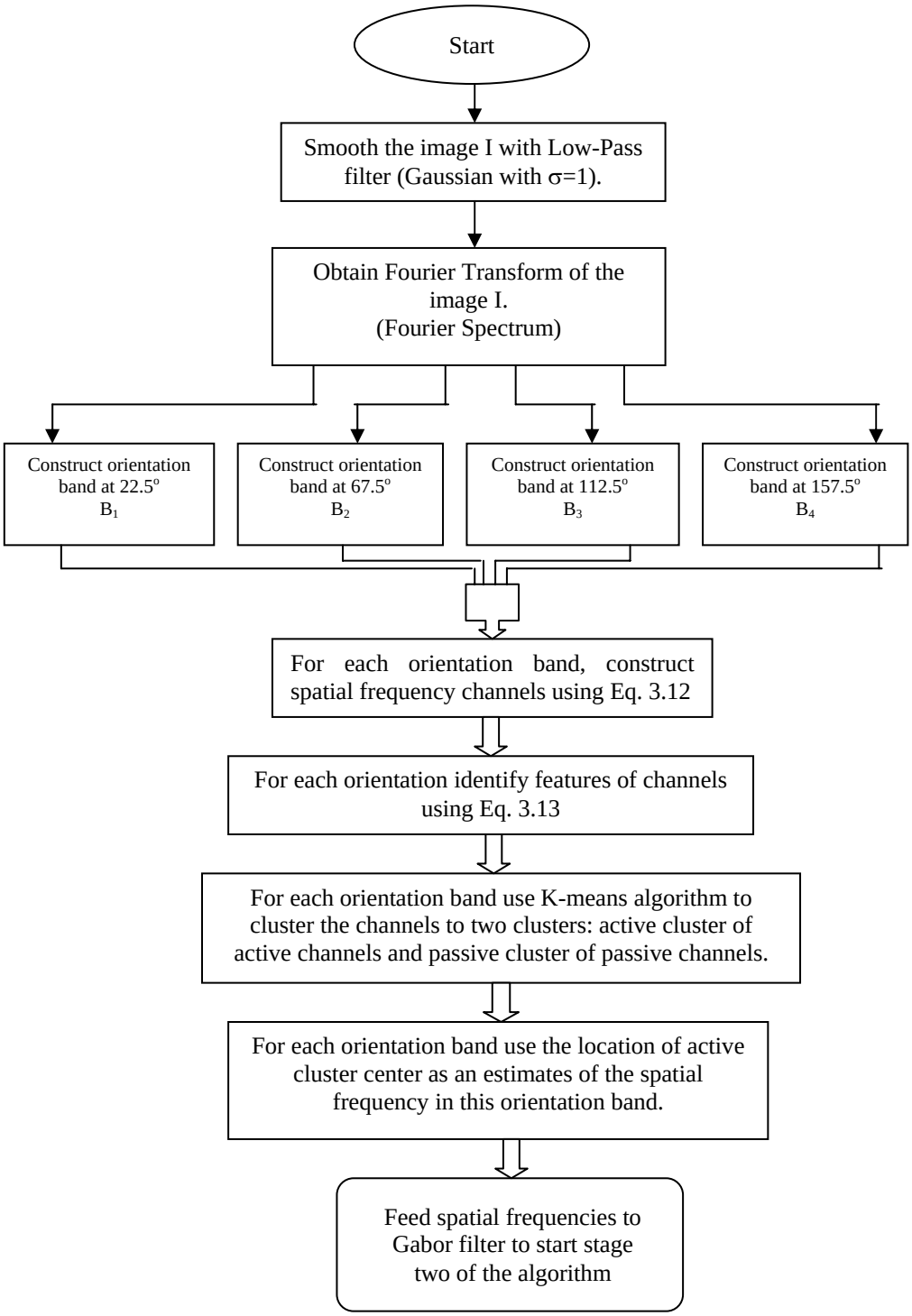


Fig. 3.7 Spatial Frequency estimation of each orientation band

Once the spatial frequency, f, of each filter is estimated, we convolve the image with each filter. Then, we use the resulting analytic signal from each filter to compute the local phase and local phase gradient, or we can estimate the local frequency of the image using Eq. 3.8. The next section presents the details of the proposed algorithm to obtain the local frequency for an image.

3.3.3 Algorithm for Local Frequency Estimation

To calculate the local phase, we can use the ATAN function on the real and the imaginary values for the analytic signal to calculate the local phase and then we compute the gradient of the local phase. In this method, we have to perform quadrant correction for the output of ATAN function based on the signs of the real part and the imaginary part of the analytic signal. This is because ATAN is periodic in π rather than 2π. The other option for estimating the local frequency without computing the local phase and the derivative of the local phase is using Eq. 3.8. To obtain a good approximation for the derivative in the discrete domain, we employed the fourth order estimator from Table 3.1 [94].

Table 3.1

Order	Coefficients
2	-1/2 0 1/2
4	1/12 –2/3 0 2/3 -1/12
6	-1/60 3/20 -3/4 0 3/4 -3/20 1/60

A brief description for the developed algorithm together with some justification will now be given.

Algorithm 3.1. Obtaining local frequency representation by using adaptive Gabor filter set.

(1) Create a set of four Gabor filters, $G_k(x, y, f_k, \theta_k, \sigma)$, k=1…4, to cover the frequency space of the image under investigation, by selecting the orientation bands of the four filters set to 22.5°, 67.5°, 112.5°, and 157.5° respectively. It may be noted that each filter has a spatial orientation bandwidth of 45°, such that the four filters cover the entire spatial bandwidth of 180°.

(2) Estimate the spatial frequency of each filter using the corresponding orientation band as described above.

(3) Convolve the input image I whose local frequency representation is desired with each filter G_k, k=1,..,4, to obtain the corresponding analytic signals

$$u^{(k)} = I \otimes G_k . \tag{3.14}$$

(4) Compute the local phase, $\Psi^{(k)}$, for the analytic signal $u^{(k)}$, k=1,..,4, by calculating

$$\psi^{(k)}(x, y) = \tan^{-1} \frac{imag(u^{(k)}(x, y))}{real(u^{(k)}(x, y))}, \tag{3.15}$$

$(x, y) \in \Omega$ (Ω is the set of image points.),

(5) Create the local frequency representation formed by each filter using

$$\Gamma^{(k)}(x, y) = \sqrt{\nabla^2{}_x(\psi^{(k)}(x, y)) + \nabla^2{}_y(\psi^{(k)}(x, y))}\,|\cos(\theta_k - \theta_\Delta{}^{(K)})|, \tag{3.16}$$

where $\nabla_x(.)$ and $\nabla_y(.)$ are the gradient estimation in the x and y directions respectively, θ_k is the orientation of the k^{th} Gabor filter, and $\theta_\Delta{}^{(k)}(x, y)$ denotes the direction of the gradient

vector with respect to the *x*-axis , $\theta_{\Delta}^{(k)}(x,y) = \tan^{-1}\frac{\nabla\psi^{(k)}{}_{y}(x,y)}{\nabla\psi^{(k)}{}_{x}(x,y)}$. It may be noted that $\Gamma^{(k)}$ provides a spatially localized estimate of the local frequency along the direction θ_k .

$(x,y) \in \Omega$ (Ω is the set of image points.).

(6) Fuse the local frequency estimates obtained from the four filters, Γ^k, *k*=1…4, to get one combined representation using

$$\Gamma(x,y) = \mathit{fuse}\{\Gamma^{(1)}(x,y), \Gamma^{(2)}(x,y), \Gamma^{(3)}(x,y), \Gamma^{(4)}(x,y)\}, \tag{3.17}$$

$(x,y) \in \Omega$ (Ω is the set of image points.). Fig. 3.8 shows a schematic for the algorithm.

Justifications for the various steps in the above scheme can be given as follows. The bandwidth σ of a Gabor filter provides a notable scalability property in the spatial domain. Motivated by the work of Fdez-Valdivia et. al. [102], the present scheme employs four Gabor filters, each covering a spatial orientation band of 45^{o} width. The biological evidence on which this development is based confirms in a qualitative sense that the four filters are sufficient to capture the variations in the spatial frequency activity characterizing the given image. How do we demonstrate that this size of the filter bank constitutes an optimal size? While an analytical determination is not possible (since different input images in general have different local frequency structures), simulation studies of the representation accuracy resulting from filter banks of different sizes can be made to justify the above selection in terms of our desire to address the tradeoff between representation accuracy and latency minimization.

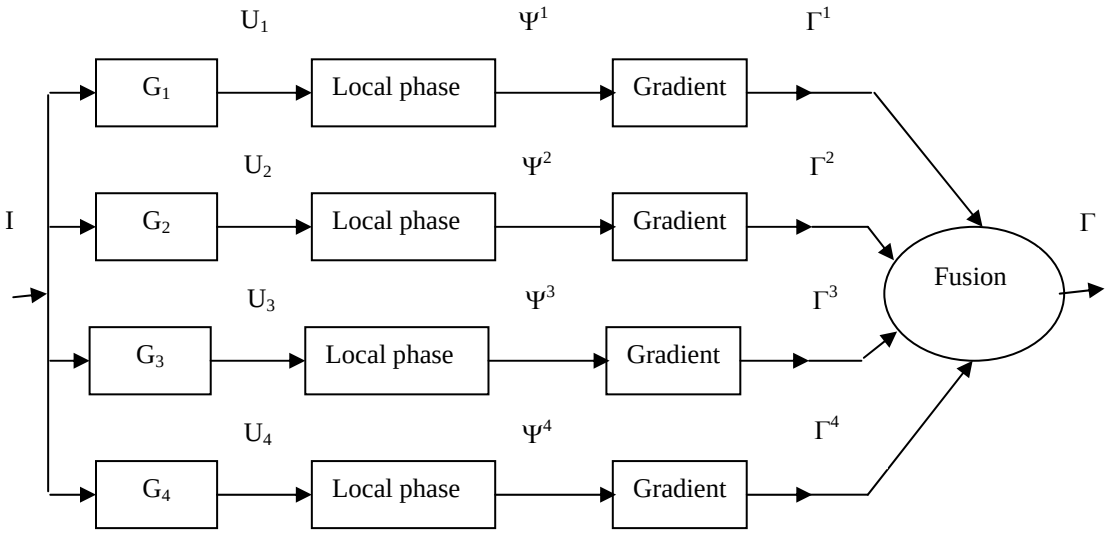


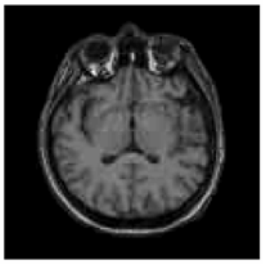
Fig. 3.8. Schematic for the local frequency extraction from a given image I process.

Fusion local frequency from the filter set

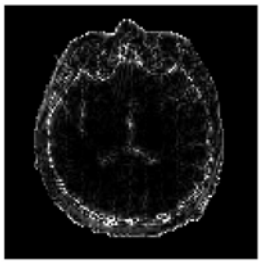
A few words justifying the fusion rule described by Eq. 3.17 are in order. Due to the different orientations used for the four filters, each filter forms a different local frequency representation of the input image. However, the local frequency variation within a certain frequency range coinciding with the spatial orientation band to which each filter is tuned is emphasized more in its output. Consequently, a fusion rule that extracts in the final result the most emphasized structures formed by the individual filters at each pixel location (in the spatial domain) will yield the most accurate estimate of the local frequency activity present in the given image. So, Eq. 3.17 will be as follows:

$$\Gamma(x,y) = Max\{\Gamma^{(1)}(x,y), \Gamma^{(2)}(x,y), \Gamma^{(3)}(x,y), \Gamma^{(4)}(x,y)\} \tag{3.18}$$

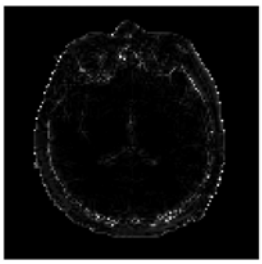
To confirm this, results of one experiment will be outlined in Fig. 3.9. Fig. 3.9a shows the input image. The local frequency representations formed along the four filter channels (*i.e.* $\Gamma^{(1)}$, $\Gamma^{(2)}$, $\Gamma^{(3)}$ and $\Gamma^{(4)}$) are shown in Figs. 3.9b-3.9e. Fig. 3.9b displays a more emphasized local frequency activity within the 45^{o} spatial orientation band centered around the 22.5^{o} orientation line compared to the representations formed along the three other channels. Figs. 3.9c, 3.9d, and 3.9e show the corresponding local frequency representations formed along the three other channels, which depict a more emphasized structure within their spatial orientation bands. Fig. 3.9f shows the final representation



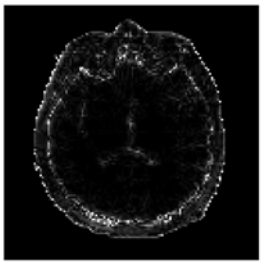
3.9a



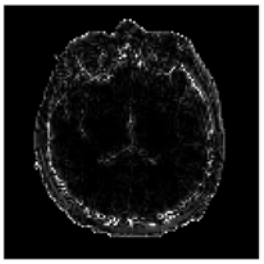
3.9b



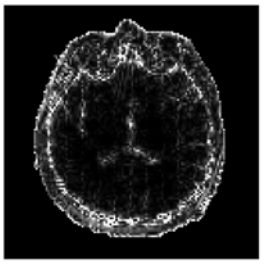
3.9c



3.9d



3.9e



3.9f

Fig. 3.9. Results from processing a T1-weighted MRI frame. *3.9a*. Input image; *3.9b –3.9e*. Local frequencies individually estimated along the four filter channels coded in gray-scale images; *3.9f.* Final local frequency representation formed by fusion of four filter outputs.

obtained by fusing the four representations according to the present fusion rule, which clearly displays a more emphasized local frequency structure in the entire image.

3.3.4 Evaluation of The Proposed Scheme

Several simulation experiments with different images were conducted to confirm in a quantitative sense the latency minimization property, *i.e.* to show that the filter bank of size 4 tailored with our approach is capable of reliably extracting most of the local frequencies. Results from these experiments will be given here for illustration. In the first experiment, the local frequency representation for a T2-weighted MRI frame of size 128 x 128 was obtained by employing the above algorithm, as well as by executing the same algorithm two more times, once with changing the number of Gabor filters used to 3 and the other with changing the number of filters to 6. In the three-filter case, the orientations of the filters were set at 30^{o}, 90^{o}, and 150^{o} in order to cover the 180^{o} range (note that each filter now covers a 60^{o} range in the spatial band), whereas in the six-filter case the filter orientations were set at 15^{o}, 45^{o}, 75^{o}, 105^{o}, 135^{o}, and 165^{o} once again to cover the 180^{o} range (note that each filter in this case covers a 30^{o} band). To ensure a fair comparison, the value of the scale parameter σ was selected to be the same in all cases. Fig. 3.10 shows the results of this experiment. The input MRI frame is shown in Fig. 3.10a and its local frequency representation obtained from the four-filter scheme is shown in Fig. 3.10b. Since for demonstrating the latency minimization we are interested in what the three-filter scheme fails to capture compared to the four-filter scheme, and also what additional information the six-filter scheme is capable of capturing compared to

the four-filter scheme, we computed the difference between the local frequency maps obtained by implementing each of these schemes. Fig. 3.10c depicts the difference between the results obtained from the three-filter and the four-filter schemes, while Fig. 3.10d shows the difference between the results obtained from the four-filter and the six-filter schemes. As can be seen from Figs. 3.10c and 3.10d, there is considerable local frequency activity captured by the four-filter scheme that did not show up in the output of the three-filter scheme, while the difference between the outputs of the four-filter and six-filter schemes is hardly significant (making the return from the additional computational effort, an almost 50% increase, not commensurate). Results from several other experiments were quite similar leading to the conclusion that the four-filter scheme shown in Fig. 3.8 offers a good tradeoff between representation accuracy and latency minimization.

In other simulations performed to support the above conclusion, we used a T1-weighted MRI frame of size 128 x 128 and a “Cameraman” image of size 128 x 128. In each simulation, the local frequency was obtained by employing the present four Gabor filters scheme, as well as by employing 3 and 6 filters in the filter bank. Then, we computed the difference between the local frequency that is obtained by 4 Gabor filter scheme and the other schemes. Fig. 3.11 and Fig. 3.12 show the results of the two simulations. Apparently in both simulations, the four-filter scheme still presents a good tradeoff between the accuracy and the latency minimization.

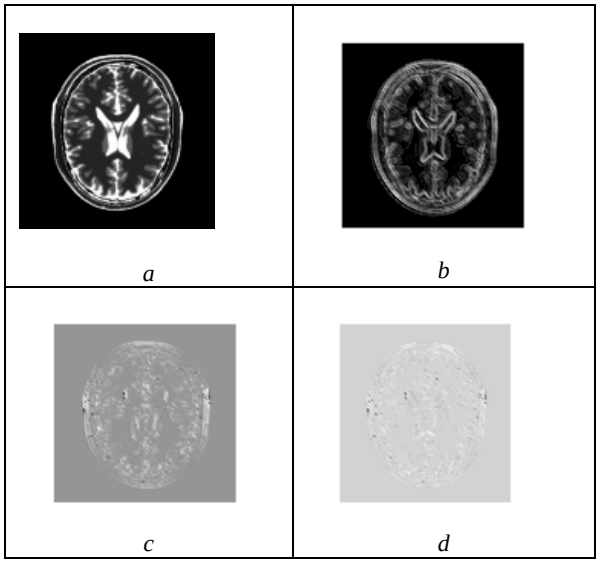


Fig. 3.10. Results from processing a T2-weighted MRI frame.
3.10a. Input image; *3.10b*. Local frequency representation obtained by present scheme; *3.10c*. Difference between local frequency representations obtained by present scheme and a three-filter scheme; *3.10d*. Difference between local frequency representations obtained by present scheme and a six-filter scheme.

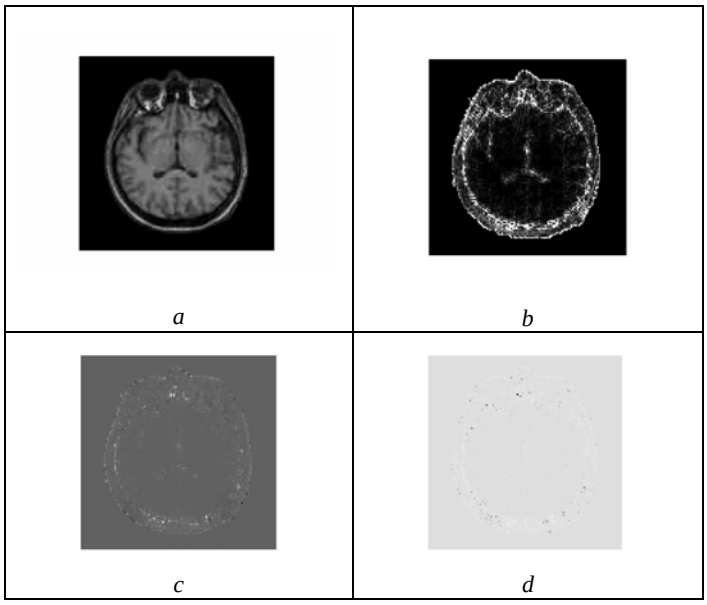


Fig. 3.11. Results from processing a T1-weighted MRI frame.
3.11a. Input image; *3.11b*. Local frequency representation obtained by present scheme; *3.11c*. Difference between local frequency representations obtained by present scheme and a three-filter scheme; *3.11d*. Difference between local frequency representations obtained by present scheme and a six-filter scheme.

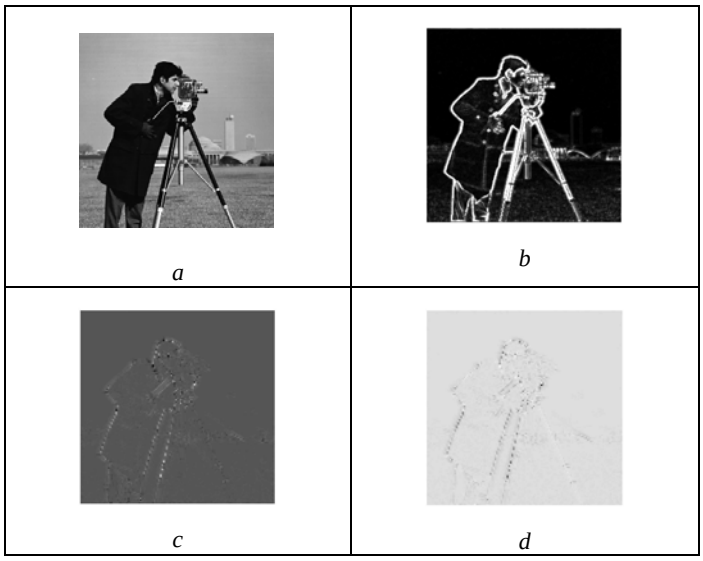


Fig. 3.12. Results from processing a Cameraman image.
3.12a. Input image; *3.12b*. Local frequency representation obtained by present scheme; *3.12c*. Difference between local frequency representations obtained by present scheme and a three-filter scheme; *3.12d*. Difference between local frequency representations obtained by present scheme and a six-filter scheme.

3.4 Image Registration Based on Local Frequency Representation

As noted before, local Frequency enjoys the following characteristics:

It is relatively invariant to illumination change in the scene. This property gives LF superiority in processing the images for many applications since the local frequency, unlike the gradient, is relatively independent of signal contrast [96].

Another superiority of local frequency is the capability of detecting the edges and the ridges in the image at the same time. Due to this property of local frequency, it is a perfect candidate to solve the problem of image registration, especially multi-modal medical image registration, because the ridges and the edges are common in the medical images. In general, to detect the edges and the ridges in an image, one typically needs an edge detector and a ridge detector [107] separately, but using local frequency can detect the edges and the ridges at the same time.

One more advantage of the local frequency representation is that it is relatively insensitive to photometric deformations that result from changes in viewing direction, surface normal and lighting conditions. Therefore, it solves the problem of establishing correspondence between stereo images with high accuracy [108].

Finally, local frequency can be extracted from a given image using systematic signal processing operations [97] and has good localization, which means that the local frequency is quiet faithful to the variations in the original signal [95].

These advantages of the local frequency provided the motivation to develop an algorithm based on local frequency representation of the image to solve the image registration problem. The developed algorithm depends on using the local frequency estimated by employing the algorithm proposed in Section 3.3.4 (Algorithm 3.1). Then, the local frequency representation is employed to extract control points from the image. After that, a matching algorithm is implemented to match these control points and to estimate the registration parameters. In addition to exploiting the advantages of local frequency, the proposed algorithm avoids the associated problems with feature-based image registration methods such as selecting the type of features, selecting an appropriate feature detection algorithm, and method of representation of the features. The next two sections will explain the employed strategies for extracting the control points and matching them.

