Image Registration Via Entropy Consideration and Data Fusion

Von Fachbereich Elektrotechnik und Informatik
der Universität Siegen zur Erlangung
des Akademisches Grades
Doktor der Ingenieurwissenschaft

vorgelegt von
Anthony Amankwah (M.Sc)
geboren am 1.Oktober 1969 in Ghana

1. Gutachter: Prof. Dr. Ing. habil. Otmar Loffeld
2. Gutachter: Prof. Dr. Ing. Mihai Datcu

CONTENTS.....

.

ZUSAMMENFASSUNG

Die Bildregistrierung ist ein Prozess, durch den die genaueste Übereinstimmung zwischen zwei Bildern bestimmt wird. Dabei können die Bilder zu gleichen oder unterschiedlichen Zeitpunkten, durch denselben oder verschiedene Sensoren vom gleichen oder von einem anderen Betrachtungswinkel aufgenommen worden sein. Es ist einer der entscheidenden Schritte in der Fernerkundung. Die Bildregistrierung erfordert einen enormen Rechenaufwand. Dieser resultiert nicht nur aus der Komplexität der Berechnungen, sondern auch aufgrund der immer weiter zunehmenden Auflösung der Bilder. Für eine hohe Genauigkeit und Robustheit als auch für einen möglichst niedrigen Berechnungsaufwand sind ein passendes Ähnlichkeitsmaß, die Verminderung der “Search Data“ und eine robuste Suchraumstrategie erforderlich.

Die Reduzierung der Search Data kann durch die Verwendung einiger Subimages für die Bildregistrierung erreicht werden. Wir schlagen ein neues Maß, genannt Alignabilty, vor, das die Fähigkeit von Subimages zeigt, robuste und zuverlässige Ergebnisse zur Verfügung zu stellen. Wir vergleichen dieses Merkmal mit der Entropie, der Varianz bzw. dem Betrag des Gradienten. Wir zeigen, dass die Verwendung von Alginability nicht nur zuverlässige, sondern auch robuste Ergebnisse in der Bildregistrierung erzeugt.

In der vorliegenden Dissertation, wird die Verwendung von Mutual Information als ein Ähnlichkeitmaß untersucht. Wir zeigen die Auswirkung der Bin-Size auf die Mutual Information. Durch die Erhöhung der Bin-Anzahl können mehr Bildintensitäten unterschieden werden. Anderseits wird durch die Verringerung der Bin-Anzahl die Mutual Information verrauscht oder schlägt fehl. Wir schlagen

zwei Methoden vor, die die Auswahl der Bin Size Größe formalisieren. Wir zeigen, wie die Verwendung der vorgeschlagenen Methoden die Robustheit der Mutual Information erhöht.

Die Mutual Information hat sich in den vergangenen Jahren zu einem populären Ähnlichkeitsmaß in der Bildregistrierung entwickelt. Leider ignoriert Mutual Information die in Bildern enthaltene Ortsinformation wie z.B. Kanten und Ecken, die zum Matching von Bildern nützlich sein könnten. Es werden nur die Beziehungen zwischen korrespondierenden Pixeln betrachtet, nicht die der näheren Umgebung. Jedoch ist es notwendig, sowohl quantitative als auch qualitative Information in der Registrierung zu betrachten. Wir stellen ein neues Ähnlichkeitsmaß vor. Dieses wird Spatial Mutual Information genannt und kombiniert die Mutual Information mit einer Gewichtsfunktion, basierend auf dem Bildgradienten, der Bildvarianz und der Bildentropie einer lokalen Bildregion. Hervorspringende Pixel in Gebieten mit einem hohen Gradienten, einer hohen Varianz und einer hohen Entropie tragen mehr zur Schätzung der Mutual Information von Bildpaaren bei die registriert werden. Wir zeigen, dass die Spatial Mutual Information gegenüber Rauschen robuster ist als die Mutual Information. Wir zeigen auch, dass die Spatial Mutual Information nicht nur robuster ist als die Mutual Information, sondern auch zuverlässiger in der Registrierung von multitemporalen Bildern ist.