3.4.1 Control Points Extraction

Solving the image registration problem using feature-based methods leads to the following questions, which are considered as sub-problems of the image registration problem:

What are the useful features in the image that can be employed in solving the problem?

What is the optimal algorithm to extract these features?

And how to represent these features to be control points?

Determining useful features in the image depends on the scene. In other words, the useful features in a scene that has homogenous regions such as lakes and fields are different

from the useful features in another scene that has roads and rivers. Moreover, the useful features in indoor scenes are different from outdoor scenes. The indoor features most likely include edges, corners and crossing points. Having identified the useful features in a scene, the extraction algorithm for these features will depend on the types of features. For example, the algorithm that can be employed to segment the homogenous regions (their centers are used as control points) in an image cannot be used to detect features such as roads. Also, the algorithm that can be used to detect roads cannot be used to detect corners and ridges. Therefore, the employed algorithm for feature extraction will depend on the scene under investigation. Also, some features such as lakes have many representations in the feature space, e.g. it can be represented by the center of gravity or by its contour, and one needs to decide which representation will be useful.

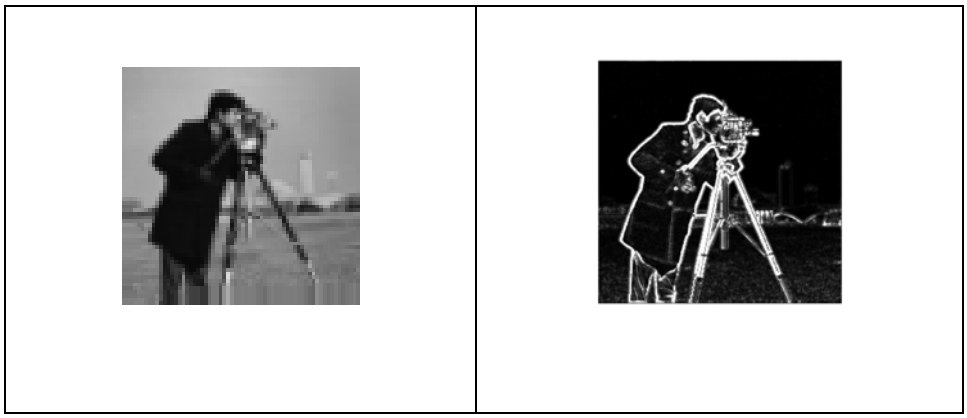
Therefore, we have developed an algorithm to extract features (control points) without depending on the scene or the types of the features. The proposed method is based on using the local frequency representation of the image to extract control points. Local frequency values of an image reflect the activities in the image scene and high values are associated with high activities in the image scene. High activities in the image scene occur where there exist features such as edges and ridges in the scene; therefore high values of the local frequency representation locate the useful features in the image. The control points are selected based on the values of the local frequency, which can be determined using a threshold or selecting the points that have higher local frequency values. Thus, we select the control points that are located at the most apparent activities in the image such as edges and ridges using the local frequency representation of the

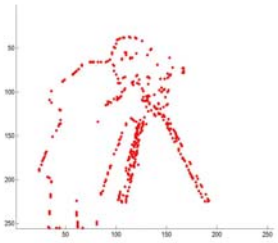
image. Obtaining the control points by this approach avoids the problems of conventional feature extraction methods, since the control points do not depend on the structure in the scene and the method of selecting the control points is the same for all types of structures. The size of the control point set is determined such that the number of points is enough to capture features in both images and to establish matches between them. From many conducted experiments, the number of the control points varied from image to image according to the size of the image. In our experiments, we found the number of control points that is enough to capture features from the image and to establish matching varied from 50 to 300 according to the image size. Fig. 3.13a shows a “Cameraman” image and Fig. 3.13b shows the corresponding local frequency representations encoded as a gray-scale image. The higher gray-scale values correspond to the higher local frequency values in the local frequency representation and vice versa. Fig. 3.13c shows the selected control points from using local frequency representation in Fig. 3.13b.

3.4.2 Control Points Matching Algorithm

Once the control points are selected for a pair of images, the correspondence problem between them has to be solved to estimate the registration parameters. While there exist many point matching procedures [109-116], we employed the algorithm presented by Gold et. al. [109]. This algorithm estimates the registration parameters and solves the correspondence problem by incorporating an optimization technique and an iterative correspondence assignment technique called “Softassign”, which is a general procedure for identifying the correspondence between two sets of points in space. The motivation

(a) (b)





(c)

Fig. 3.13 (a) ”Cameraman” image., (b) the extracted local frequency representation using the proposed algorithm coded as a gray-scale image, (c) the extracted control points using the proposed scheme.

for using this algorithm comes from its capability that it detects the outliers from matched pairs while at the same time solves the correspondence problem and estimates the transformation parameters. The algorithm minimizes the following objective function:

$$\min_{m,A} \; E(m,A) = \sum_{i=1}^{M}\sum_{j=1}^{N} m_{ij} \left\| X_i - AY_j \right\|^2 - \alpha \sum_{i=1}^{M}\sum_{j=1}^{N} m_{ij} \tag{3.19}$$

In Eq. 3.19, m_{ij} denotes the correspondence variables that define the match matrix of dimension $M \times N$. If m_{ij} is one, point *i* in image X and point *j* in image Y have correspondence, and if they do not correspond, m_{ij} equals zero. If a point in one image does not have correspondence, then it is discarded as an outlier. The second term with the multiplier α biases Eq. 3.19 towards matches. It acts as a threshold error distance, indicating how far apart the two points must be before they may be treated as outliers. $\{X_i\}$ and $\{Y_j\}$ are two 2-D point sets, related by the affine transformation $X = AY$. A is a 3x3 2-D affine transformation matrix in the plane and defined by six parameters *a,b,c,e, f,* and *g* in the form

$$A = \begin{bmatrix} a & b & c \\ e & f & g \\ 0 & 0 & 1 \end{bmatrix} \tag{3.20}$$

These six parameters specify the translation, the scaling and the rotation in the plane. Eq. 3.19 describes an optimization problem whose solution yields the transformation matrix, *A*.

Softassign

For M image points in a set X and another N image points in a set Y, the correspondence between the two sets is given by an (M+1) x (N+1) assignment matrix *m*, where $0 \leq m_{ij} \leq 1$. The value of m_{ij} ($1 \leq i \leq M$, $1 \leq j \leq N$) determines how well the i^{th} point in a set of points matches the j^{th} point in the other set of points. At the beginning of the algorithm, all m_{ij} have the same values indicating that the correspondences between the two point sets are unknown (for example $m_{ij} = 0.1 \quad \forall i \text{ and } j$). The last row, M+1, and the last column, N+1, of the matching matrix, m, are the slack row and slack column, respectively. The slack variables receive large values when the corresponding point in a set doesn't match any point in the other set, which means that the point is an outlier. The goal in this algorithm is to determine the matching matrix $m = \{ m_{ij} \}$ in a zero-one form to explicitly specify the matches between the two sets of points. The matching matrix satisfies the two way constraint that each point in one set matches at most one point in the other set and vice versa. Softassign algorithm can be summarized as follows:

1- initialize m by an arbitrary value

2- update the matching matrix m by iterative row and column normalization until m converges . The constraints of row and column normalization are $m_{ij}^{o} \leftarrow \frac{m_{ij}}{\sum_{j=1}^{N} m_{ij}}$,

$m_{ij} \leftarrow \frac{m_{ij}}{\sum_{i=1}^{M} m_{ij}}$, respectively. Here m_{ij}^{o} denotes the previous value of m_{ij} before the current iteration of normalization. The criterion for the convergence is $\sum_{i=1}^{M}\sum_{j}^{N}\left|m_{ij}^{o} - m_{ij}\right| \prec \varepsilon_1$, where ε_1 is an arbitrarily small number.

3- update m using $m_{ij} = \exp[\beta q_{ij}]$, where q is updated using $q_{ij} = -(\left\|X_i - AY_j\right\|^2 - \alpha) = -\frac{\partial E}{\partial m_{ij}}$ and β is initialized using β^o, an arbitrarily small value, and the following constraint holds $\forall i \sum_{j=1}^{N} m_{ij} \le 1$, $\forall j \sum_{i=1}^{M} m_{ij} \le 1$, and increased in every updating to q .

The Softassign algorithm can be employed alone to solve the correspondence problem and β^f can be used to stop the algorithm, where β^f is an arbitrary value such that $\beta^f \succ \beta^o$. The slack variables are introduced from linear programming techniques and are determined using the following constraint $\forall i \sum_{j=1}^{N+1} m_{ij} = 1$, $\forall j \sum_{i=1}^{M+1} m_{ij} = 1$.

Point pattern match algorithm summary:

The point matching algorithm including the softassign algorithm can be summarized as follows:

- Initialize the control parameters and matching matrix m_{ij} including the stack variables.
- Update the matching matrix using the Softassign algorithm.

- Update matrix *A* using Eq. 3.19. Updating the matrix *A* is achieved by minimizing the objective function, Eq. 3.19, with respect to *A* and assuming anything else is fixed as follows:

$$A = (\sum_i \sum_j m_{ij} X_i Y_j^T)(\sum_i \sum_j m_{ij} Y_j Y_j^T)^{-1} \tag{3.21}$$

- Update the control parameter β.
- Iterate until convergence. The convergence criterion is

$|\theta^o - \theta| + |s^o - s| + |t^o - t| \prec \varepsilon_2$. Where ε_2 is an arbitrarily small number. The $\theta, s,$ and t are the rotation angle, the scaling, and the translation parameters of the affine transformation respectively.

After the algorithm converges, the transformation parameters are obtained from A.

To summarize the proposed image registration algorithm, given two images and distortion (translation, scaling, rotation) between them, the local frequency representation is extracted for both images and the control points are selected. Finally, the matching algorithm is employed to solve the correspondence problem and to estimate the registration parameters. A brief description of the various steps is given below and a schematic overview of the algorithm is shown in Fig. 3.14.

Algorithm 3.2

1- Use Algorithm 3.1 to extract the local frequency representation of the input image.

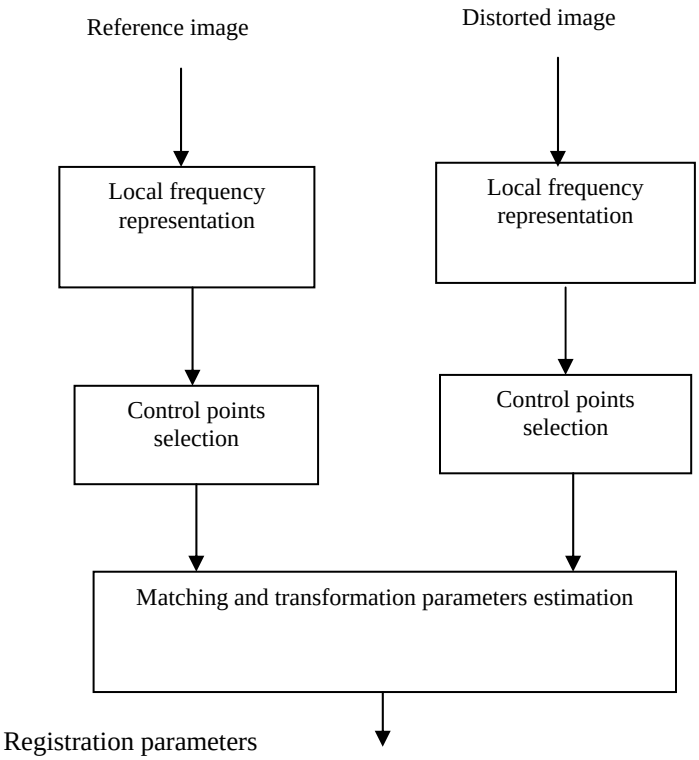


Fig.3.14 Schematic overview of the proposed image registration algorithm.

2- Select a set of control points in each image based on the values of the local frequency representation for that image.
3- Employ the matching algorithm to solve the correspondence problem and to estimate the registration parameters.

3.5 Experimental Results

In this section, we describe three different types of experiments. The first type of experiments is to evaluate the performance of the proposed algorithm in extracting the local frequency of different image modalities. The second type of experiments is to evaluate the results of the proposed scheme for image registration. The last part from experiments is to evaluate the efficiency of the proposed algorithm for image registration in dealing with images of different complexities.

A number of experiments were conducted to evaluate the performance resulting from the present approach to extract local frequency representation from different types of image data and to establish its robustness to scene details. We also tested the performance resulting from the use of the present approach to image registration problems. For this purpose we conducted experiments on images that contain real objects, as well as, images that contain synthetic objects. Results from these experiments will be presented here to illustrate the strong points of the present approach.

Fig. 3.15 shows the results of an experiment where a number of images captured by sensors of different modalities were processed by the present algorithm. The images and

the corresponding local frequency representations formed by the algorithm are shown. The first of these images is an EO image, “Cameraman” image, of size 128 x 128. The corresponding local frequency representation is encoded in a gray-scale image and it is shown in first row in right column. The higher the gray-scale values, the higher the local frequency values in the local frequency representation and vice versa.

From the local frequency representation shown, one can clearly see the activities in the EO image. In the middle row, the second image is a T1-weighted MRI image of size 128 x 128, whose local frequency representation that is encoded as gray-scale image reflects all the edges and the ridges characterizing the image. In the third row, the image is a synthetic one containing a trapezoidal object with sharp edges. Once again, the local frequency representation that is shown in 3-D view clearly displays these edges, which are extracted perfectly by the algorithm. These results serve to illustrate the versatility of the approach used here to process images of different types and the robustness of the algorithm to variations in the scene.

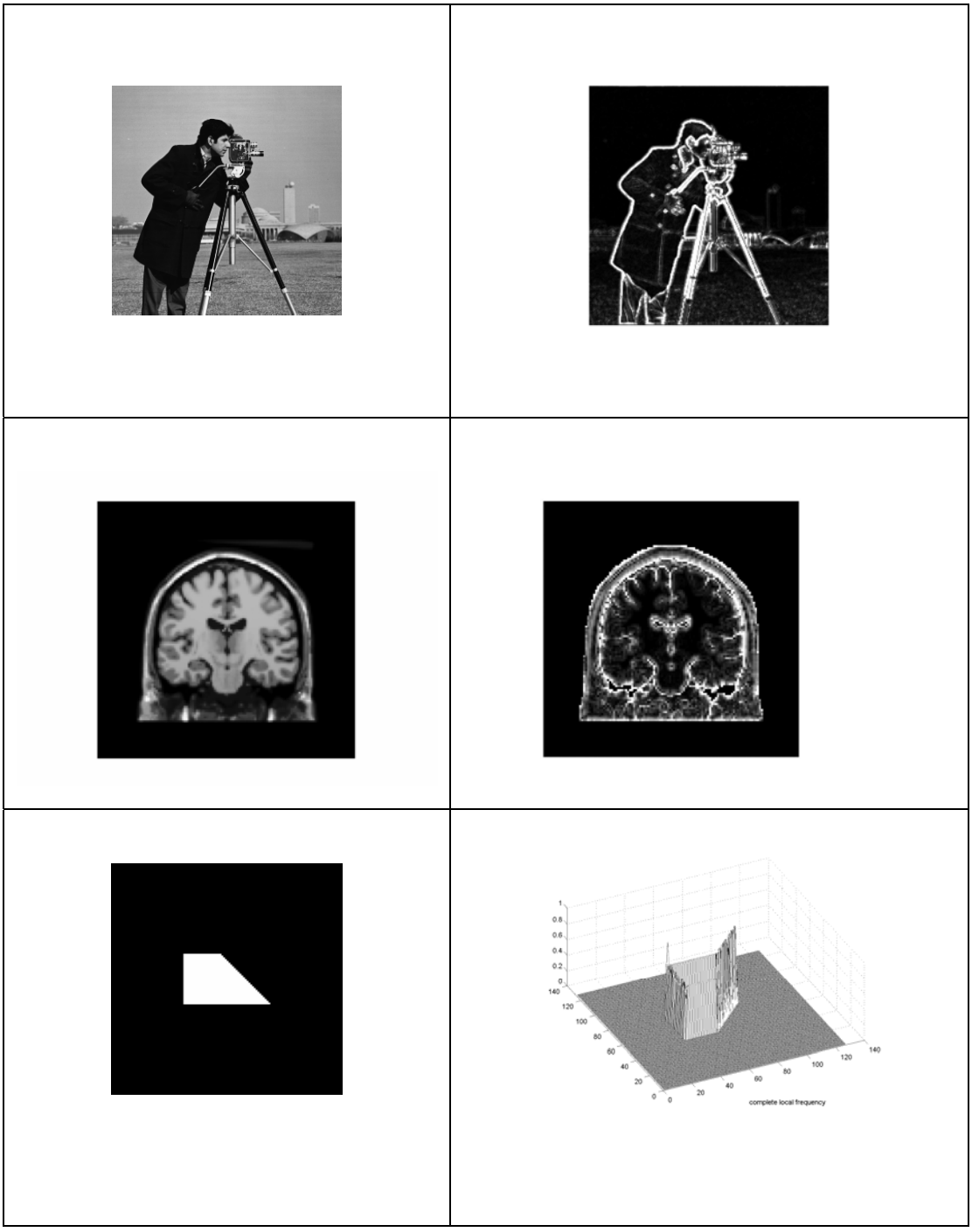


Fig.3.15. Three different types of images and their corresponding local frequency representations formed by the present algorithm. *Left column: Top* – EO "Cameraman" image of size 128 x 128, *Middle* - T1-weighted MRI image of size 128 x 128, *Bottom* - a synthetic image of size 128 x 128 containing a trapezoidal object. Right column: *Top and middle* - Gray- scale coding of the corresponding local frequency representations, *Bottom* - the corresponding local frequency representation in 3D view.

For demonstrating the accuracy of the local frequency representation and its utility in image registration, we conducted the following controlled experiments. In the first experiment, we used the 128 X 128 T1-weighted MRI frame shown in Fig. 3.16 as a reference image and we distorted it manually by known transformation parameters (rotation and translation along *x*- and *y*-directions) to create a second frame. The two frames are then registered by forming local frequency representations from the two images with the present algorithm (Algorithm 3.1), and implementing a point matching registration procedure. While there exists a number of specific algorithms in the literature that employ point matching ideas for registration [46,54], the registration performance resulting from these critically depend on the quality (accuracy) of the points that can be used for matching. This is indeed where the accuracy of the local frequency representation formed by our algorithm makes a major difference in the outcome of the registration process. From the local frequency representations obtained from the two frames (the reference and the distorted), we selected a set of points that have a significant local frequency magnitude. For the present experiment it was seen that about 200 points extracted from the local frequency representations were sufficient to capture the important features in order to establish the correspondence between the two images (the number of points to be chosen depends in general on the size of the images). Fig. 3.16 shows the result of this step of the procedure, where the reference T1-weighted MRI frame is shown in Fig. 3.16a and the set of 200 control points extracted for matching from the local frequency representation is shown in Fig. 3.16b.

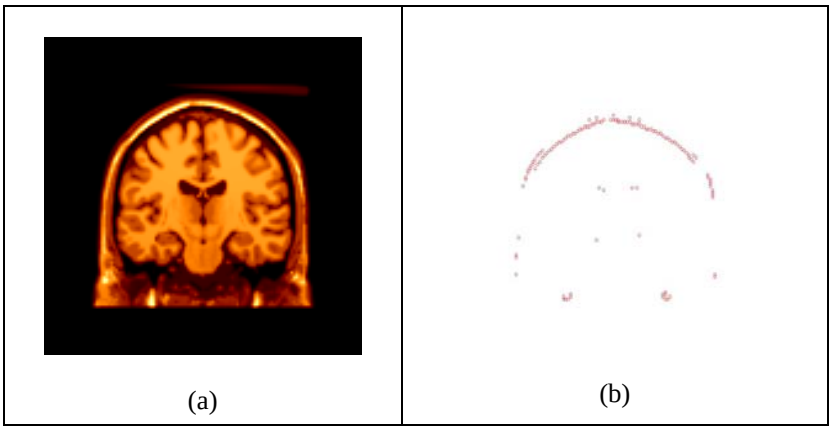


Fig. 3.16. Extraction of control points for registration.
a. Reference T1-weighted MRI frame; *b*. Extracted control points from the local frequency representations.

Employing these 200 control points and implementing the point matching procedure given by Gold *et al*., [109], the registration problem could be solved with a high degree of accuracy. Results of this experiment are summarized in Table 3.1, which shows the estimated transformation parameters (rotation ϕ, translation T_x and translation T_y) as well as the ground truth values of these parameters under Registration Experiment #1.

To further highlight the local frequency extraction performance of our algorithm when applied to images from different modalities and its usefulness in accurately solving registration problems, a similar experiment was conducted with a 64 x 64 CT image

frame that was distorted with different transformation parameters. These frames were then used as inputs to our algorithm and 100 points with the higher magnitude values were selected from the local frequency representations in order to establish correspondence between the two frames. Fig. 3.17 shows the reference CT image frame and the 100 extracted control points. The registration performance resulting from this selection of control points is summarized in Table 1 under Registration Experiment #2. The accuracy with which all registration parameters (rotation ϕ, translation T_x and translation T_y) have been estimated attests to the quality of the control points extracted for matching, which in turn confirms the accuracy of the local frequency representations formed by the present algorithm.

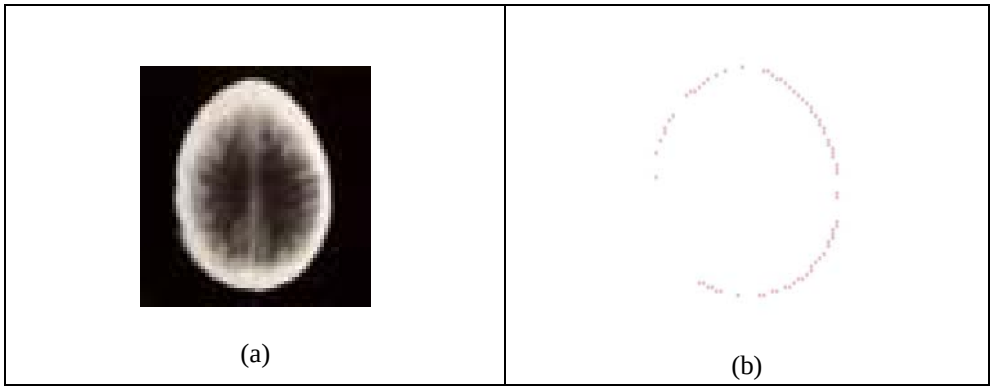


Fig. 3.17. Control point extraction for image registration.
a. Reference CT image frame; *b*. 100 control points extracted.

Table 3.2: Results of Registration Experiments

Registration experiment #	Registration parameters	Ground truth data	Estimated data
1	Rotation	2.0 ^{o}	2.02 ^{o}
	Translation T_x T_y	0.0 4.0	0.054 4.117
2	Rotation	1.0 ^{o}	0.9 ^{o}
	Translation T_x T_y	3.0 -2.0	2.97 -2.006

Local frequency representation based registration of complex images

While some efficient approaches to image registration have been developed lately, the registration algorithms resulting from these approaches generally remain application dependent and may require operator-assisted tuning for different images to achieve equal accuracy levels. In this section of the dissertation, we present the results of a few experiments in image registration using local frequency representation to demonstrate that the proposed algorithm offers scene-independent registration performance and is efficient for different scenes ranging from complex, highly-varying gray-scale images to simpler, low-varying gray-scale images. Experimental results reported here indicate that the proposed registration technique is accurate and yields promising results for the alignment and fusion of complex images. In short, the purpose of this part of the experiments is to confirm our claim that the local frequency representation can be used to register images regardless the complexity of the structure of the image, in contrast with other schemes in the literature, which are application dependent.

All the images considered in these experiments were obtained from a public database [130] to enable checking of the reported results. In order to create ground truth data with which the registration parameters estimated by the present algorithm could be compared for a quantitative evaluation, the original image frame in each case was distorted by a known affine transformation in order to obtain a second frame that was then registered with the first (undistorted) frame.

In the first experiment, a "Cameraman" image was rotated by 4^{o} (using Matlab's function "maketform" to resample the image based on bicubic interpolation), with respect

to the standard Cameraman image shown in Fig. 3.15, in order to obtain the distorted frame shown in Fig. 3.18. The Cameraman image is of size 128 x 128. Estimating the registration parameters (rotation angle, in this case) between the two images by the proposed algorithm occurs in three steps. First, we obtain the local frequency representations for the two images using the procedure outlined in Algorithm 3.1. Then a set of control points are selected from each local frequency representation for execution of the matching step by isolating points with higher local frequency values. For the images under consideration, it was found that a set of 200 control points extracted from each local frequency representation is enough to cover all principal features in the image, as well as, to ensure that there are adequate numbers of matching pairs to implement the matching algorithm. Determination of an appropriate number of control points is generally dependent on the activity within the image (and hence is image-dependent). However, in all experiments that were performed, matching sets containing about 300 control points were found to be adequate to give reasonably accurate estimates of the registration parameters. It must be emphasized that selection of the least number of control points necessary for the specific image being processed is useful for minimizing the computational effort in the matching step. However, extracting a matching set that may have more than the required minimum number of points would eliminate user intervention and makes the registration process fully automatic for all considered images. In the third step, we apply the matching algorithm to establish correspondence between the two extracted sets and to estimate the transformation parameters at the same time. The estimated rotation angle in this experiment was 3.9^{o}, which agrees with the ground

true value quite well. Fig. 3.18(b) shows the distorted frame in Fig. 3.18(a) after it is registered by the estimated parameters (i.e. de-rotated by 3.9^o).

Fig. 3.19 shows the results of an experiment performed with a different type of image. The original image, of size 256 x 256 and shown in Fig. 3.19(a), is distorted by scaling to 85% of the original and a rotation of 3^o. The resulting distorted image is then registered with the original frame. The estimated registration parameters by the present algorithm in this case are 3.1^o for rotation and 0.85 for scaling. Fig. 3.19(b) shows the local frequency representation of the frame in Fig. 3.19(a), and the extracted control points for matching are shown in Fig. 3.19(c). The registered image after de-rotating by 3.1^o and prior to applying scaling is shown in Fig. 3.19(d).

To increase the challenge to the present algorithm, two other experiments with remotely sensed earth data were conducted. Fig 3.20(a) shows a 256 x 256 image of part of a city that was scaled by a factor of 0.9 (and no rotation) to obtain a distorted image, which was then registered with the original frame. The estimated registration parameters obtained were 0.901 for scaling and 0.07^o for rotation, which agree with the true value quite well. Fig. 3.20(b) shows the local frequency representation of the frame in Fig. 3.20(a). Fig. 3.21(a) shows another frame of remote sensing data of size 256 x 256. A distorted image, shown in Fig. 3.21(b), was obtained by rotating this frame by 6^o (and no scaling), which was then registered with the original frame. The estimated registration parameters obtained were 6.01^o for rotation and 1 for scaling. The distorted image after de-rotation by 6.01^o is shown in Fig. 3.21(c).

(a) (b)

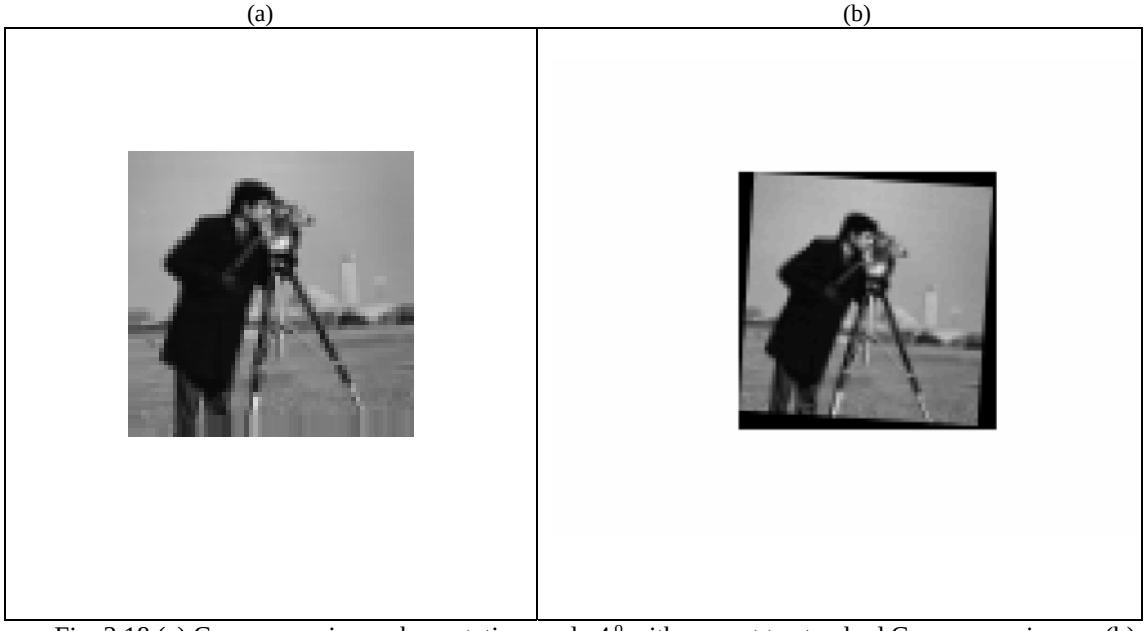
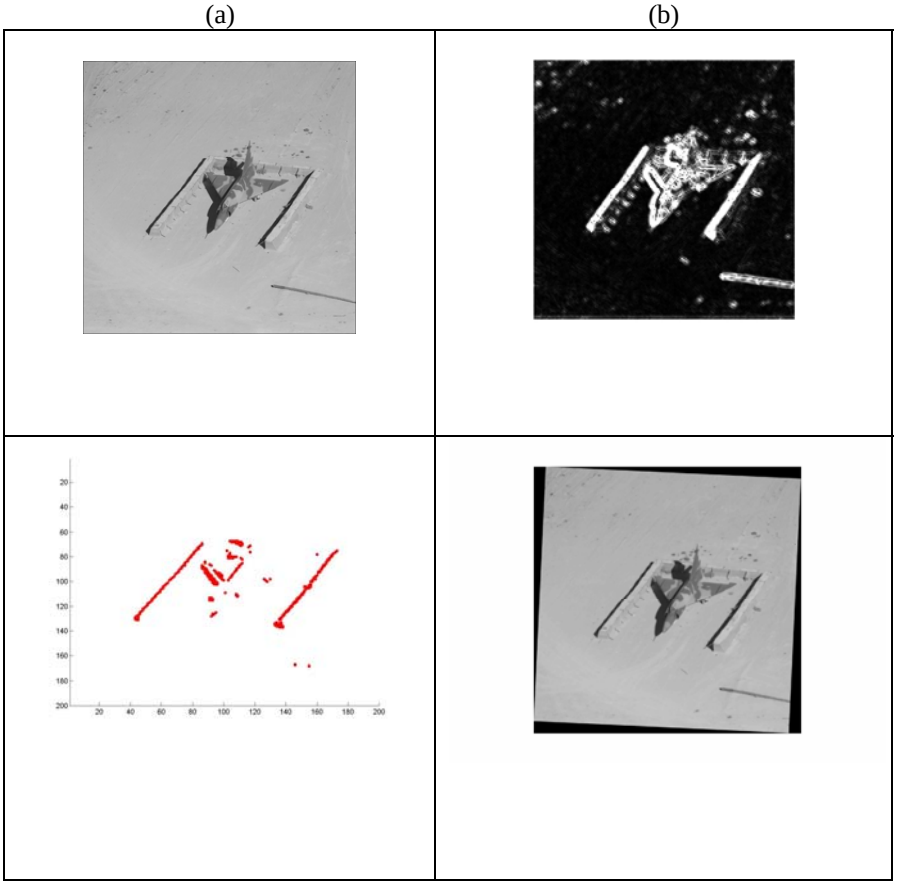


Fig. 3.18 (a) Cameraman image has rotation angle 4^{o} with respect to standard Cameraman image. (b) Cameraman image in (a) after it is de-rotated by the recovered angle 3.9^{o}

(a) (b)



(c) (d)

Fig. 3.19. (a) Original airplane image. (b). Local frequency representation. (c). The selected control points from the local frequency representation. (d). Distorted image registered with image in (a) by applying 3.1^{0} de-rotation.

(a) (b)

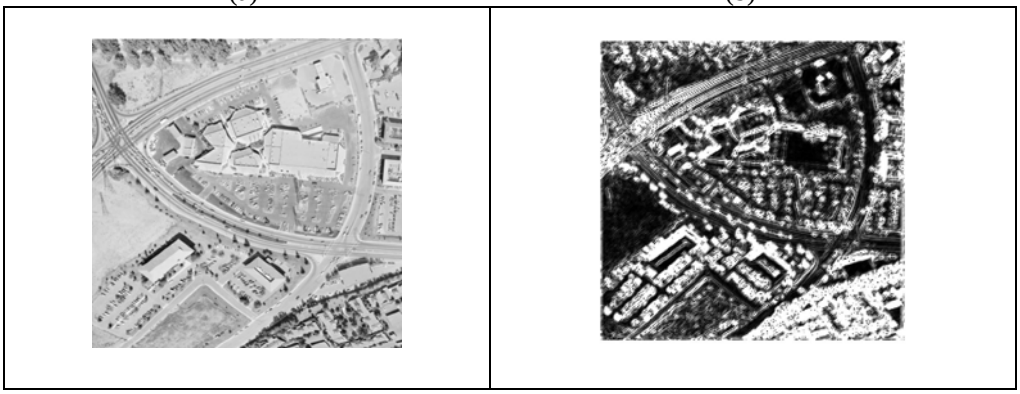
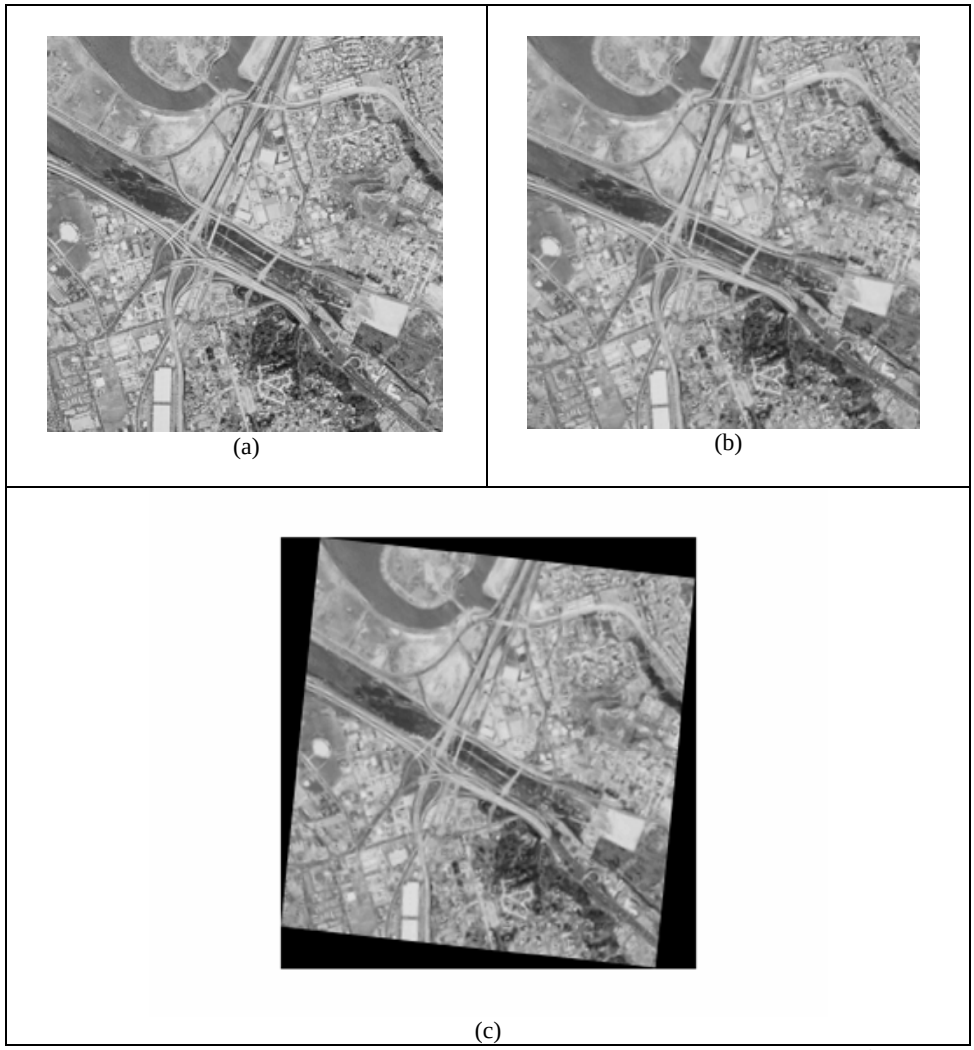


Fig. 3.20. (a) Aerial image of part of a city. (b) Local frequency representation for image in (a).



(a) (b) (c)

Fig. 3.21. (a) Another aerial image of part of a city. (b) Distorted image by applying 6^{o} rotation to the reference image. (c) The registered image obtained by applying 6.01^{o} de-rotation to image in (b).

CHAPTER 4

Extension of Local Frequency To Sub-Pixel Registration

4.1 Introduction

Recently, local frequency representation (LFR) has been employed to solve the problem of image registration at the pixel level, and the algorithm presented in the last chapter provides a solution to this problem. In this chapter, we propose an extension of using the LFR to solve the problem of sub-pixel image registration. The proposed method is based on using the normalized cross-correlation between the local frequency representations derived from the pair of images. A significant improvement in the accuracy of the results of the algorithm can be achieved using the so-called scaled local frequency representation (SLFR) of the image, instead of using LFR only. Scaled local frequency of an image is the local frequency representation of the image where the local frequency value at an individual spatial location is scaled by the image energy at that location. We evaluate the proposed algorithm using numerous experiments, some of which are reported here. These results indicate that SLF is superior to LF and yields comparable results as other state-of-the art sub-pixel image registration algorithms.

4.2 Background and Motivations

Image registration is an important preprocessing operation for the analysis and fusion of data in many applications. Pixel-level registration is adequate for many applications, however sub-pixel registration is essential for some important problems in biomedical

imaging and remote sensing. In this work, we assume that the images have pixel-level registration and we need to extend the accuracy of the registration to sub-pixel level. In addition, we assume that sub-pixel deformation between images is translation only.

There are several approaches that have been investigated for solving the problem of sub-pixel registration. Most of the methods are based on some form of interpolation. Examples include intensity interpolation and phase correlation interpolation [117-119]. One commonly used method is based on the differential properties of the image [120,121] under an intensity conservation assumption between images [122]. This assumption sometimes is violated in real images. Other methods solved the problem by employing optimization techniques [123,124]. Again these methods are based on an intensity conservation assumption, which is violated if the luminance has changed between the images, or if the images are from different spectral bands. Recently, the authors in [125,126] extended the phase correlation (PC) method to sub-pixel registration. In a slightly modified procedure, the authors in [127, 128] used a gradient correlation (GC) method to solve the problem and they reported superior results than the phase correlation technique in a few experiments. The gradient correlation technique is based on the cross-correlation between the computed spatial gradients for the pair of images. In other words, the method is based on using correlation between salient features in both images, such as the horizontal and vertical gradients.

Motivated by these developments, we propose extension of the local frequency representation to solve the problem of sub-pixel image registration. The LFR is successfully used to solve the problem of pixel-level image registration [129,95]. However, using LFR to solve the problem of sub-pixel registration is not investigated yet in the image registration literature. The proposed technique is based on computing the phase correlation between the local frequency representations of images.

All the advantages of local frequency representations make it a promising candidate in solving the problem of image registration. However, the results from the conducted experiments were not satisfactory when compared with the state-of-the art method, Foroosh's method [125] that employs phase correlations between the images to be registered.

The unsatisfactory results from directly using local frequency representations provided a motivation to improve the proposed scheme. The idea of improvement is based on using more information from the image in addition to the local frequency and its orientation in the representation of the image. The image energy at various spatial locations is used as this additional information. Which is used to construct the so-called scaled local frequency (SLF) of the image. Then, the SLF representation is employed to solve the problem of sub-pixel image registration. The results of conducted experiments show that using SLF is superior to LF and offers results which compare favorably with Foroosh's method.

4.3 Estimation of Local Frequency for Sub-Pixel Registration

A practically efficient procedure for computing the local frequency representation of a given image was developed in Chapter 3. This procedure utilizes some important recent findings inspired by biological data (efficiency of the model human image code and orientation bandwidth of visual cortex cells) of splitting up the two-dimensional spatial frequency plane of a given image into 4 orientation bands, with orientation bandwidths of 45 degrees (to cover the 2-D frequency plane of the image), in order to construct a multi-channel filtering scheme. For the proposed sub-pixel image registration algorithm, we adapted the four-filter scheme presented in Section 3.2 to build a more efficient algorithm by including the orientation of the local frequency in addition to the magnitude of the local frequency to obtain more information about the image. Hence, the orientation and the magnitude of the local frequency are employed in one representation for the image. Algorithm 3.1 presented in Section 3.3.3 will be employed for this purpose with some modifications as follows:

Consider the analytic signal in Eq. 3.14 represented in polar form as

$$u^{(k)}(x,y) = |u^{(k)}(x,y)|\, e^{j\psi^{(k)}(x,y)}. \tag{4.1}$$

Then, the local frequency estimation in x and y directions, $\frac{\partial \psi^{(k)}(x,y)}{\partial x}$, $\frac{\partial \psi^{(k)}(x,y)}{\partial y}$, can be obtained from the analytic signal $u^{(k)}$, k=1,..,4, without explicitly computing the

phase as follows:

$$\frac{\partial \psi^{(k)}(x,y)}{\partial x} = \frac{imag\left\{u^{(k)^{\bullet}}(x,y)\partial u^{(k)}(x,y)/\partial x\right\}}{\left|u^{(k)}(x,y)\right|^2},$$

$$\frac{\partial \psi^{(k)}(x,y)}{\partial y} = \frac{imag\left\{u^{(k)^{\bullet}}(x,y)\partial u^{(k)}(x,y)/\partial y\right\}}{\left|u^{(k)}(x,y)\right|^2}, \qquad (4.2)$$

where $\bullet$ denotes the complex conjugate.

Proof.

Consider the analytic signal $u^{(k)}$ in Eq. 4.1

$$u^{(k)}(x,y) = \left| u^{(k)}(x,y) \right| \mathrm{e}^{(j\psi^{(k)}(x,y))}$$

Computing the partial derivative with respect to x

$$\frac{\partial u^{(k)}(x,y)}{\partial x} = \frac{\partial \left| u^{(k)}(x,y) \right|}{\partial x} \mathrm{e}^{j\psi^{(k)}(x,y)} + \left| u^{(k)}(x,y) \right| j \frac{\partial \varphi^{(k)}(x,y)}{\partial x} e^{j\psi^{(k)}(x,y)}$$

multiplying by $e^{(-j\psi^{(k)}(x,y))}$ and rearranging

$$\left\{ \frac{\partial u^{(k)}(x,y)}{\partial x} e^{(-j\psi^{(k)}(x,y))} - \frac{\partial \left|u^{(k)}(x,y)\right|}{\partial x} \right\} \Big/ \left|u^{(k)}(x,y)\right| = j \frac{\partial \psi^{(k)}(x,y)}{\partial x}$$

multiplying the left side by $\dfrac{\left|u^{(k)}(x,y)\right|}{\left|u^{(k)}(x,y)\right|}$

$$\frac{\left|u^{(k)}(x,y)\right|}{\left|u^{(k)}(x,y)\right|}\left\{\frac{\partial u^{(k)}(x,y)}{\partial x}e^{(-j\psi^{(k)}(x,y))}-\frac{\partial\left|u^{(k)}(x,y)\right|}{\partial x}\right\}\Big/\left|u^{(k)}(x,y)\right|=j\frac{\partial\psi^{(k)}(x,y)}{\partial x}$$

$$\left\{\frac{\partial u^{(k)}(x,y)}{\partial x}\left|u^{(k)}(x,y)\right|e^{(-j\psi^{(k)}(x,y))}-\left|u^{(k)}(x,y)\right|\frac{\partial\left|u^{(k)}(x,y)\right|}{\partial x}\right\}\Big/\left|u^{(k)}(x,y)^2\right|=j\frac{\partial\psi^{(k)}(x,y)}{\partial x}$$

$$imag\left\{\ \left\{\frac{\partial u^{(k)}(x,y)}{\partial x}\left|u^{(k)}(x,y)\right|e^{(-j\psi^{(k)}(x,y))}-\left|u^{(k)}(x,y)\right|\frac{\partial\left|u^{(k)}(x,y)\right|}{\partial x}\right\}\Big/\left|u^{(k)}(x,y)^2\right|\ \right\}=\frac{\partial\psi^{(k)}(x,y)}{\partial x}$$

$$imag\left\{\ \left\{\frac{\partial u^{(k)}(x,y)}{\partial x}\left|u^{(k)}(x,y)\right|e^{(-j\psi^{(k)}(x,y))}\right\}\Big/\left|u^{(k)}(x,y)^2\right|\ \right\}=\frac{\partial\psi^{(k)}(x,y)}{\partial x}$$

$$imag\left\{\ \left\{u^{(k)}(x,y)^{\bullet}\frac{\partial u^{(k)}(x,y)}{\partial x}\right\}\Big/\left|u^{(k)}(x,y)\right|^2\ \right\}=\frac{\partial\psi^{(k)}(x,y)}{\partial x}$$

Similarly we can prove $\dfrac{\partial\psi^{(k)}(x,y)}{\partial y}$.

Having obtained the local frequency in both *x*-direction and *y*-direction for each filter G_k, k=1,..,4, we combine them in one complex representation to retain the magnitude and the orientation of the local frequency estimated by each filter G_k, k=1,..,4, as

$$\Gamma_o^{(k)}(x,y)=\frac{\partial\psi^{(k)}(x,y)}{\partial x}+j\frac{\partial\psi^{(k)}(x,y)}{\partial y}, \tag{4.3}$$

where $j=\sqrt{-1}$. It is easy to note that Γ_o^k gives the local frequency in the rectangular form

and hence includes both magnitude and orientation of the local frequency estimate.

We then fuse the results of the four filters to construct one combined representation of the LF estimate, Γ_o using

$$\Gamma_o(x,y) = \mathit{fuse}(\Gamma_o^{(1)}(x,y), \Gamma_o^{(2)}(x,y), \Gamma_o^{(3)}(x,y), \Gamma_o^{(4)}(x,y)) \,, \tag{4.4}$$

where the *fuse* operator is defined by Eq. 3.18 in Algorithm 3.1.

4.4 Sub-Pixel Estimation Based on Local Frequency

Phase correlation (PC) is a computationally efficient method for translation estimation $(\Delta x, \Delta y)$ between a pair of images sharing some mutual support. For a brief description of this method let $I_1(x,y)$ and $I_2(x,y)$ be two given images and their mutual support satisfies the following equation:

$$I_2(x,y) = I_1(x+\Delta x, y+\Delta y) \,. \tag{4.5}$$

Using the *Fourier shift theorem*, we have

$$\begin{aligned} FFT2\{I_2(x+\Delta x, y+\Delta y)\} &= \hat{f}_2(\omega_x, \omega_y) \,, \\ &= \hat{f}_1(\omega_x, \omega_y) e^{j(\omega_x \Delta x + \omega_y \Delta y)} \end{aligned} \tag{4.6}$$

where $\hat{f}_1(\omega_x, \omega_y) = FFT2\{I_1(x,y)\}$ and *FFT2* stands for the fast Fourier transform in 2-D. The normalized cross power spectrum is then given by

$$\frac{\hat{f}_2(\omega_x,\omega_y)\,\hat{f}^{\bullet}_1(\omega_x,\omega_y)}{\left|\hat{f}_2(\omega_x,\omega_y)\,\hat{f}^{\bullet}_1(\omega_x,\omega_y)\right|} = e^{j(\omega_x\Delta x+\omega_y\Delta y)}. \tag{4.7}$$

where $\bullet$ denotes the complex conjugate. Foroosh et al [125] show analytically how to compute the sub-pixel translation from Eq. 4.7.

To employ Eq. 4.7 for obtaining the sub-pixel shifts between images using the LFR for both images, we obtain the fast Fourier transform of the local frequency representation for each image and compute the normalized cross power spectrum as

$$\frac{FFT2(\Gamma_{o_2}(x,y))FFT2^{\bullet}(\Gamma_{o_1}(x,y))}{\left|FFT2(\Gamma_{o_2}(x,y))FFT2^{\bullet}(\Gamma_{o_1}(x,y))\right|} = e^{j(\omega_x\Delta x+\omega_y\Delta y)} \tag{4.8}$$

where Γ_{o_1} and Γ_{o_2} are the local frequency representation for the images I_1 and I_2, respectively. The sub-pixel translations Δx and Δy can then be evaluated using routine calculations.

4.4.1 Accuracy Evaluation of the Proposed Algorithm

To evaluate the proposed algorithm, we conducted an experiment using “Aerial” image from [130]. The image is shown in Fig. 4.2. The experiment is conducted in two steps. The first step is to generate a ground truth data set and the second step is to use the ground truth data to evaluate the proposed algorithm.

Ground Truth

In order to evaluate the accuracy of the proposed algorithm, we prepared ground truth using the scheme described below. The idea is to use a high-resolution image, I, and create down-sampled images of low-resolution. Creation of low-resolution images is accomplished in two steps as follows:

Step 1: Convolve the image, I, with a Gaussian filter Ga_w with kernel w. The function of the Gaussian filter is to reduce the aliasing in the down-sampled images.

$$I_{ga} = Ga_w * I \tag{4.9}$$

where I_{ga} is the image I after convolution.

Step 2: Synthesize sub-pixel shifts by down-sampling I_{ga} by factor l ($w > l$) in x-direction and in y-direction. The down-sampled images are shifted with respect to each other by integer amounts x_o and y_o in the high- resolution grid. After down sampling by factor of l in each dimension, the relative shifts become sub-pixel shifts of size x_o/l and y_o/l, respectively.

Using "Aerial" image of 512 x 512 size and a down sampling factor $l = 8$, we synthesized a set of 64 testing images, $\left\{ I_{down\_sampled_i} \right\}_{i=1}^{64}$, of size 64 x 64. The synthesized sub-pixel shifts are (1/8;2/8;3/8;4/8;5/8;6/8;7/8). We selected an arbitrary image $I_{down\_sampled_r}$ from the set of the synthesized images and we designated it as a

reference image for the rest of the image set. Then, we used the proposed algorithm to estimate the sub-pixel shifts between the selected reference image and the rest of images in both x and y directions. From the results, we computed the average of the absolute error of the estimated sub-pixel shifts for all images within the set relative to the reference according to

$$e = \frac{1}{2n}\sum_{i=1}^{n}\left|dx_i - dx_r\right| + \left|dy_i - dy_r\right|, \tag{4.10}$$

where n is the cardinal number of the synthesized image set excluding the reference image, $I_{down\_sampled_r}$, dx_r and dy_r are the synthesized sub-pixel shifts between the images in x and y directions, respectively relative to the chosen reference (note that they take values from the set (1/8, 2/8, 3/8, 4/8, 5/8, 6/8,7/8)). dx_i and dy_i are the estimated sub-pixel shifts between the reference image and image i in x and y directions, respectively. We plotted the results in Fig. 4.3 and we compared them with the results of the phase correlation (PC) method [125]. Investigating the graphs, the average of the absolute error of the estimated sub-pixel shifts using PC is much smaller than the average of the absolute error of the estimated sub-pixel shifts using LF representation.

Figure 4.1. “Aerial” image

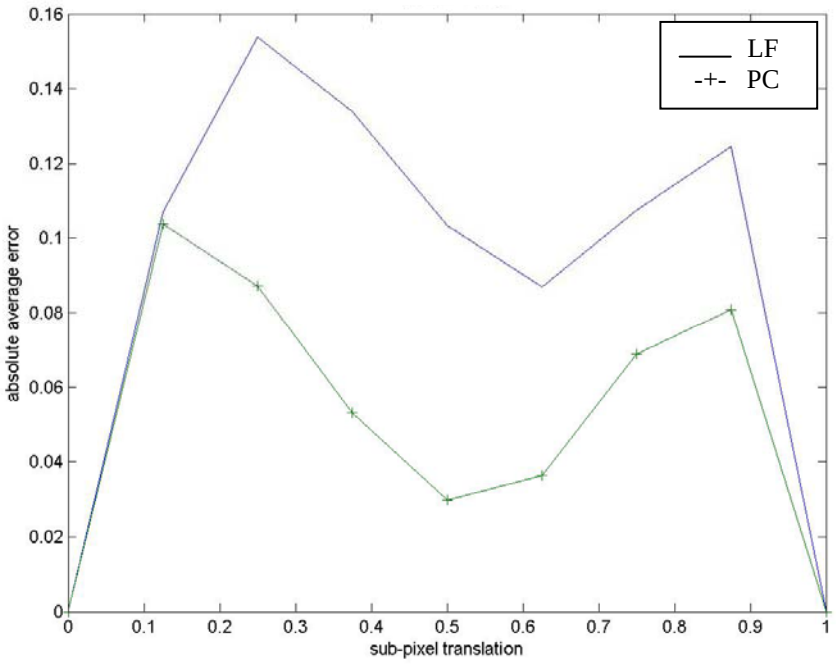


Figure 4.2. Comparison of the average of the absolute error for the estimated sub-pixel shifts using the proposed algorithm and the phase correlation algorithm.

4.5 Sub-Pixel Estimation Based on Scaled Local Frequency

To further improve the accuracy of the proposed algorithm we incorporated additional information from the images in Eq. 4.8. in the form of image energy at individual spatial locations. The image energy with the local frequency representation defines the so-called scaled local frequency (SLF). The scaled local frequency in the *x*-direction using the k[th] filter is defined by

$$\Omega_x^{(k)}(x,y) = \frac{\partial \psi^{(k)}(x,y)}{\partial x}.\left|u^{(k)}(x,y)\right|^2$$

$$= imag\left\{ \left\{ u^{(k)}(x,y)^{\bullet} \frac{\partial u^{(k)}(x,y)}{\partial x} \right\} \right\}, \tag{4.11}$$

where $\frac{\partial \psi^{(k)}(x,y)}{\partial x}$ is the LF for the image at the spatial position (x,y) in the *x*-direction using the k[th] filter and $\left|u^{(k)}(x,y)\right|^2$ is the energy of the image, $E^{(k)}(x,y)$, at the spatial position (x,y) using the k[th] filter. Similarly scaled local frequency in the *y*-direction using the k[th] filter is defined by

$$\Omega_y^{(k)}(x,y) = \frac{\partial \psi^{(k)}(x,y)}{\partial y}.\left|u^{(k)}(x,y)\right|^2$$

$$= imag\left\{ \left\{ u^{(k)}(x,y)^{\bullet} \frac{\partial u^{(k)}(x,y)}{\partial y} \right\} \right\}. \tag{4.12}$$

The scaled local frequency estimate formed by the k[th] filter can then be described in the rectangular form, as

$$\Omega^{(k)}(x,y) = \Omega_x^{(k)}(x,y) + j\Omega_y^{(k)}(x,y) . \quad (4.13)$$

The scaled local frequency estimates from all the filters, $\Omega^{(k)}$, k=1,..,4, are then fused to construct one combined scaled local frequency representation for the image, Ω. The *fuse* operator in Eq. 3.18 will be employed again using the scaled local frequency instead of the local frequency.

To employ SLF for an image to estimate the sub-pixel shifts, Eq. 4.8 will take the form

$$\frac{FFT2(\Omega_2(x,y))FFT2^{\bullet}(\Omega_1(x,y))}{\left|FFT2(\Omega_2(x,y))FFT2^{\bullet}(\Omega_1(x,y))\right|} = e^{j(\omega_x\Delta x+\omega_y\Delta y)} \quad , \quad (4.14)$$

where Ω_1 and Ω_2 are the scaled local frequency representations for the two images I_1 and I_2 , respectively.

While an analytical proof that SLF is superior over LF (in solving the problem of sub-pixel registration) is not possible, simulation studies can be used to justify the above selection.

To show that SLF is superior to LF in dealing with the problem of sub-pixel image registration, we conducted again the previous experiment using “Aerial” image, but we used SLF representation to replace LF representation. Then, we estimated the absolute average error using Eq. 4.10. Fig. 4.3 shows the results of using SLF compared with results of using LF. Using SLF, the error is up to 30% lower.

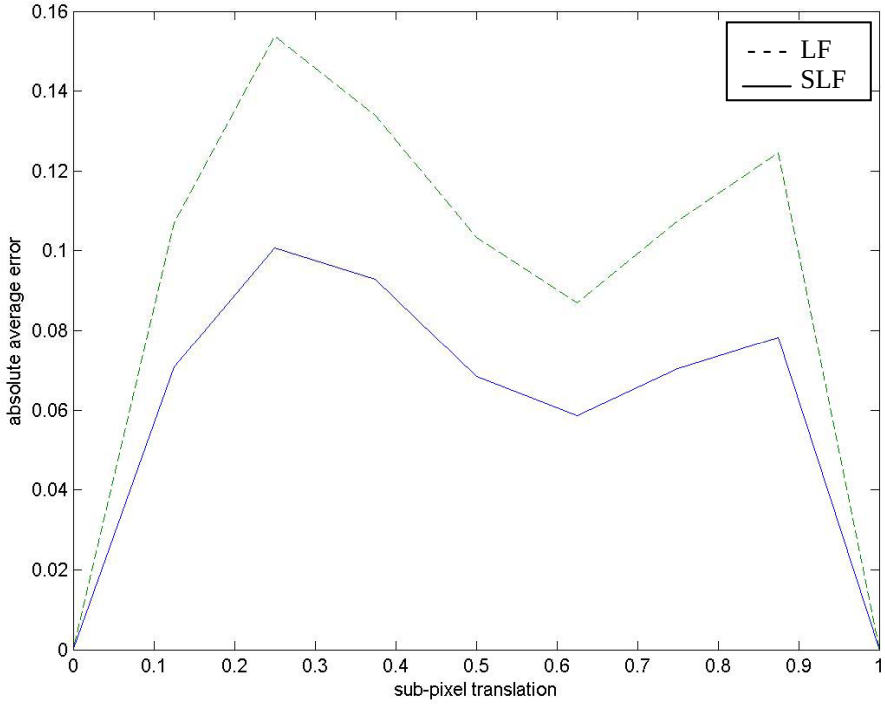


Figure 4.3. Comparison of average of the absolute error for the estimated sub-pixel shifts using SLF representation and using LF representation. The units of absolute average error in pixels.

4.6 Experimental Results

In order to evaluate the precision of the proposed algorithm, we prepared a controlled simulation using the scheme described in Section 4.3. The method involves using a single high-resolution image to represent the scene and use it to create a set of smaller size images such that there is known sub-pixel shifts between each pair of two images. Then we use the known sub-pixel shift as ground truth to compare with the estimated sub-pixel shifts by the proposed algorithms.

We conducted two types of experiments to evaluate the proposed algorithms. In the first type of experiments, we compared the estimated sub-pixel shifts with the ground truth data by calculating the average absolute error. In the second type of experiments, we used the estimated sub-pixel shifts by the proposed algorithms to construct the super-resolution image from the down- sampled images set. Then we calculated peak signal to noise ratio (PSNR) for each constructed image to evaluate quantitatively the proposed algorithms.

4.6.1 Performance of LF and SLF Algorithms

In the first two experiments, we compared the precision of the proposed algorithms. In the first experiment, we used “Lenna” image of size 512 x 512 to create the set of down-sampled images. Each down-sampled image is of size 64 x 64. The ground truth data of sub-pixel shifts for each low-resolution image with respect to another low-resolution image in the set comprises of two values from (1/8, 2/8, 3/8, 4/8, 5/8, 6/8,7/8) for the x and y directions. We selected one image from the low-resolution image set as a

reference image in estimating the sub-pixel shifts. Using Eq. 4.10, we calculated the absolute average error of the estimated sub-pixel shifts using the LF algorithm and then using the SLF algorithm. The results of this experiment appear in Fig. 4.5. It can be seen that the result of the SLF algorithm is superior to the result from LF algorithm.

In the second experiment we used “Plane” image of size 512 x 512 to create the low-resolution images. The down sampling factor is 8 and the low-resolution image of size 64 x 64. Fig. 4.6 shows the results. Once again the average absolute error with use of SLF algorithm is less than the average absolute error using the LF algorithm at all sub-pixel shifts.

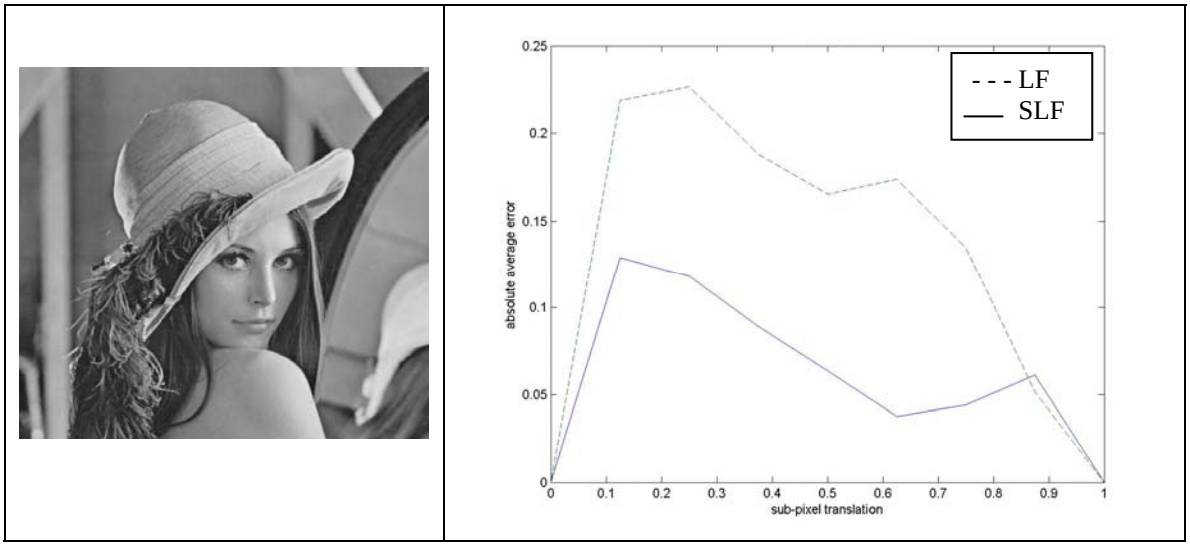


Fig. 4.4. (a) “Lenna” image, (b) comparison of absolute average error of the estimated sub-pixel shifts using LF algorithm and using SLF algorithm.

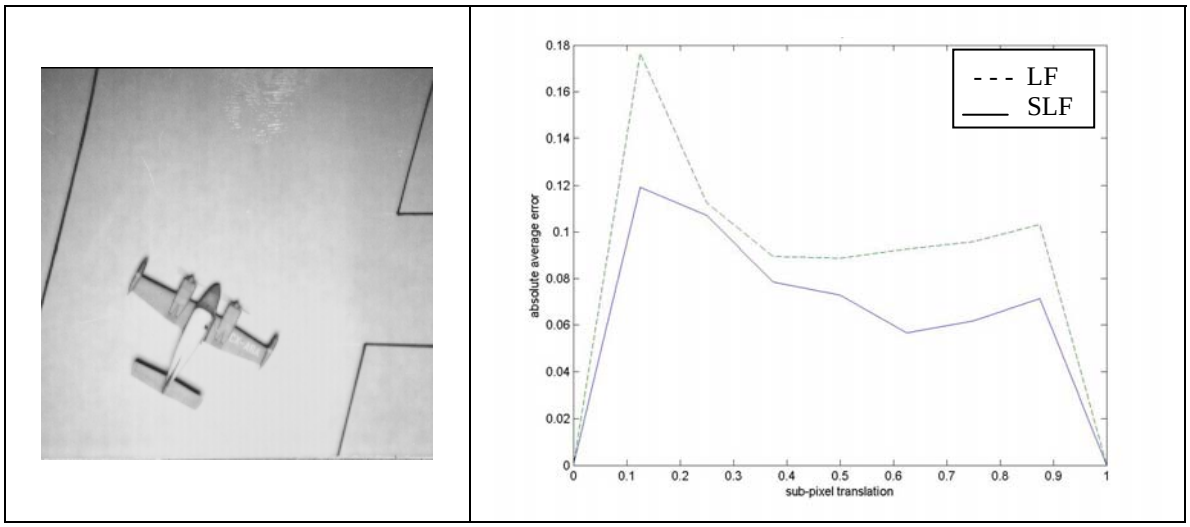


Fig. 4.5. (a) “Plane” image, (b) comparison of absolute average error of the estimated sub-pixel shifts using LF algorithm and using SLF algorithm.

Influence of noise on LF and SLF performances

At the end, we conducted a different experiment to show the influence of noise on the performance of the SLF and LF algorithms in sub-pixel image registration. For this purpose, we synthesized sub-pixel shifts between set of images generated from “Man” image, Fig. 4.7, as before and we used the proposed algorithm to estimate these sub-pixel shifts. In the process of synthesizing, we added Gaussian noise with zero mean to the image “Man” before the down-sampling process. We added noise with different level values such that the signal-to-noise ratio (SNR) varies from (5 dB to 20 dB), where SNR (in dB) is defined by, $SNR(dB) = 10\log\frac{\sigma_s^2}{\sigma_n^2}$ σ_s, σ_n are the standard deviations for the signal (image) and the added noise, respectively. We then estimated the sub-pixel shifts by the developed algorithms and we computed the average of the error. Fig. 4.8 shows the computed average error against SNR. Obviously, performance of the SLF is better, however the error in the estimated sub-pixel shifts for both algorithms is decreasing with increasing in the SNR.

Fig. 4. 6 “Man” Image

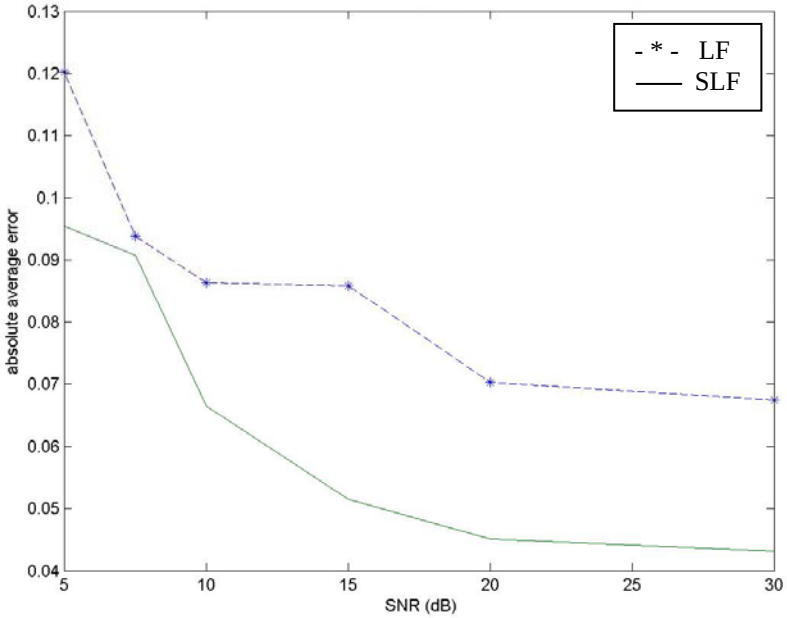


Fig.4.7 The average error vs. SNR

4.6.2 Comparison with Other Sub-pixel Registration Algorithms

This section presents a comparison of the proposed algorithm with the GC algorithm [127] and the PC algorithm [125]. For this purpose, we conducted two types of experiments. In the first type of experiments, a set of low-resolution images is generated as described above in Section 4.3. Then, each algorithm is used to estimate the sub-pixel shifts between the images and the average absolute error is calculated for the results of each algorithm. Figure 4.9 shows the results for both “Lenna” and “Plane” images. The performance of the proposed algorithm is comparable with the performance of PC and is better than the performance of GC.

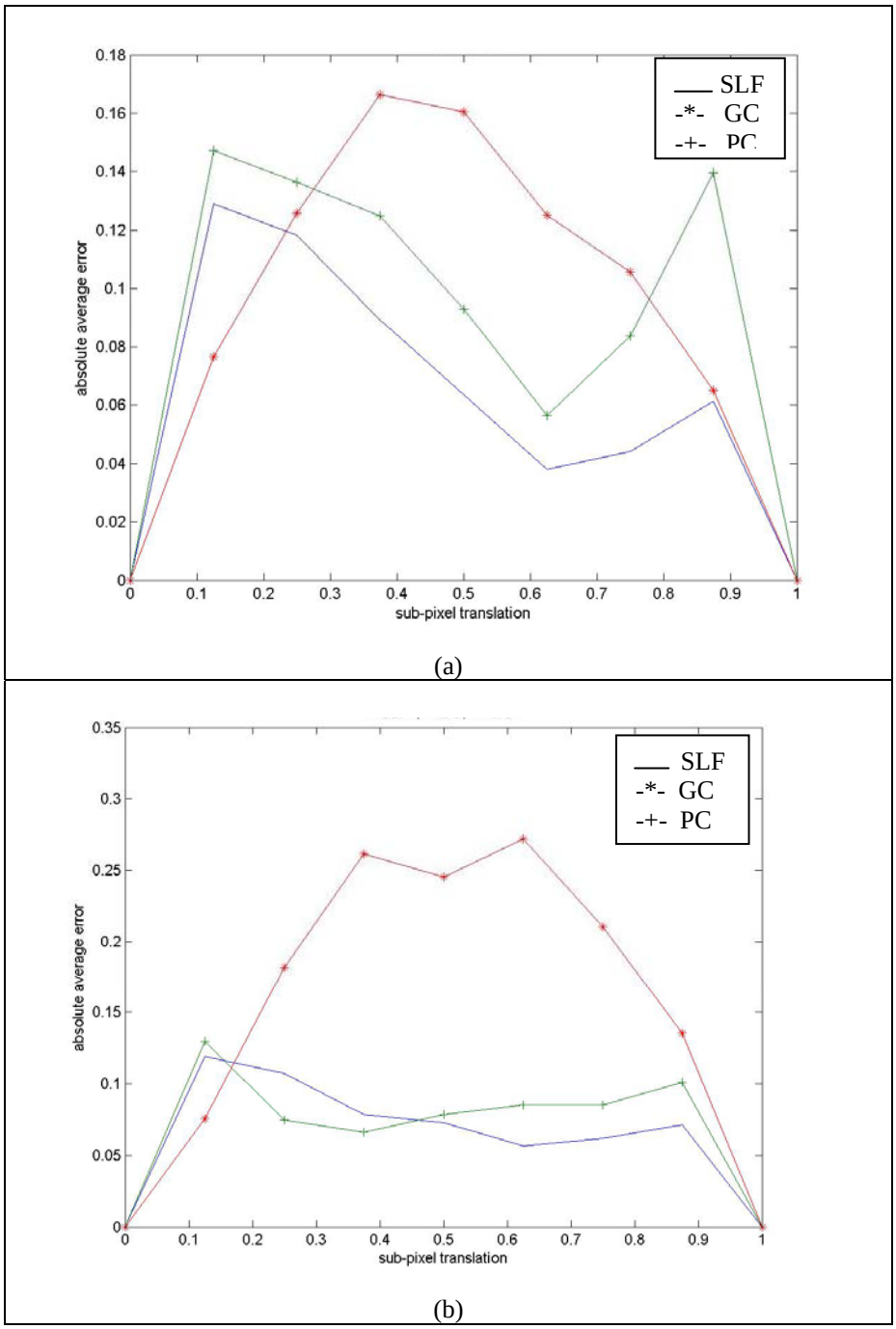


Fig. 4.8. (a) Using "Lenna" image, a comparison of absolute average error of the estimated sub-pixel shifts using PC, GC, and SLF algorithms. (b) Using "Plane" image, a comparison of absolute average error of the estimated sub-pixel shifts using PC, GC, and SLF algorithms.

4.7 Super-Resolution Imaging.

In all image-based systems the demand for high-resolution images is gradually increasing. However, in many imaging systems, the quality of image resolution is limited by physical constraints. The imaging systems yield aliased and under-sampled images if their detector array is not sufficiently dense. Digital image processing approaches have been investigated to reconstruct a high-resolution image from aliased low-resolution images. Super-resolution is defined as the processing of an image so as to recover object information from beyond the spatial frequency bandwidth of the optical system that formed the image. Super-resolution is an essential operation for performing better diagnosis in medical applications and is necessary for surveillance and observation systems. Super-resolution is accomplished by using multi-frames for the same scene with sub-pixel shifts between them. The task of registration techniques in super-resolution imaging is to estimate the sub-pixel shifts between those frames of the scene, so that the high-resolution frame can be reconstructed by summing the low-resolution frames.

4.7.1 Importance of Sub-Pixel Estimation to Super-Resolution Imaging

It is known that a high-resolution image can be estimated from a sequence of low-resolution frames with sub-pixel motion between them. These frames are captured by mounting an imager on a moving or a vibrating platform (uncontrolled microscanning). Thus, estimating the sub-pixel shift between the frames is an essential step towards implementing super-resolution algorithms. This is obvious by investigating the model in Eq. 4.15, where all the matrices are known except the matrix F_k, which stands for the

geometric warp operation that exists between the low resolution images $\left\{\overline{Y_k}\right\}_{k=1}^{N}$. So, implementing an algorithm for super-resolution and using the proposed algorithm for estimating the sub-pixel registration parameters in order to compute F_k is the goal in this section.

4.7.2 Super-Resolution Imaging Model

In this section, we briefly present the super-resolution model. Given N measured images, $\left\{\overline{Y_k}\right\}_{k=1}^{N}$, for the same scene such that there exist sub-pixel shifts between each pair of images. These images are employed to form a single high-resolution image, denoted as $\overline{X}$. The images with low resolution and the high resolution image are related by the following equation

$$\overline{Y_k} = D_k H_k F_k \overline{X} + \overline{V_k} \quad , \quad k = 1,...,N \ , \tag{4.15}$$

where the matrix D_k stands for the decimation operation, representing the reduction of the observed pixels in the low resolution images. The matrix H_k is the blurring matrix, representing the point spread function (PSF) of the camera. The matrix F_k stands for the geometric warp operation that exists between the high-resolution image vector $\overline{X}$ and the interpolated version of the low-resolution image vector $\overline{Y_k}$. The vectors $\left\{\overline{V_k}\right\}_{k=1}^{N}$ represent Gaussian additive noise with zero mean and the auto-correlation matrix given by $E\left\{ \overline{V_k}\ \overline{V_k}^T \right\}$. In most applications the noise is assumed to be white, which means that

$$E\{\ \overline{V_k}\ \overline{V}_k^T\ \} = \sigma^2 I\ .$$

In fact, the problem of super-resolution reconstruction from summing multiple frames was first addressed in the literature by Tsai and Huang [131] and since then the problem is well known and extensively treated in the literature [132 - 141]. The problem is known to be an inverse problem, where an unknown image is to be reconstructed, given low-resolution images related to it through linear operators and additive noise as shown in the above model. In our work, we adopted the algorithm given by [134] to reconstruct the high-resolution image from given low-resolution images in order to evaluate the proposed sub-pixel image registration. The employed super-resolution reconstruction algorithm assumes that the blur is space-invariant and is the same for all measured low-resolution images, the geometric warps between the low-resolution images are pure translations, and the additive noise is white. The algorithm can be briefly outlined as follows:

4.7.3 Fast Super-Resolution Image Reconstruction Algorithm

The Maximum-likelihood estimation of $\overline{X}$ can be obtained through the following least squares expression

$$\hat{\overline{X}} = Arg\underset{\overline{X}}{Min}\left\{\sum_{k=1}^{N}[\overline{Y_k} - D_k H_k F_k \overline{X}]^T W_k^{-1}[\overline{Y_k} - D_k H_k F_k \overline{X}]\right\}$$

(4.16)

Taking the first derivative of the objective function in Eq. 4.16 with respect to $\overline{X}$ and equating to zero yields

$$R\overline{\hat{X}} = \overline{P} \tag{4.17}$$

where $R = \sum_{k=1}^{N} F_k^T H_k^T D_k^T W_k^{-T} D_k H_k F_k$ and $\overline{P} = \sum_{k=1}^{N} F_k^T H_k^T D_k^T W_k^{-1} \overline{Y}_k$.

Solving $R\overline{\hat{X}} = \overline{P}$ directly is not practically possible due to its size. For example, if the size of image $\overline{\hat{X}}$ is 100 x 100 pixels, the matrix R is of size 10000 x 10000, which requires huge memory consumption. Another approach to solve for $\overline{\hat{X}}$ is using iterative methods. One common and simple method is using the steepest descent (SD) algorithm. Using the SD algorithm, $\overline{\hat{X}}$ is estimated as following

$$\overline{\hat{X}}_{j+1} = \overline{\hat{X}}_j + \mu[\overline{P} - R\overline{\hat{X}}_j], \tag{4.18}$$

where $\overline{\hat{X}}_0$ is any initialization vector.

The algorithm assumes the following properties.

1) The decimation operations are assumed to be the same, i.e. $\forall k, D_k = D$
2) The blur operations are assumed to be the same, i. e. $\forall k, H_k = H$. Moreover, H is assumed to be block circulant, representing a linear space invariant blur.
3) All the warp operations are pure translations.
4) The additive noise is white and is the same for all the measurements, i. e. $\forall k, W_k = \sigma^2 I$

Using these assumptions, the updating equation in Eq. 4. 18 can be rewritten as

$$\overline{\hat{X}}_{j+1} = \overline{\hat{X}}_j + \mu \sum_{k=1}^{N} F_k^T H^T D^T [\overline{Y}_k - DHF_k \overline{\hat{X}}_j] \quad . \tag{4.19}$$

Exploiting the fact that block circulating matrices commute yields that $F_k^T H^T = H^T F_k^T$ and $HF_k = F_k H$, and hence Eq. 4.19 becomes

$$\overline{\hat{X}}_{j+1} = \overline{\hat{X}}_j + \mu H^T \sum_{k=1}^{N} F_k^T D^T [\overline{Y}_k - DF_k H \overline{\hat{X}}_j] \ . \tag{4.20}$$

Multiply both sides by H and define $\overline{\hat{Z}}_j = H \overline{\hat{X}}_j$ to obtain the updating equation

$$\overline{\hat{Z}}_{j+1} = \overline{\hat{Z}}_j + \mu HH^T \sum_{k=1}^{N} F_k^T D^T [\overline{Y}_k - DF_k \overline{\hat{Z}}_j] \tag{4.21}$$

$$\overline{\hat{Z}}_{j+1} = \overline{\hat{Z}}_j + \mu HH^T [\tilde{\overline{P}} - \tilde{R} \overline{\hat{Z}}_j], \tag{4.22}$$

where $\tilde{\overline{P}} = \sum_{k=1}^{N} F_k^T D^T \overline{Y}_k$ and $\tilde{R} = \sum_{k=1}^{N} F_k^T D^T DF_k$.

Assuming that we somewhat found $\overline{\hat{Z}}_\infty$ from Eq 4.22, then we must apply an image restoration process in order to remove the effect of the blur matrix H, this is based on $\overline{\hat{Z}}_\infty = H \overline{\hat{X}}_\infty$ and then we get $\overline{\hat{X}}_\infty$. More details on this reconstruction algorithm can be found in [134].

4.8 Experimental Results

The goal in this part of the experiments, is to measure the performance of the proposed sub-pixel estimation algorithms quantitatively in super-resolution reconstruction. For this purpose, we implemented the super-resolution reconstruction

algorithm outlined in the previous section. We used a set of low-resolution images obtained from “Lenna” and we conducted two experiments. In one experiment, we used the LF algorithm to estimate the sub-pixel shifts in the super-resolution reconstruction algorithm. In the second experiment, we used the SLF algorithm to replace the LF algorithm to estimate the sub-pixel shifts. Having obtained the super-resolution image in both experiments, we calculated PSNR in dB for each super-resolution reconstructed image. PSNR is defined by

$$\mathbf{PSNR} = 10\log_{10}\left\{\frac{255^2.N}{\left\|\hat{I}-I\right\|^2}\right\}, \tag{4.23}$$

where N is the total number of pixels in the high-resolution image; $\hat{I}$ and I are the reconstructed super-resolution image and the original high-resolution image, respectively. The norm $\left\|\hat{I}-I\right\|$ is defined as $\left\|\hat{I}-I\right\| = \sqrt{\sum_{i,j\in N}(\hat{I}(i,j)-I(i,j))^2}$. The calculated PSNR using the LF algorithm to estimate the sub-pixel shifts is 23.6642 dB, while PSNR is 25.6309 dB using the SLF algorithm.

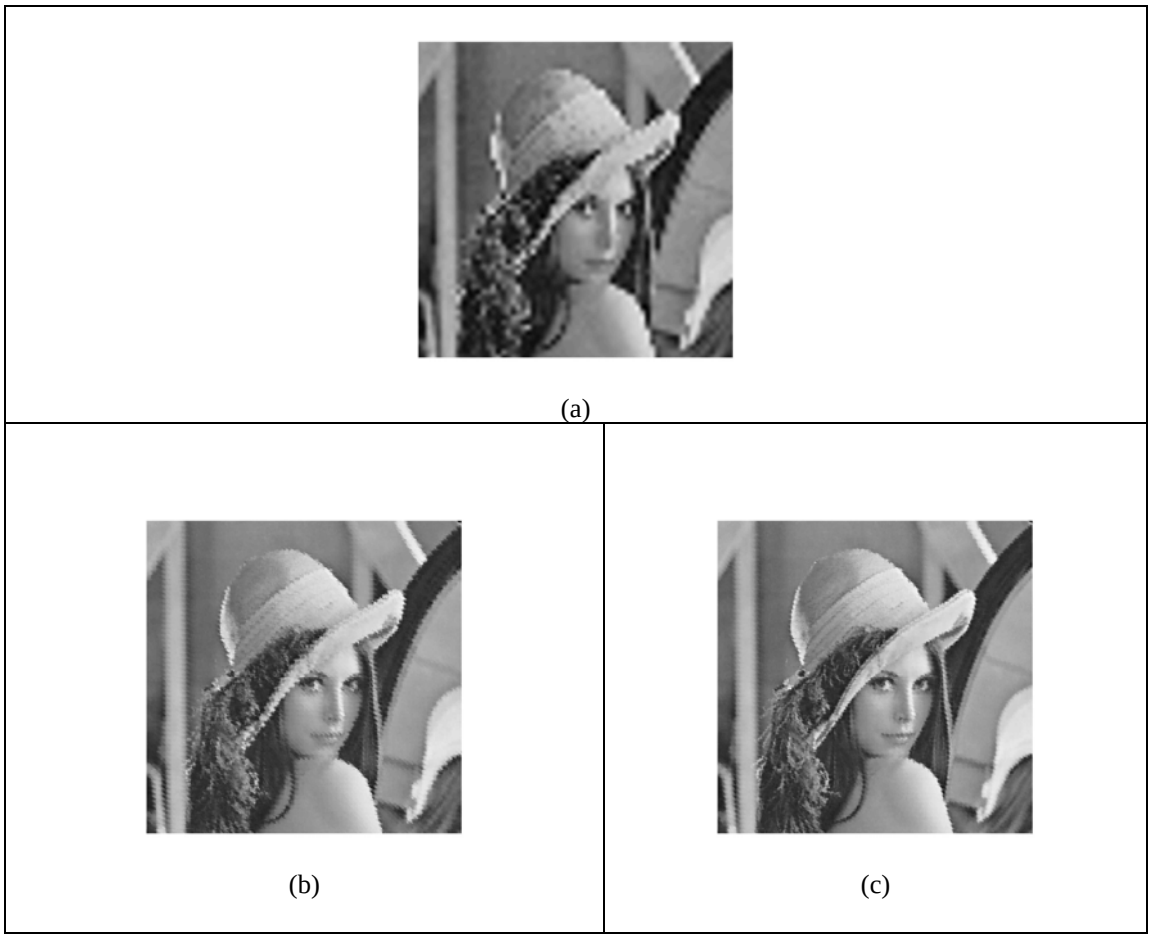
Similar to this experiment with the “Lenna” image, we used the low-resolution images obtained from down-sampling the “Plane” image, and we reconstructed the super-resolution image by using both LF and SLF algorithms. PSNR was 28.3430 dB with LF and 29.4804 dB when the SLF algorithm replaced the LF algorithm. Table 4.1 shows the

results of calculating PSNR from the two experiments with “Lenna” image and “Plane” image.

Table 4.1 PSNR of the reconstructed image using LF to estimate the sub-pixel shifts and using SLF to estimate sub-pixel shifts.

	PSNR(dB) LF	PSNR(dB) SLF
Lenna Image	23.6642	25.6309
Plane Image	28.3430	29.4804

For visual comparison, the reconstructed super-resolution images in the above experiments are shown in Fig. 4.9 and Fig. 4.10 for the “Lenna” and “Plane” images, respectively. For comparison, an upsampled image obtained by using a bicubic interpolation of the selected reference low-resolution images in both experiments are shown in Figs. 4.10(a) and 4.11(a).

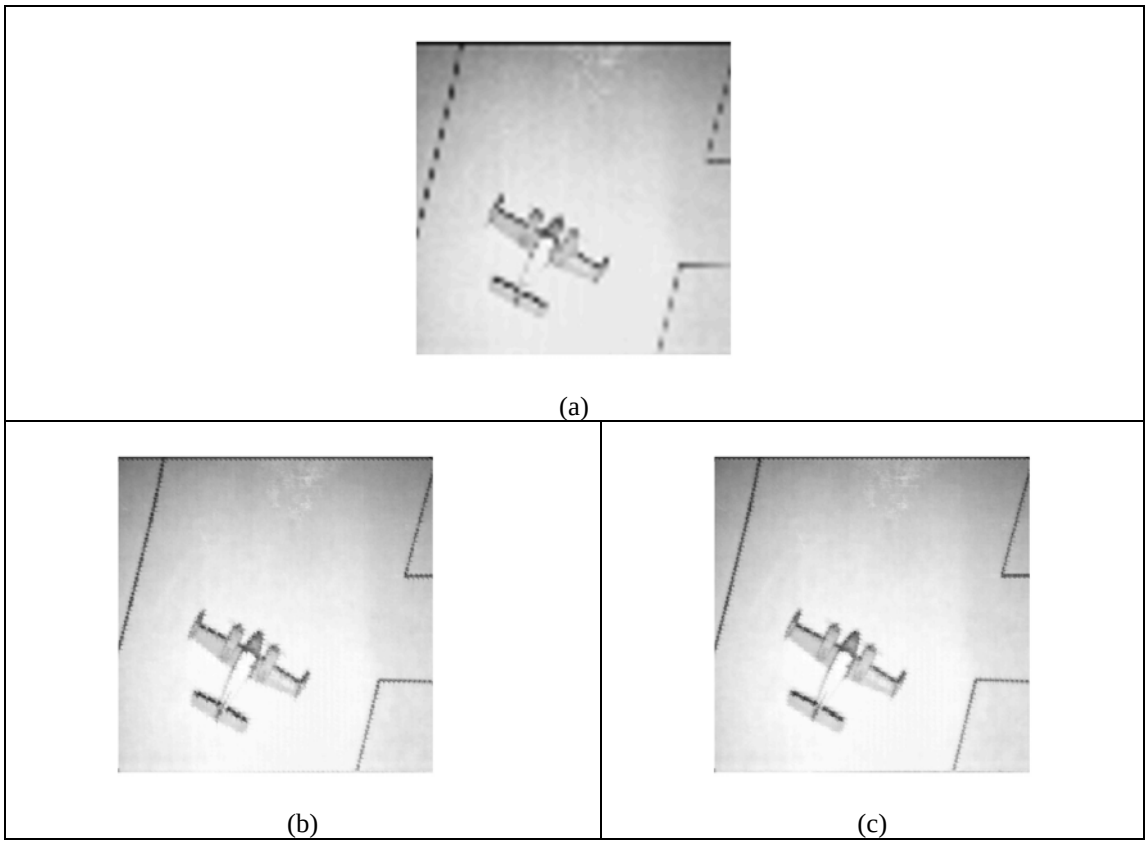


(a)

(b)

(c)

Fig. 4.9. Results of "Lenna" image. (a) A bicubic interpolation image, (b) the result of reconstruction based on using LF sub-pixel shifts estimation, (c) the result of reconstruction based on using SLF sub-pixel shifts estimation.



(a)

(b)

(c)

Fig. 4.10. Results of “Plane” image. (a) A bicubic interpolation image, (b) the result of reconstruction based on using LF sub-pixel shifts estimation, (c)the result of reconstruction based on using SLF sub-pixel shifts estimation

CHAPTER 5

Multi-Modal Image Registration

5.1 Introduction

While some efficient procedures have been developed lately, the registration of images acquired from sensors operating in different modalities is still a challenging problem. In general, such images have different gray level characteristics and the features in the two images to be registered are often not well preserved, rendering registration techniques such as those based on feature extraction and area correlation generally not efficient, and hence not feasible in all cases. In this dissertation, we propose a new algorithm for multi-sensor image registration based on the local frequency representation of the images together with image representation by Computer-Aided Design (CAD) models that permit use of region-of-interest-to-region-of-interest, ROI-to-ROI, space to solve the image registration problem (instead of relying only on the captured images to solve image registration problem using image-to-image space). The key point underlying the proposed approach is the employment of local frequency representations of images where CAD models are used to replace objects in the scene in order to efficiently determine sets of matching points from the images to be registered, which in turn enables obtaining correspondence between these sets for estimating the transformation parameters. Performance evaluation results reported here indicate that the proposed technique is robust and yields promising results for multi-modal image registration.

5.2 Background and Motivations

In many image-based systems, it is necessary to use many sensors operating in different modalities to improve the reliability of the overall system. The information captured by sensors of different types are inherently different and have different gray level characteristics, thus accentuating the registration problem [89]. Although a number of new registration algorithms have been developed in the recent past, they typically encounter difficulties in accurately registering images acquired from different sensing modalities on a consistent basis [90]. Among some recent attempts made at multi-sensor registration, one may cite the use of the point matching algorithm in [91] to register electro-optical (EO) and infrared (IR) images. The work reported in [90, 92] uses contours instead of features, but this method works only with image pairs in which the contour information is preserved. An alternate approach in [89] uses segments to form matching triangles, and the registration method is based on the assumption that the only segment-related invariant and robust characteristic between the two images to be registered is the straight line defined by the segment.

Modeling the content of images from using models of objects is a topic that is receiving a lot of attention at present by researchers in image analysis [142-147]. Especially useful in this task are Computer-Aided Design (CAD) models of man-made objects such as buildings, tanks, and trucks. For image registration, such models can provide a mechanism for image-to-ROI space transformations and ROI-to-ROI space transformations rather than relying on image-to-image space transformations only [148]. In this dissertation, we shall develop a new multi-sensor image registration algorithm that

combines the capabilities resulting from employing CAD models with the local frequency extracted from the images to be registered. The use of local frequency for multi-modal image representation is motivated by the fact that it is relatively robust to gray level changes and can reliably detect the edges and ridges simultaneously in the various images. Results from several performance evaluation experiments are included to illustrate the capabilities of the algorithm in registering imagery data acquired from multiple diverse sensors.

5.3 Motivations for Using Region-of-Interest Space Transformations

Before exploring the motivations behind using what is called ROI-to-ROI space, first we need to introduce the following definitions.

-*Image space of a scene*. It is the 2D image for that scene (e.g. Electro-optical image and IR image). Fig. 5.1(a) shows the image space for a scene.

-*ROI space of a scene*. It is the 2D image that is generated from the image space for that scene by a high-level segmentation of selected object(s) in the scene and by eliminating all of the background. ROI space contains only selected object(s) from the image space. Fig. 5.1(b) shows an example for *ROI space* generated from *the image space* in Fig. 5.1(a).

-*Image representation of CAD model*. It is the 2D image of the CAD model for a specific object. Fig. 5.2 shows the image representation of the CAD model for a car.

In the approach that will be proposed for segmentation, the image of CAD model will be used to replace the ROI after the CAD model is successfully aligned to the object's ROI

and its pose is recovered. The reasons and the details will be given below. Fig. 5.3 shows a scene captured by an IR sensor and the corresponding ROI space using image of a CAD model for a car in the scene after the alignment and recovering the pose.

1(a) *1(b)*

Fig.5.1. (a)Image space. 1(b) The object space corresponding to image space in 5.1(a).

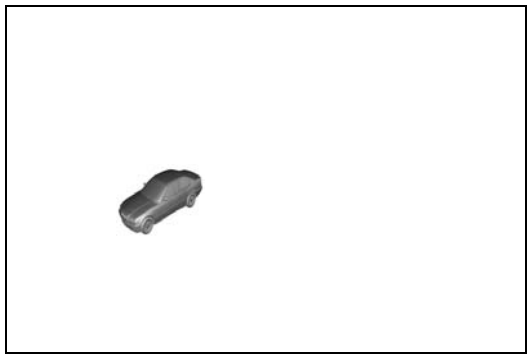
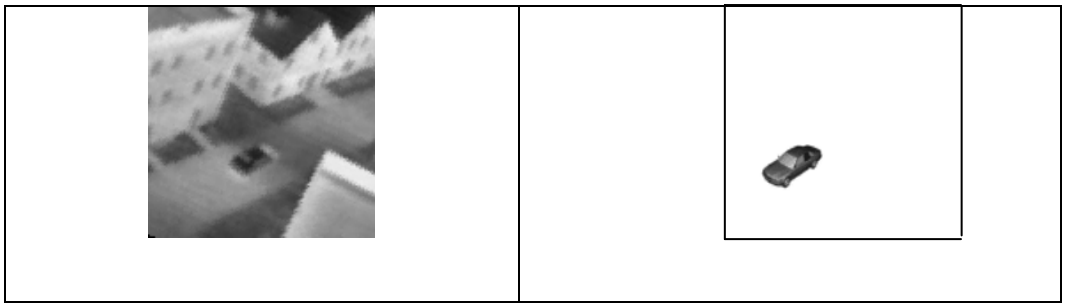


Fig. 5.2 The image representation generated from the CAD model with a specific pose.



(a) (b)

Fig. 5.3. (a) IR image of a real scene, image space. (b) The corresponding ROI space contains the car in the IR image space.

It should be noted that, extensive research performed during the past decade has given rise to many image registration techniques. However, no algorithm provides a satisfactory level of performance for accurately registering images acquired from diverse sensors consistently. This is due to the fact that the gray levels or texture characteristics of data captured by different types of sensors do not necessarily match well. Thus, the features extracted from these images are often not well preserved, rendering registration techniques such as those based on feature extraction and area correlation generally not efficient, and hence not feasible to use in all scenes. However, many of the scenes that are imaged include several man-made objects, such as buildings, trucks, cars, and tanks. Fig. 5.4 shows some illustrative real images that do not include clear features for registration, but contain certain man-made objects (such as a truck or a tank, etc.).

Fortunately more recently, successful image analysis algorithms to recover man-made objects from the images based on using CAD models for these objects are being developed for applications in object recognition and computer vision. These techniques enjoy a certain degree of robustness in recovering objects even when the features of the objects present in the different images are not identical, as in the case of images acquired from different kinds of sensors such as IR, LADAR, MMW, and CCD [142-147,151]. Use of this idea for image registration will be the primary focus of this paper.

Employing the image representation for the CAD model that is aligned to the object in the scene and has its pose, provides a mechanism for employing the ROI image space to solve the correspondence problem, and to estimate the parameters defining image-to-

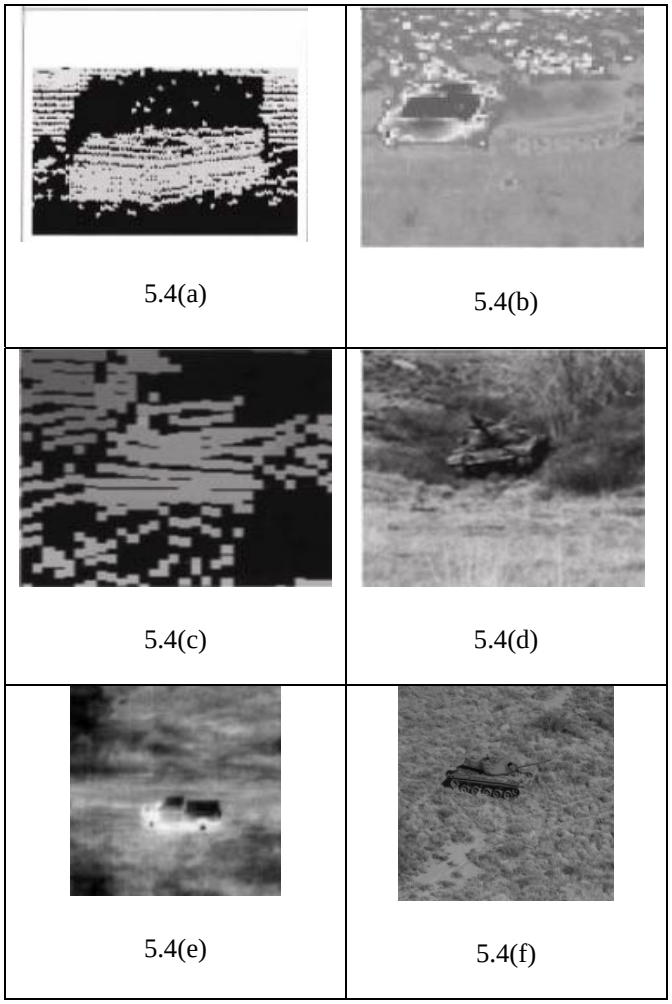
image registration. In fact, to solve the registration problem between two images, we can use the two images in three different ways as follows:

1) *Image Space to Image Space (image-to-image)*, by using the image space.

2) *ROI Space to ROI Space (ROI-to-ROI),* by generating the corresponding ROI images and then using them to solve the image registration problem.

3) *Image Space to ROI Space*, by using the image space of one of the two images against the ROI space of the other image to solve the image registration problem.

Fig. 5.5 illustrates the possible combinations between the image space and the ROI space to solve the image registration problem. The arrows indicate which pair of images can be used to solve the problem.



Figs. 5.4(a) and 5.4(b) LADAR and FLIR images for the same scene. 5.4(c) LADAR and 5.4(d) EO for the same scene, 5.4(e) FLIR, and 5.4(f) EO different scenes. Illustrative images acquired by different sensors and contain man-made objects.((a)-(d) from [144].

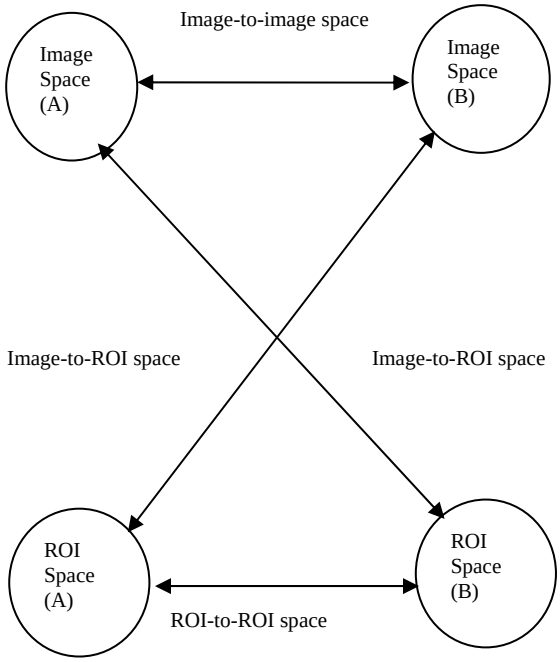


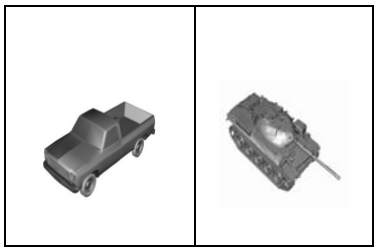
Fig5.5. Employing the image space and the ROI space to solve the image registration problem.

Given an image, using image of the CAD model to construct the corresponding ROI image space of that image has many advantages. It overcomes the lack of feature consistency in images acquired by different types of sensors. Of particular relevance to our work is the possibility it presents for estimating the registration parameters between the ROIs extracted from the images to be registered. We may note that since the registration process can now attempt to estimate the correspondence between the ROIs (and not the original images themselves), the procedure can be applied with the same efficiency irrespective of the imaging modality used to acquire the image. Performance evaluation results given later in this paper provide confirmation of this basic idea. Fig. 5.6(a) shows the image representation of a CAD model for a truck that can be used to model the object present in the scene depicted in Fig. 5.4(a) to construct the ROI space. Similarly, Fig. 5.6(b) shows the image of a CAD model of a tank that can be used to replace the object in Fig. 5.4(b) to construct the ROI space.

5.3.1 Region-of-Interest Space Using Image Representation of CAD Model

To obtain the image representation of CAD model, first we need to solve the problem of simultaneous pose and alignment determination of CAD model to the corresponding object in the scene. This is an old problem in the literature, specifically for the computer vision community and model-based automatic target detection (ATR) community [143, 149, 150, 151]. It can be defined as follows: given 3D set of points, CAD model, and its 2D image and we need simultaneously to determine the alignment and the pose of the 3D points to its 2D image.

5.6(a) *5.6(b)*



Figs. 5.6(a) and 5.6(b): CAD model images for a truck and a tank.

Solving this problem depends on the application [144,147,151]. In [151], the authors used the general pattern theory to detect, to track, and to classify the targets in a unified process. They used synthesized templates from CAD models of targets to match the measured data and through certain geometric transformations they determine the orientations and the positions of the target. Their work is guided by *Jump-diffusion* processes, which provide the dynamic flexibility to accommodate higher and lower complexity scenes. *Jump-diffusion processes* provide a mathematical foundation for a sequential detection and computationally efficient target hypothesis during recognition.

5.4 Local Frequency Estimation of Region-of-Interest Image

While using image representation of CAD models of the objects present in the images provides a mechanism for obtaining ROI-to-ROI space transformation, estimation of the required transformation parameters will require determining a set of points extracted from each image to be used for matching purposes. Use of the local frequency representation can significantly aid in this process. To estimate the local frequency of an CAD model image, Algorithm 3.1 will be employed for this purpose such that the input image for the algorithm it will be the CAD image.

For illustration purpose, Fig. 5.7 shows the gray levels coding for the local frequency representations for the images shown in Fig. 5.6. The brighter pixels reflect higher local frequency values and vice versa.

5.5.Matching the Local Frequency Representation

Having obtained the local frequency representations of a pair of ROI images, we selected a set of control points. We select the size of the control points set such that the selected number of points are enough to capture features in the pair of images and to establish the matching between them. To determine the control points, we select the points with higher local frequency values. For this purpose, we have conducted many experiments to study the relation between the number of points and the extracted features from an image. Figs. 5.9-5.11 Show the results from different objects shown in Fig. 5.8.

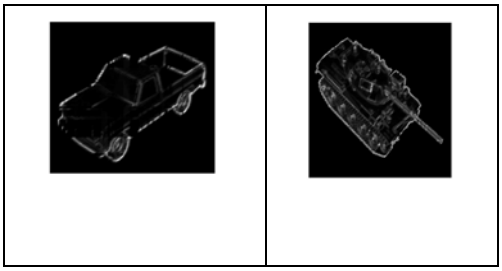


Fig. 5.7. Gray levels coding of Local frequency representation for the images in Fig. 5.6.

(a)

(b)

(c)

Fig. 5.8 CAD model images. (a) Car, (b) tank, (c) truck.

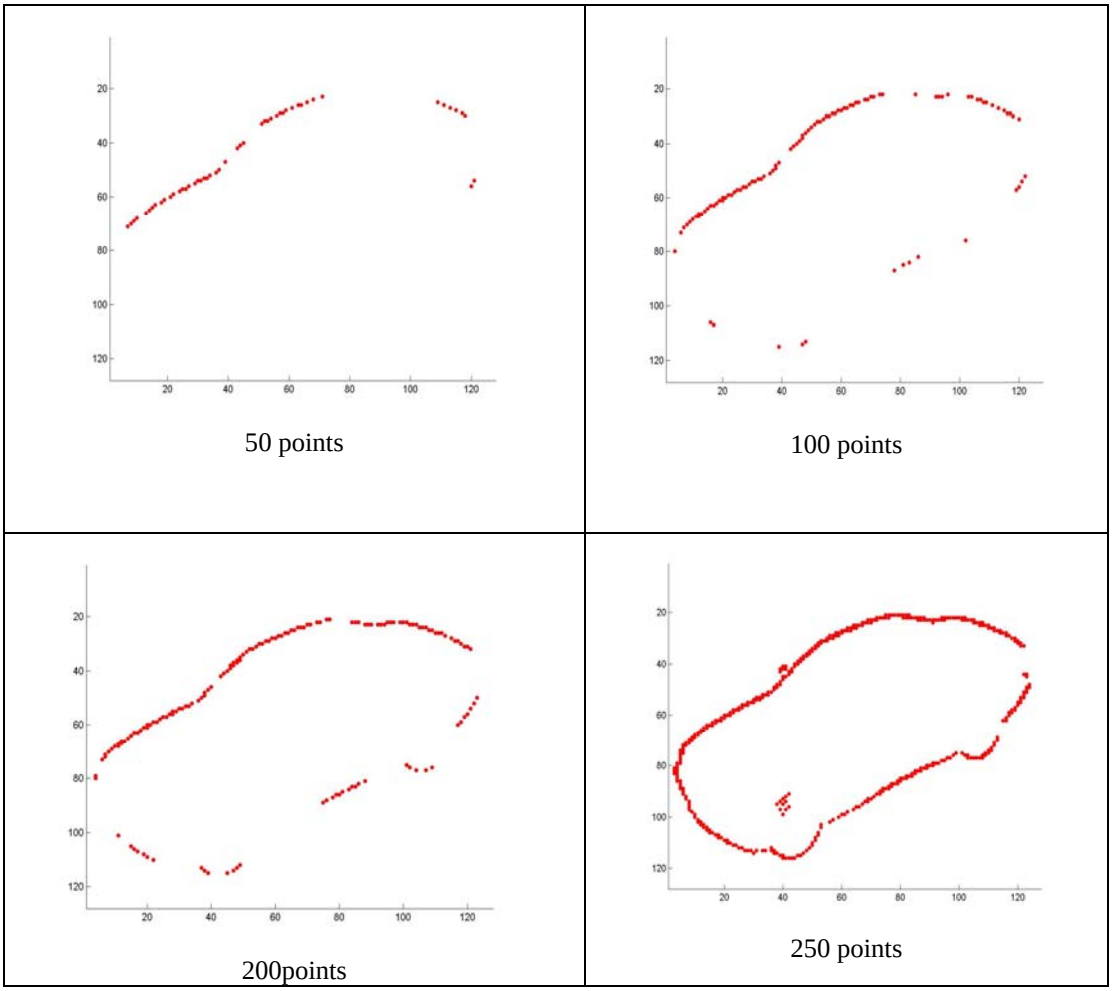


Fig. 5.9 shows different number of extracted control points from local frequency estimation of CAD model image in Fig. 5.8(a).

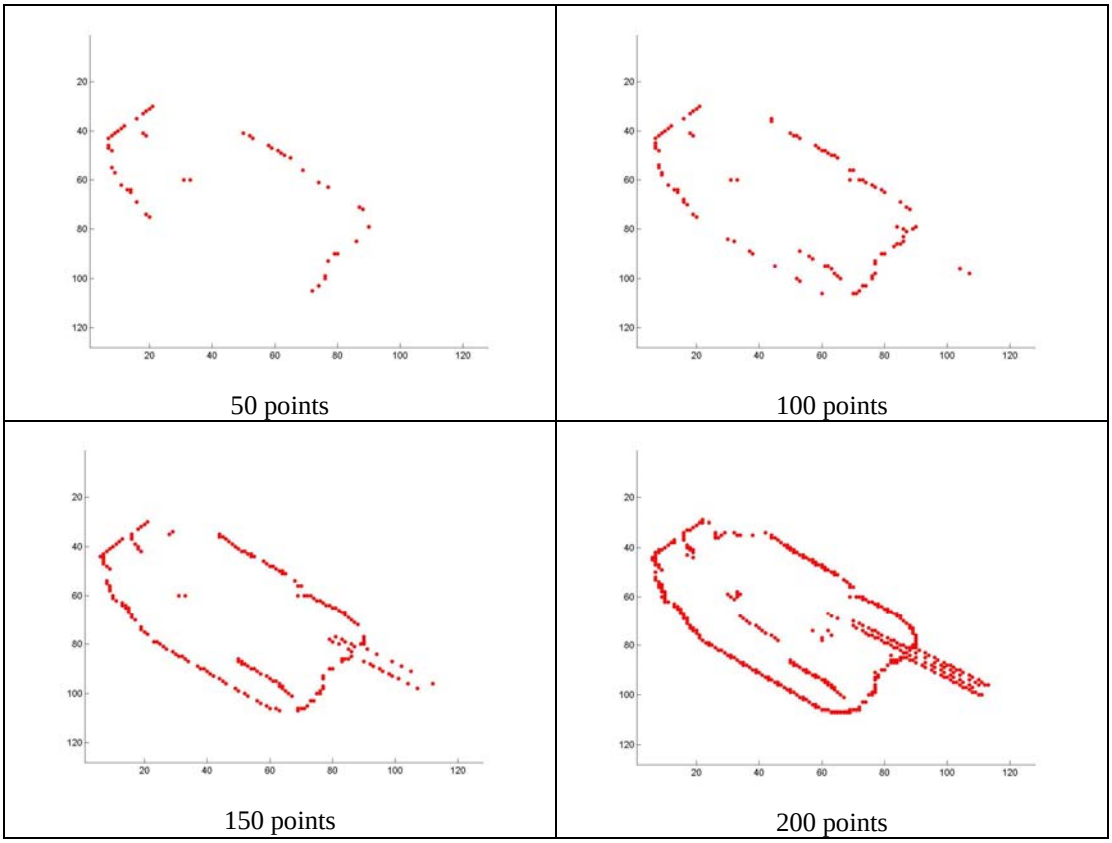


Fig. 5.10 shows different number of extracted control points from local frequency estimation of CAD model image in Fig. 5.8(b).

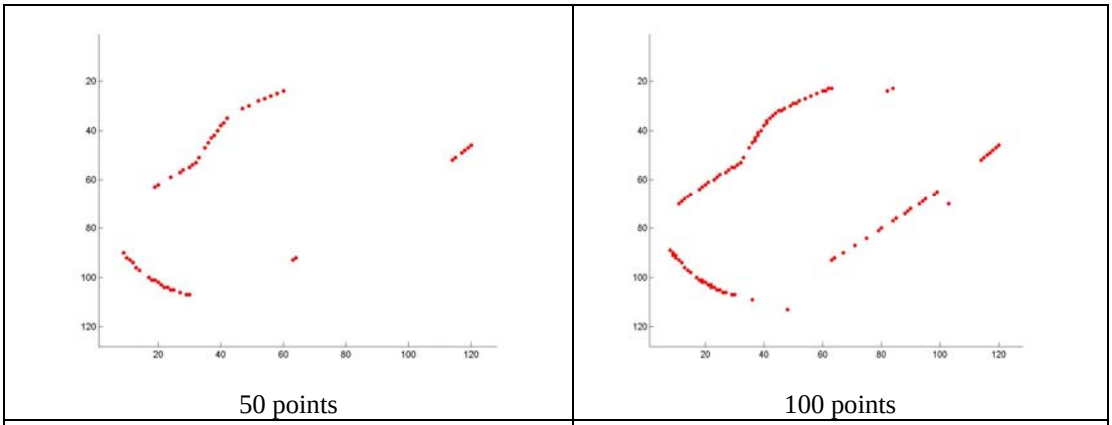


Fig. 5.11 shows different number of extracted control points from local frequency estimation of CAD model image in Fig. 5.8(c).

Once the control points are selected for a pair of object CAD images, the correspondence problem between them has to be solved. For this purpose, we have employed Algorithm 3.2 in Chapter 3. Fig. 5.12 shows a schematic diagram of the proposed registration algorithm .

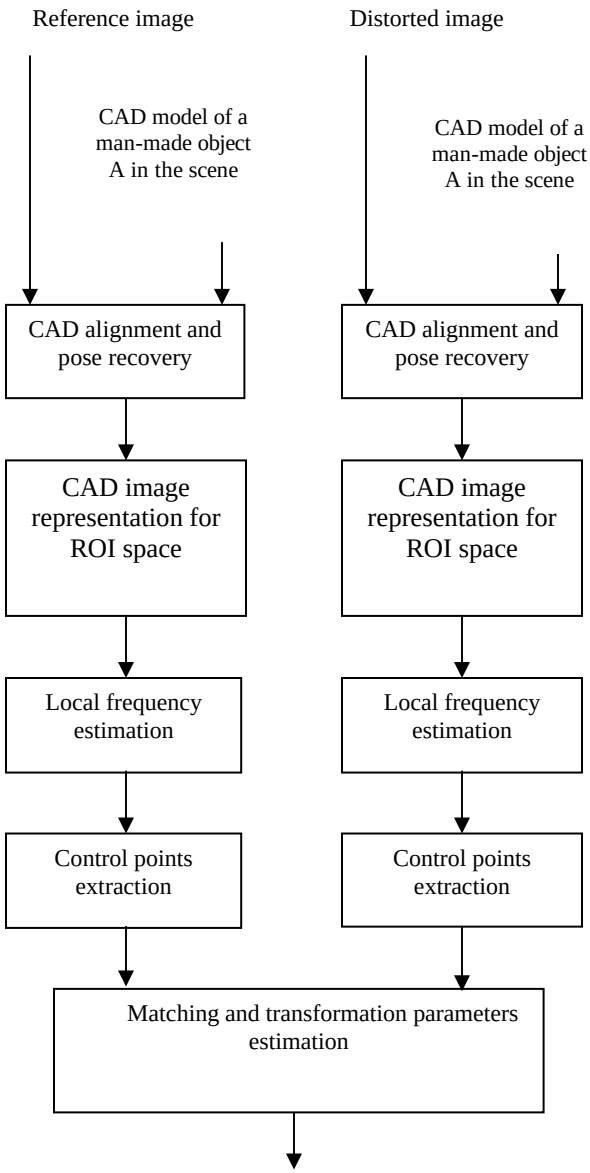


Fig. 5.12 Schematic overview of the proposed algorithm.

5.6 Experimental Results

In this section, from many conducted experiments to evaluate the proposed technique, we provide two illustrative results for registration real images. The images and the CAD models used in the experiments are from public databases. In the first experiment, Figs. 5.13(a) and 5.13(b) shows two images of size 256×256 for real scene. The two images captured by electro-optical (EO) and infrared (IR) sensors, respectively. The registration parameters between the two images are unknown. Estimating registration parameters using the image space directly might not possible since the features of the scene in both images are not consistent due to the difference in the working principles of the sensors. However, the scene contains man-made objects such as the truck indicated by the arrow in the images. The Truck is man-made object and its corresponding CAD model is available, which can be aligned to this truck and recovers its pose. Having aligned CAD model to the truck in the images, we can exploit this aspect by using image representation of these aligned CAD models to establish the ROI-to-ROI space transformation in order to solve the registration problem as shown in Figs. 5.13(c) and 5.13(d). Figs. 5.13(c) and 5.13(d) show the corresponding ROI space images for image space in Figs. 5.13(a) and 5.13(b), respectively. We have to say here that for image registration there is a general assumption that the pair of images used to recover the registration parameters has been captured with the same viewing angle to the scene. This means that the objects in the pair of images have the same viewing angle with respect to the capturing sensors and hence no chance to see the same object in the two images from two different viewing

angles. So, the object space images using the image representation for CAD model have the same object CAD model image but there exist mis-registration between them.

Having obtained the object space images, we estimate the local frequency representations for these object images using the procedure outlined in Section 5.4. In order to perform the matching, we select a set of points that are the higher values for the local frequency estimation. In our experiments, we selected 200 points for each object to cover most of the features for that object and also to establish the matching. Then the matching algorithm in Section 3.4.2 (Algorithm 3.2) has been applied to solve correspondence problem to estimate the transformation parameters at the same time. Using the proposed technique, the recovered registration scaling was 0.62 and the translations are 0.01pixel and 69.37 pixel in x and y directions, respectively. To evaluate the results of proposed technique, we registered the two images by the estimated parameters and then averaging the summation is used to fuse the two images. The result of the fusion is shown in Fig. 5.13(e). Apparently, the two images are aligned quite well which reflects the accuracy of the proposed technique.

In another experiment, Figs. 5.14(a) and 5.14(b) show a pair of images for a real field, the images are captured by EO and IR sensors, respectively. Also, here the registration parameters are unknown. Due to the features in IR image are not clear as in EO image, recovering the registration parameters using these images directly is not easy and might not be possible. So, we applied our algorithm to estimate the registration parameters after constructing the object space images by using a suitable CAD model for the tank, which is indicated by the arrow, in the scene with alignment and pose recovering process. The

estimated registration parameters are then used to register the two images for fusion purpose. Using the average fusion approach, the result in fused image is shown in Fig. 5.14(c). Again, the result of the fusion using the estimated parameters emphasized the accuracy of the proposed technique.

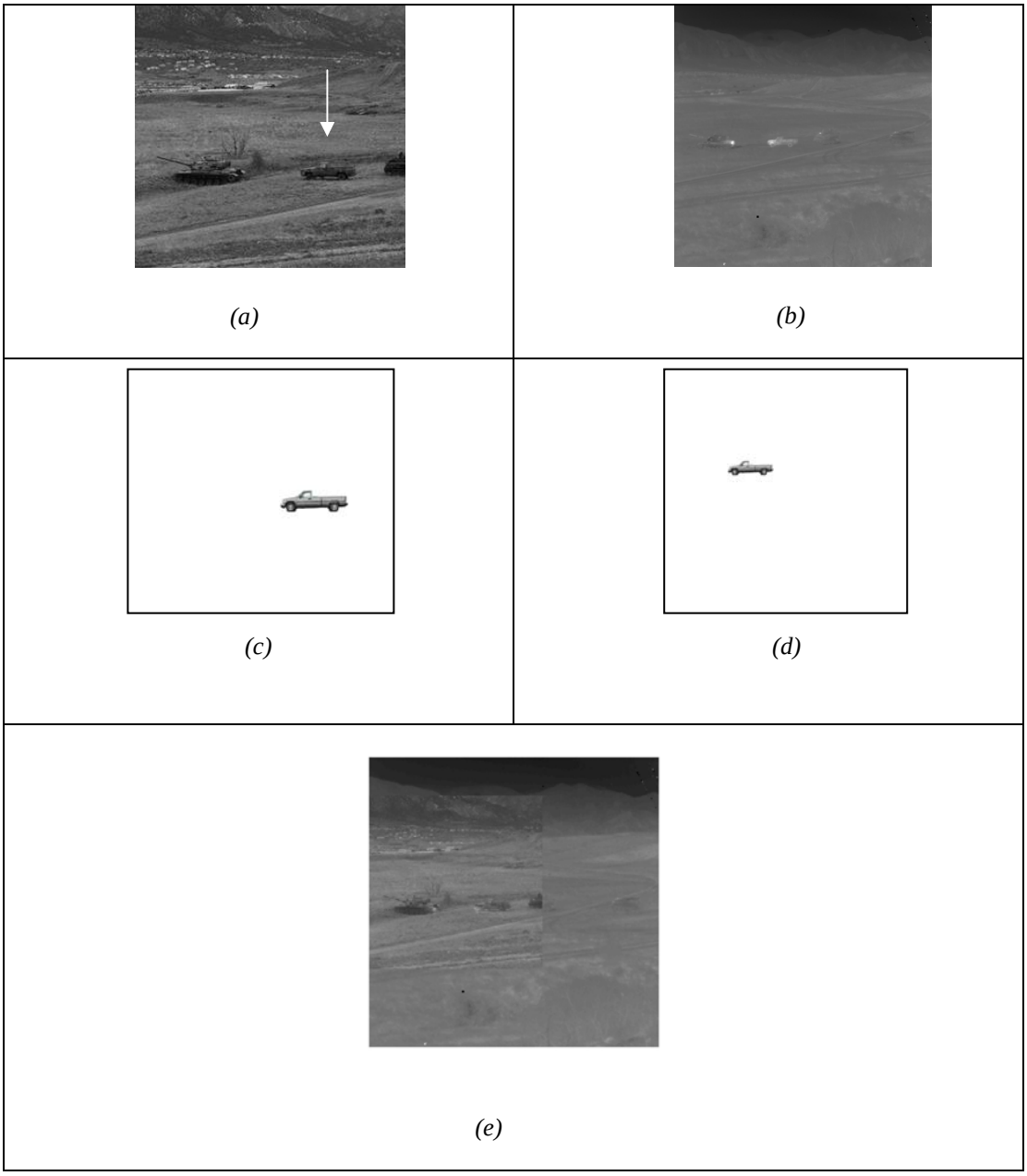
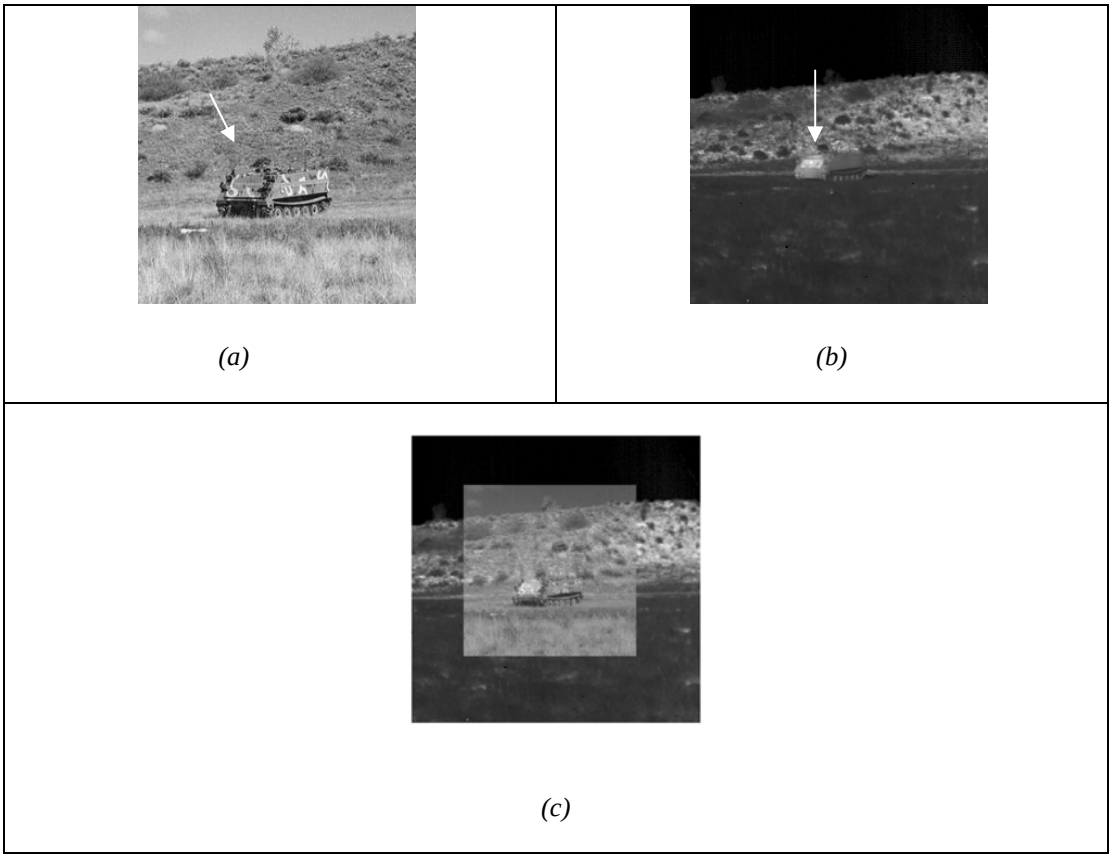


Fig. 5.13. (a) and (b) are EO and IR images respectively, (c) and (d) are the corresponding object space images for a truck in the scene, (e) the fused image using IR and EO images after registration by the proposed technique.



(a) *(b)* *(c)*

Fig. 5.14. (a) and (b) are EO and IR images respectively, (c) the result image of fusing EO and IR images after registration by the proposed technique.

CHAPTER 6

Conclusions

This chapter briefly summarizes the results presented in the dissertation and outlines some directions for further research.

6.1 Summary of Contributions

The first contribution of this dissertation is a novel algorithm for the estimation of the local frequency representation of a given image. The algorithm attempts to retain a high level of representation accuracy while minimizing the latency of the Gabor filter bank, which in turn ensures computational efficiency making it ideally suited for the registration and fusion of images. Results from various performance evaluation experiments reported in this dissertation confirm the versatility of the present algorithm to handle images acquired from different modality sensors and its robustness to variations in the scenes.

Employing local frequency representation offers a promising approach for solving image registration problems. For this reason, the dissertation introduced a new algorithm to solve the image registration problem. The developed algorithm employs the local frequency representation of the image to extract control points to solve the registration by matching the control points.

Results presented in this dissertation demonstrate that the developed technique can be efficiently utilized to register diverse images with differing complexity levels and that the algorithm is quite robust to scene details. The present algorithm enables one to obtain a fully automatic approach for image registration and hence can find many applications in image registration, sensor fusion, object recognition, and detection and tracking of surveillance targets.

The third contribution of this dissertation is introducing an extension of using local frequency representation of the image to solve the problem of sub-pixel image registration. The proposed approach is based on using the normalized cross-power spectrum of the local frequency representation of the images. The local frequency representation is presented in vector form by including the orientation of the local frequency in addition to the magnitude in order to enhance the correlation. Further, to improve the results of the approach, we employed scaled local frequency to replace local frequency. The experimental results show that using scaled local frequency is superior to using local frequency.

As the fourth and the last contribution in the dissertation, we introduce a novel algorithm to solve the problem of registration between multi-modal images. The proposed algorithm is based on using a CAD model to replace a man-made object in the scene of the image and then using the local frequency representation of the CAD model to solve the registration problem by matching the regions-of-interest corresponding to the

object extracted from the images. Employing the image representation of CAD models for facilitating use of region-of-interest-to-region-of-interest space transformation is a promising approach for solving image registration problems, particularly when the images to be registered come from different sensing modalities and hence have significant gray scale differences. Using the local frequency representation of the object space data in generating matching points proved to be quite successful. In addition, the conversion of the given images to be registered to region-of-interest space prior to extraction of matching points makes the process robust to scene details. The algorithm hence can find many applications in multi-sensor image registration, multi-sensor fusion, and multi-sensor target tracking.

6.2 Some Directions for Further Work

Local frequency representation was employed to solve pixel level and sub-pixel level registration problems. However, because of the numerous advantages enjoyed by the local frequency representation of an image, it can be employed to replace the direct image in many other tasks. In particular, local frequency of an image can be employed for tasks such as target detection, target tracking, detection and extraction of features for image fusion.

For pixel-level registration, the performance of the proposed algorithm depends on the accuracy of the matching algorithm selected to match the control points extracted. To reduce the dependency on the matching algorithm, we can introduce employing the

corresponding values of the local frequency at the control points in addition to the coordinates of the control points. This step may reduce the dependency on the matching algorithm because there is another source of matching.

Also, despite the good performance demonstrated by the proposed sub-pixel registration algorithm in Chapter 4, the sub-pixel registration problem is still a challenging one. In general, it is known that the performance of the various algorithms available for this problem offer non-uniform performance when applied to different types of scenes (high spatial activity scenes vs. scenes where the level of activity is not significant) and different amounts of sub-pixel shifts. One may hence conjecture selection of an appropriate algorithm based on classifying the images according to the spatial activity present and their frequency content. To facilitate such a selection process, one needs to perform an exhaustive study of different scene patterns and conducting registration experiments with different amounts of sub-pixel shifts introduced into these images.

References

[1] L. A. Klein, *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*. SPIE-The International Society for Optical Engineering, Washington, USA, 2004.

[2] D. L. Hall and James Llians, *Handbook of Multisensor Data Fusion*. CRC Press, 2001.

[3] X. E. Gros, *NDT Data Fusion*. Co-published in North America by John Wiley & Sons, Inc., New York, 1997.

[4] L. A. Klein, *Sensor and Data Fusion Concepts and Applications*. SPIE-The International Society for Optical Engineering, 2nd edition, Washington, USA, 2004.

[5] D. L. Hall and J. Llinas, "Introduction to Multi-sensor Data Fusion," *Proceedings of the IEEE* vol. 85, no. 1, pp.6-23, Jan. 1997.

[6] L. G. Brown, "A Survey of Image Registration Techniques," *ACM Computing surveys*, vol. 24, no. 4, pp. 325-376,1992.

[7] E. J. M. Rignot, R. Kowk, J. C. Curlander, and S. S. Pang, "Automated Multisensor Registration: Requirements and Techniques," *Photogrammetric Engineering and Remote Sensing*, vol 57, no. 8, pp. 1029-1038, Aug. 1991.

[8] A. Maintz, J. B. and Viergever, M. A., "A Survey of Medical Image Registration," *Medical Image Analysis*. Vol. 2, no. 1, pp. 1-36, 1998.

[9] B. Zitova and J. Flusser, "Image Registration Methods: A Survey," *Image and Vision Computing* vol. 21, no. 11, pp. 977-1000, 2003.

[10] B. Jahn, *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*. Springer-Verlag berlin Herlin Heidelberg, Germany, 1997.

[11] W. K. Pratt, *Digital Image Processing*. Second edition, John Wiley & Sons, inc. Press, 1991.

[12] L. G. Shapiro and A. Rosenfeld, *Computer Vision and Image Processing*. Boston: Academic Press, 1992.

[13] R. Lewis, *Practical Digital Image Processing*. Ellis Horwood Limited, West Sussex, 1990.

[14] W. Niblack, *Digital Image Processing*. Prentice-Hall International , (UK) Ltd, 1985.

[15] J. P. Lewis, "Fast Template Matching," *Vision Interface*, pp. 120-123,1995.

[16] P. J. Burt, C. Yen and X. Xu, "Local Correlation Measures for Motion Analysis: A Comparative Study," *Proc. of IEEE Conf. on Pattern Recognition and Image Processing*, , Las Vegas, pp. 269-274, 1982.

[17] B. K. Ghaffary and A. A. Sawchuk, "A Survey of New Technique for Image Registration and Mapping," *Proceedings of the SPIE: Applications of Digital Image Processing*, vol. 432, pp. 222-239, 1983.

[18] H. Hanaizumi and S. Fujiimura, "An Automated Method for Registration of Satellite Remote Sensing Images," *Proceedings of The International Geoscience and Remote Sensing Symposium, IGARSS'93*, Tokyo, Japan, pp. 1348-1350, 1993.

[19] R. Berthilsson, "Affine Correlation," *Proceedings of The International Conference on Pattern Recognition, ICPR'98*, Brisbane, Australia, pp. 1458-1461, 1998.

[20] A. Simper, "Correcting General Band to Band Mis-Registrations," *Proceeding of the IEEE International Conference on Image Processing, ICIP'96*, Lausanne, Switzerland, pp. 597-600, 1996.

[21] L. Ding and A Goshtasby, M. Satter, "Volume Image Registration by Template Matching," *Journal of Image and Vision Computing*, vol.19, pp. 821-832, 2001.

[22] W. K. Pratt, "Correlation Techniques of Image Registration," *IEEE Trans. On Aerospace and Electronics Systems*, vol. 10, pp. 353-358, 1974.

[23] A. Rosenfeld, G J. Vanderbrug "Coarse-Fine Template Matching," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 7, pp. 104-107, 1977.

[24] E. D. Castro and C. Morandi, "Registration of Translated and Rotated Images Using Finite Fourier Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.9, pp. 700-703, 1987.

[25] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric Phase-only Matched Filtering of Fourier-Melline Transform for Image Registration and Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 1156-1168, 1994.

[26] B. S. Redy and B. N. Chatterji, "An FFT-based Technique for Translation, Rotation and Scale-Invariant Image Registration," *IEEE Trans. on Image Processing*, vol. 5, no. 8, pp. 1266-1271, 1996.

[27] G. Wolberg and S. Zoka, "Robust Image Registration Using Log-polar Transform," *Proceedings of the IEEE International Conference on Image Processing*, Canada, pp. 493-496, September 2000.

[28] L. Lucchese, G. Doretto, and G.M . Cortelazzo, "A Frequency Domain Technique for Range Data Registration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1468-1484, 2002.

[29] A. Averbuch, and Y. Keller, "FFT Based Image Registration," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 13-17, May 2002.

[30] C. D. Kuglin and D. C. Hines, "The Phase Correlation Image Alignment Method," *In Proc. of International Conf. on Cybernetics and Society*, pp163-165, 1975.

[31] P. Viola and W. Wells, "Alignment by Maximization of Mutual Information," *Proc. Fifth Int. Conf. On Computer Vision*, pp.16-23, Boston, 1995.

[32] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Trans. on Medical Imaging*, vol. 16, no. 2, pp.187-198, 1997.

[33] K. Johnson, A Rhodes, J. Le Moigne, and I. Zavorin, "Multi-Resolution Image Registration of Remotely Sensed Imagery Using Mutual Information," *SPIE AeroSense "Wavelet Applications VIII"*, vol. 4391, pp. 330-339, April 2001.

[34] P. Thevenaz and M. Unser, "An Efficient Mutual Information Optimizer for Multiresolution Image Registration," *Proc. of IEEE Inter. Conf. on Image processing, ICIP'98*, Chicago, IL., pp. 833-837, 1998.

[35] P. Thevenaz and M. Unser, "A Pyramid Approach to Sub-Pixel Image Fusion Based on Mutual Information," *Proc. of IEEE International Conference on Image Processing, ICIP'96,* Lausanne, Switzerland, pp. 256-268, 1996.

[36] P. Thevenaz and M. Unser, "Spline Pyramids for Inter-Modal Image Registration Using Mutual Information," *Proceeding of SPIE*, vol. 3169, pp. 236-247, 1997.

[37] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, " Multimodality Image Registration by Maximization of Mutual Information," *IEEE Trans. on Medical Imaging*, vol. 16, pp. 187-198, 1997.

[38] J.P. W. Pluim, J. B. Maintz, and M. A. Viergever, "Mutual Information Matching in Multiresolution Contexts," *Journal of Image and Vision Computing*, vol. 19, pp. 45-52, 2001.

[39] P. Fua and Y. Leclerc, "Image Registration without Explicit Point Correspondences," *Proc. DARPA Image Understanding Workshop*, pp. 981-992, 1994.

[40] A. Cole-Rhodes, and A. Adenle, "Automatic Image Registration by Stochastic Optimization of Mutual Information," *IEEE Inter. Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp.313-316, April 2003.

[41] G. Wolberg and S. Zokai, "Image Registration for Perspective Deformation Recovery," *SPIE Conference on Automatic Target Recognition X*, Orlando, Florida, USA, vol. 4050, pp. 259-270, April 2000.

[42] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The Correlation Ratio as A New Similarity Measure for Multimodal Image Registration," *Proc. of first Inter. Conf. on Medical Image Computing and Computer-assisted Intervention, MICCAI'98*, lecture Notes in Computer Science, Cambridge, USA, vol. 1496, pp. 1115-1124, 1998.

[43] P. Van Wie and M. Stein, "A Landsat Digital Image Rectification System," *IEEE Trans. on Geoscience Electronics*, vol. 15, pp. 130-137, 1977.

[44] P. Thevenaz, M. Unser, "An Efficient Mutual Information Optimizer for Multiresolution Image Registration," *Proc. of the IEEE Inter. Conf. on Image Processing*, *ICIP'98*, Chcago, IL, pp. 833-837,1998.

[45] M. Irani. and P. Anandan, "Robust Multi-sensor Image Alignment," *Proc. of Inter. Conf. on Computer Vision, ICCV'98* , Bombay, India, pp 959-965,1998.

[46] J. Ton and A. K. Jain, "Registration Landsat Images by Point Matching," *IEEE Trans. Geoscience and Remote. Sensing* , vol. 27, no. 5, pp.642-651, 1989.

[47] A Goshtasby and J. Le Moigne, "Special Issue on Image Registration," *Pattern Recognition*, vol. 32, no.1, 1999.

[48] J. Le Moigne, W. Xia, P. Chalermwat, T. ElGhazawi, Manohar Mareboyana, Nathan Netanyahu, James C. Tilton, William J. Campbell, and Robert F. Cromp, "First Evaluation of Automatic Image Registration Methods," *IEEE International Geoscience and Remote Sensing Symposium Proceedings*, vol. 1, pp.315-317, 1998.

[49] T. Kim and Y. J. Im, "Automatic Satellite Image Registration by Combination of Matching and Random Sample Consensus," *IEEE Trans. On Geoscience and Remote Sensing*, vol. 41, no. 5, pp.1111-1117, 2003.

[50] F. H. Cheng, “Point Patterns Matching Algorithm Invariant to Geometric Transformation and Distortions,” *Pattern Recognition Letters*, vol.17, pp. 149-159,1984.

[51] J. W. Hsieh, H. Y. M. Liao, K. C. Fan, M. T. Ko, “Image Registration Using a New Edge-based Approach,” *Computer Vision and Image Understanding*, vol. 67, pp. 112-130, 1997.

[52] N. R. Pal and S. K. Pal, “A Review on Image Segmentation Techniques,” *Pattern Recognition*, vol. 26, pp. 1277-1294, 1993.

[53] S. H. Chang, F. H. Chang, W. H. Hsu, and G.Z. Wu, “Fast Algorithm for Point Pattern Matching: Invariant to Translations, Rotations and Scale Changes,” *Pattern Recognition*, vol. 30, pp. 311-320, 1997.

[54] S. Banerjee, D. P. Mukherjee, and D. D. Majumdar, “Point Landmarks for Registration of CT and MR Images,” *Pattern Recognition Letters*, vol. 16, pp. 1033-1042, 1995.

[55] X. Dai and S. Khorram, “Developedment of a Feature-Based Approach to Automated Image Registration for Multi-temporal and Multisensor Remotely Sensed Imagery,” *Inter. Geo. and Rem. Sen. Symposium Proc., IGARSS’97*, Singapore, pp.243-245, 1997.

[56] W. Beil, K. Rohr, and H. S. Stiehl, “Investigation of Approaches for the Localization of Anatomical Landmarks in 3-D Medical Images,” *Conf. Computer Assisted Radiology and Surgery*, Germany, pp. 265 – 270, 1997.

[57] N. M. Alpert, J. F. Bradshaw, D. Kennedy, and J. A. Correia, “The Principal Axes Transformation – A Method for Image Restoration,” *The Journal of Nuclear Medicine*, vol. 31, no. 10, pp 1717-1722, 1990.

[58] A. Giachetti, “Matching Techniques to Compute Image Motion,” *Image and Vision Computing*, vol. 18, pp. 247-260, 2000.

[59] L.-Y. Hsu, and M. H. Loew, “Fully Automatic 3D Feature-based Registration of Multi-modality Medical Images,” *Image and Vision Computing*, vol.19, pp 75-85, 2001.

[60] J. Williams and M. Bennamoun, “Simultaneous Registration of Multiple Corresponding Point Sets,” *Computer Vision and Image Understanding*, vol. 81, pp 117-142, 2001.

[61] V. Govindu and C. Shekhar, “Alignment Using Distributions of Geometric

Properties," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1031-1043,1999.

[62] A Goshtasby, G. C. Stockman, and C. V. Page, "A Region-Based Approach to Digital Image Registration with Sub-pixel Accuracy," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 24, pp. 390-399, 1986.

[63] Qinfen Zheng and Rama Cheeappa, " A Computational Vision Approach to Image Registration," *IEEE Trans. On Image Processing*, vol. 2, no. 3, July 1993.

[64] R. Safari, N. Narasimhamurthi, M. Shridhar, and M . Ahmadi, "Affine Point Pattern Matching and Its Application to Form Registration," *Proc. of the IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 656-659, 1996.

[65] J. Jianchao ,"Image Registration Based on Both Features and Intensity Matching," *Proc. of IEEE Inter. Conf. on Acoustics, Speech and Signal Processing, ICASSP'01*, pp. 1693-1696, 2001.

[66] S. Meshoul and M. Batouche, "A Fully Automatic Method for Feature –Based Image Registration," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pages 5, 2002.

[67] X. Dai and S. Khorram, "A Feature-based Image Registration Combined with Invariant Moments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2351-2362, Sept. 1999.

[68] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard, "A Level-set Based Approach to Image Registration," *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pp.86-93,11-12 June, 2000.

[69] Z. Zheng, H. Wang, and E. K. Teoh, "Analysis of Gray Level Corner Detection," *Pattern Recognition Letters*, vol. 20, pp. 149 – 162, 1999.

[70] M. Trajkovic and M. Hedley, "Fast Corner Detection," *Image and Vision Computing*, vol. 16, pp. 75 – 87, 1998.

[71] M. Fidrich and J.-P. Thirion, "Multiscale Extraction and Representation of Features from Medical Images," *INRIA Technical Report*, 1994.

[72] B. S. Manjunath, C. Shekhar, and R. Chellappa, "A New Approach to Image Feature Detection with Applications," *Pattern Recognition*, vol.29, no. 4, pp. 627 – 640, 1996.

[73] K. Rangarajan, M. Shah, and D. Van Brackle, "Optimal Corner Detection," *Computer Vision, Graphics, and Image Processing*, vol. 48, pp 230-245, 1989.

[74] K. P. Ngoi and J. C. Jia, "An Active Contour Model for Colour Region Extraction in Natural Scenes," *Image and VisionComputing*, vol. 17, pp. 955-966, 1999.

[75] J. B. A. Maintz, P.A. Van Den Elsen, and M. A. Viergever, "Evaluation of Ridge Seeking Operators for Multimodality Medical Image Matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 353-365, 1996.

[76] B Zitova, J. Kautsky, G Peters, and J. Flusser, " Robust Detection of Significant Points in Multiframe Images," *Pattern Recognition Letters*, vol. 20, pp. 199-206, 1999.

[77] D. I. Barnea, and H. F. Silverman, "A Class of Algorithms for Fast Digital Image Registration," *IEEE Transactions on Computers*, vol. 21, no. 2, pp 179-186, 1972.

[78] Jan F. Andrus, C. Warren Campbell, and Robert R. Jayroe, "Digital Image Registration Using Boundary Maps," *IEEE Transactions on Computers*, vol. 24, pp. 935-940, September 1975.

[79] A. Rosenfeld, and G. J. Vanderbrug, "Coarse-Fine Template Matching," *IEEE Transactions on Systems, Man and Cybernetics*, pp 104-107, 1977.

[80] P. E. Anuta, "Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques," *IEEE Transactions on Geoscience Electronics*, vol. 8, no. 4, pp 353-368, 1970.

[81] J. Le Moigne, and I. Zavorin, "Use of Wavelets for Image Registration," *SPIE AeroSense 2000, "Wavelet Applications VI*", Orlando, FL, April 24-28, 2000.

[82] H. Stone, J. Le Moigne, and M. McGuire, "Image Registration Using Wavelet Techniques," *Proc. of SPIE*, vol. 3240, pp. 116-125, 1998.

[83] H. S. Stone, J. L. Moigne, and M. McGuire, "The Translation Sensitivity of Wavelet-Based Registration," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1074-1081, October 1999.

[84] L. Fonseca, B.S. Manjunath, and C. Kindy, "Translation Invariant Wavelets to Image Registration," *In Proc. Of Image Registration Workshop, IRW'97*, NASA, Goddard Space Flight Center, MD, pp.13-28, Nov.20-21,1997.

[85] R. L. Allen, F. A. Kamangar, and E. M. Stokely, "Laplacian and Orthogonal Wavelet Pyramid Decompositions in Coarse-To-Fine Registration," *IEEE Trans.*

Signal Processing, vol. 41, no. 12, pp. 3536-3541, Dec. 1993.

[86] S. Ranade and A. Rosenfeld "Point Pattern Matching by Relaxation," *Pattern Recognition*, vol. 12, pp. 269-275, 1980.

[87] P. Chalermwat, and T. El-Ghazawi, "Multi-resolution Image Registration Using Genetics," *International Conference on Image Processing, ICIP'99*, vol.2, pp. 452-456, 1999.

[88] H. H. Li and Y. Tzhou, "A Wavelet-based Point Feature Extractor for Multi-sensor Image Registration," *in SPIE AeroSense "Wavelet Applications III"*, Orlando, Florida, pp. 524-534, 1996.

[89] E. Coiras, J. Santamaria, and C. Miravet, "Segment-Based Registration Technique for Visual-Infrared Images," *Optical Eng*ineering, vol. 39, no. 1, pp. 282-289, January 2000.

[90] M. A. Ali, and D. A. Clausi, "Automatic Registration of SAR and Visible Band Remote Sensing Image," *IEEE IGARS'02* , vol. 3, pp. 1331-1333, 24-28 June, 2002.

[91] H. Li, and Y. Zhou, "Automatic EO/IR Sensor Image Registration," *IEEE Inter. Conf. On Image Processing*, vol. 3, pp. 240-243, Oct. 1995.

[92] H. Li, B. S. Manjunath, and K. Mitra, "A Contour-Based Approach to Multisensor Image Registration," *IEEE Trans. On Image Processing*, vol. 4, no. 3, pp.320-334, 1995.

[93] B. Boashash, "Estimating and Interpreting The Instantaneous Frequency of a Signal – Part 1: Fundamentals," *Proc. of IEEE*, vol. 80, no. 4, pp. 520-538, April 1992.

[94] B. Boashash, "Estimating and Interpreting The Instantaneous Frequency of a Signal – Part 1: Algorithms and Applications," *Proc. of IEEE,* vol. 80, no. 4, pp. 540-568, April 1992.

[95] G. H. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Dordrecht, The Netherlands: Kluwer Academics, 1995.

[96] J. Liu, B. C. Vemuri, and J. L. Marroquin, "Local Frequency Representations for Robust Multimodal Image Registration," *IEEE Trans. on Medical Imaging*, vol. 21, no. 5, pp. 462-469, May 2002.

[97] J. Liu, B. C. Vemuri and F. Bova, "Multi-modal Image Registration Using Local Frequency**,"** *Fifth IEEE Workshop on Applications of Computer Vision*, pp. 120–125, 2000.

[98] D. J. Fleet and A. Jepson, "Computation of Component Image Velocity from Local Phase Information," *Inter. Journal of Computer Vision*, vol. 5, no.1, pp. 77-104, 1990.

[99] D. Gabor,. "Theory of Communications," *Journal of International Electrical Engineers*, vol. 93, pp. 429-457, 1946.

[100] J. G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by 2D Visual Cortical Filters," *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160-1169, 1985.

[101] J. P. Jones and L. A. Palmer, "An Evaluation of two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex," *J. Neurophysiol.*, vol. 85, pp. 1233-1258, 1987.

[102] J. G. Daugman, "Spatial Visual Channels in the Fourier Space," *Vision Research*, vol. 24, pp. 891-910, 1984.

[103] N. V. S. Graham. *Visual Pattern Recognition*. Oxford university Press, 2001

[104] J. Fdez-Valdivia, J.A. Garcia, J. Martinez-Baena and Xose R. Fdez-Vidal, "The Selection of Natural Scales in 2D Images Using Adaptive Gabor Filtering," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 458-469, 1998.

[105] J. Martinez-Baena, J.A. Garcia, and Xose R. Fdez-Vidal, "A New Image Distortion Measure Based on a Data-Driven Multisensor Organization," *Pattern Recognition*, vol. 31, no. 8, pp. 1099-1116, 1998.

[106] J. Kamarainen, V. Kyrki, and H. Kalviainen, "Fundamental Frequency Gabor Filters for Object Recognition," *IEEE Proc. Of 16th Inter. Conf. On Pattern Recognition*, vol. 1, pp. 628-631, 2002.

[107] J. B. Antoine Maintz, P. A. Van den Elsen, and M. A. Viergever, "Comparison of Edge-Based and Ridge-Based Registration of CT and MR Brain Images," *Medical Image Analysis*, vol. 1, no. 2, pp. 151-161, 1996.

[108] W. M. Theimer, "Phase-based Binocular Vergence Control and Depth Reconstruction Using Active Vision," *Journal of Computer Vision, Graphics, and Image Processing*, vol. 60, no.3, pp.343-358, Nov. 1994.

[109] S. Gold, A. Rangarajan, C. Lu, S. Poppu, and E. Majolseness, "New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence," *Pattern Recognition*, vol. 31, no. 8, pp. 1019-1031, 1998.

[110] A. Rangarajan, E. Mjolsness, S. Pappu, L. Davachi, P. S. Goldman-Rakic, and J. S. Duncan. "A Robust Point-Matching Algorithm for Autoradiograph Alignment," *The 4th International Conference on Visualization in Biomedical Computing*, Hamburg, Germany, September 22-25, pp.277-286, 1996.

[111] P. David. D. DeMenthon, and R. Duraiswami, and H. Samet, "Simultaneous Pose and Correspondence Determination Using Line Features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 424-431, 2003.

[112] D. M. Mount, N. S. Netanyahu, and J. Le Moigne, "Improved Algorithms for Robust Point Matching and Applications to Image Registration," *Proc. of the Fourteenth Annual Symposium on Computational Geometry*, Minneapolis, Minnesota, pp. 155-164, 1998.

[113] D. M. Mount, N. S. Netanyahu, and J. Le Moigne, "Efficient Algorithms for Robust Feature Matching," *Pattern Recognition*, vol. 32, pp. 17-38, 1999.

[114] L. S. Shapiro and J. M. Brady, "Feature-based Correspondence: An Eigenvector Approach," *Journal of Image and Vision Computing*, vol. 10, no.5, pp. 283-288, June 1992.

[115] S. Ranade and A. Rosenfeld, "Point Pattern Matching by Relaxation," *Pattern Recognition*, vol.12,pp. 269-275, 1980.

[116] S. Umeyama, "Least-Square Estimation of Transformation Parameters Between Two Points Patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376-380, 1991.

[117] Q. Tian and M. N. Huhns, "Algorithms for Sub-pixel Registration," *Journal of Computer Vision, Graphics, and Image Processing*, vol. 35, pp. 220-223, 1986.

[118] V. N. Dvorcenko, "Bounds on (Deterministic) Correlation Functions with Applications to Registration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 206-213, 1983.

[119] J. J. Pearson, D. C. Hines, S. Golosman, and C. D. Kuglin, "Video Rate Image Correlation Processor," *Proc. SPIE*, vol. 119, pp. 197-205, 1977.

[120] H. H. Nagel, "On the Estimation of Optical Flow: Relations between Different Approaches and Some New Results," *Artificial Intelligence*, vol. 33, pp. 299-324, 1987.

[121] A. Verri, F. Girosi, and V. Torre, "Differential Techniques for Optical Flow," *Journal of Optical Society of America: A*, vol. 7, pp. 912-922, 1990.

[122] D. J. Heeger, "Model for Extraction of Image Flow," *Journal of Optical Society of America: A*, vol. 4, pp. 1455-1471, 1987.

[123] S. P. Kim and W. Y. Su, "Sub-pixel Accuracy Image Registration by Spectrum Cancellation," *In Proc. IEEE Intern. Conf. On Acoustics, Speech, and Signal Processing*, vol. 5, pp. 153-156, 1993.

[124] P. Thevenaz, U. E. Ruttimann, and M. Unser, "A Pyramidal Approach to Sub-pixel Registration Based on Intensity," *IEEE Trans. Image Processing*, vol. 7, pp. 27-41, Jan, 1998.

[125] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of Phase Correlation to Subpixel Registration," *IEEE Trans. On Image Processing*, vol. 11, no. 3, March 2002.

[126] W. S. Hoge, "A Subspace Identification Extension to the Phase Correlation Method," *IEEE Trans. On Medical Imaging*, vol. 22, no. 2, pp. 277-280, Feb. 2003.

[127] V. Argyriou and T. Vlachos "Estimation of Sub-pixel Motion Using Gradient Cross-Correlation," *Electronics Letters*, vol. 39, no. 13, 26th June 2003.

[128] V. Argyriou and T. Vlachos, "Using Gradient Correlation for Sub-pixel Motion Estimation of Video Sequences," *In Proc. IEEE Intern. Conf. On Acoustics, Speech, and Signal Processing*, pp. III-329 - III-332, 2004.

[129] M. Elbakary and M. K. Sundareshan, "Accurate Representation of Local Frequency Using Computationally Efficient Gabor Flter Fusion Approach With Application to Image Registration," *to Appear in Pattern Recognition Letters*.

[130] The USC-SIPI image database, http://sipi.usc.edu/database/.

[131] R. Tsai and T. Huang, "Multiframe Image Restoration and Registration," *In Advances in Computer Vision and Image Processing*, vol. 1, pp. 317-339, 1984.

[132] D. Capel and A Zisserman, "Computer Vision Applied to Super Resolution," *IEEE Signal Processing Magazine*, pp. 75-86, May 2003.

[133] M. Elad and A. Feuer, "Super-Resolution Reconstruction of Image Sequences," *IEEE Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 817-834, 1999.

[134] M. Elad and Y. Hel-Or, "A Fast Super-Resolution Recognition Algorithm for Pure Translation Motion and Common Space-Invariant Blur," *IEEE Trans. On Image Processing*, vol. 10, no. 8, pp. 1187-1193, August 2001.

[135] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-Resolution Images from Low-Resolution Compressed Video," *IEEE Signal Processing Magazine*, pp. 37-48, May 2003.

[136] S. P. Kim, N. K. Bose and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image from Noisy under-sampled Multiframes," *IEEE Trans. On Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013-1027, June 1990.

[137] N. Nguyen, P. Milanfar, and G. Golub, "A Computationally Efficient Super-resolution Image Reconstruction Algorithm," *IEEE Trans. On Image Processing*, vol. 10, no. 4, pp. 573-583, April 2001.

[138] E. S. Lee and G. Kang, "Regularized Adaptive High-Resolution Image Reconstruction Considering Inaccurate Sub-pixel registration," *IEEE Trans. on Image Processing*, vol. 12, no. 7, pp. 826-837, July 2003.

[139] S. C. Park, M. K. Part, and M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, pp. 21-36, may 2003.

[140] M. Elad and A. Feuer, "Restoration of a Single Super-Resolution Image from Several Blurred, Noisy, and Undersampled Measured Images," *IEEE Trans. on Image Processing*, vol. 6, no. 12, Dec. 1997.

[141] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Under-sampled Images," IEEE *Trans. On Image Processing*, vol. 6, no. 12, pp. 1621-1633, Dec. 1997.

[142] A. Gruen, E. P. Baltsavias, and O, Henricsson, *Automatic Extraction of Man-made Objects from Aerial and Space Images I.* Birkhauser Verlag, Switzerland 1997.

[143] P. Wunsch and G. Hirzinger, "Registration of CAD Models to Images by Iterative Inverse Perspective Matching," *Proc. of IEEE Int. Conf. on Pattern Recognition, ICPR'96*, vol. 1, pp. 78-83, 1996.

[144] M.R. Stevens and J. Ross Beveridge, "Precise Matching of 3-D Target Models to Multisensor Data," *IEEE Trans. on Image Processing*, vol. 6, no. 1, pp 126-142, January 1997.

[145] Q. Zheng, S. Der, and R. Chellappa, "Model-based Target Recognition in Pulsed LADAR Imagery," *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, pp. 515-520, 1998.

[146] Q. Zhen, S. Der, H. Mahmoud, "Model-based Target Recognition in Pulsed LADAR Imagery," *IEEE Trans. on Image Processing*, vol. 10, no. 4, pp. 565-572, 2001.

[147] M. I. Miller, U. Grenander, J. A. O'sulliran, and D. L. Snyder, " Automatic Target Recognition Via Jump-diffusion Algorithms", *IEEE Trans. on Image Processing*, vol. 6, no. 1, pp. 157-174, 1997.

[148] T. Tipdecho, "Automatic Image Registration Between Image and Object Spaces," *Proc. of the Open Sources GIS-GRASS Users Conferences*, Trento, Italy, Sept. 2002.

[149] M. K. Hamilton and T.A Kipp. "Model-based Multi-sensor Fusion," *IEEE Proc. of The Twenty-Seventh Asilomar Conference on Signal, Systems and Computers*, vol. 1, pp. 283-289, Nov. 1993.

[150] P. Bhattacharya. "Aspect Angle Estimation of Targets in Forward-Looking Infrared Imaging Using the Model-based Vision Approach," *Optical Engineering*, vol. 33, no. 10, pp.3334-3341, October 1994.

[151] A. D. Lanterman, M. I. Miller, and D. L. Snyder "The Unification of Detection, Tracking, and Recognition for Millimeters Wave and Infrared Sensors," *RADAR/LADAR processing, Proc. of SPIE*, vol. 2562, pp.150-161, 1995.