Eine auf dem Robbins-Monro stochastischen Approximationsalgorithmus beruhende Suchraumstrategie wird auch eingeführt. Die Ergebnisse zeigen, dass der Algorithmus nicht nur schnell, sonder auch robust ist.

ABSTRACT

Image registration is a process by which the most accurate match is determined between two images, which may have been taken at same or different times, by the same or different sensors, from the same or different viewpoints. It is one of the crucial steps in remote sensing. Image registration requires intensive computational effort, not only because of its computational complexity but also the increasing resolution of images. For high accuracy and robustness as well as low computational cost a suitable similarity metric, a reduction in search data and a robust search space strategy is needed.

Reduction in search data can be achieved by using a few subimages for image registration. We propose a new measure, called alignability, which shows the ability of subimages to provide robust and reliable results. We compare this feature to entropy, variance and gradient magnitude respectively. We show that using alignability produces not only reliable but robust results in image registration.

In the dissertation, we investigate mutual information as a similarity metric. We show the effect of bin size on mutual information. Increasing the number of bins discriminates more image intensities, on the other hand, in decreasing the number of bins mutual information becomes noisy or even fail. We propose two methods which formalize the selection of bin size. We show that using the proposed methods increases the robustness of mutual information.

Mutual information has emerged in recent years as a popular similarity metric in the registration of images. Unfortunately, it ignores the spatial information contained in the images such as edges and corners that might be useful in the matching of images. It takes into account only the relationships between corresponding individual pixels and not those of each pixel's neighborhood. However, it is essential to consider both quantitative and qualitative information in the registration of images. We propose a new similarity metric, called spatial mutual information, which combines mutual information and a weighting function based on image gradient, image variance, and image entropy of local regions. Salient pixels in the in regions with high gradient, high variance and high entropy contribute more in the estimation of mutual information of image pairs being registered. We show that spatial mutual information is more robust to noise than mutual information. We also demonstrate that the spatial mutual information is not only more robust than mutual information but more reliable in the registration of multitemporal images.

A search space strategy based on Robbins-Monro stochastic approximation algorithm is also introduced. Results show that the algorithm is not only fast but robust as well.

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Prof Otmar Loffeld and Prof Mihai Datcu for providing me with constant guidance and inspiration throughout the entire period of my thesis work.

They helped me to understand the basic concepts and theories of signal processing that were the building blocks for this research. Their suggestions also helped me shape my ideas into reality. This research work was a great learning experience for me, and I am grateful to them for providing me with an opportunity to work with them.

I wish to express my gratitude to my colleagues, H. Nies and V. Peters for their technical support. I thank S. Knedlik, Mrs Niet-Wunram and Amaya who has been most helpful to me during the various difficult phases of this work.

There are of course many more people I want to thank. In no particular order: Tomi, Frank, Arun, Yvonne, Ulrich.

Last but not the least, many thanks to my parents and siblings Clement, Paul, Evelyn and Bernard for their prayers and support during the trying times.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Image registration is the process by which the most accurate match is determined between two images, which may have been taken at the same or different times, by the same or different sensors, from the same or different view points. It is one of the essential steps in many image processing applications such as medical imaging, computer vision and remote sensing. In all the areas image registration can be used to find changes in multitemporal images, build 3D models from 2D images, taken from different view points and track objects. In remote sensing, with the increasing number of platforms and remote sensing missions, different sensors may observe the same scene at the same time. These sensors produce images in different spectral ranges or at different resolutions, over multiple times and thus give a large amount of redundant or complementary data. The fusion of all these data will not only allow for better analysis but also very good results. Georeferencing can be used as a first step in integrating the information from different sources in remote sensing. Remote sensing is, however, georeferenced to within a few pixels. We focus on automatic image registration, which refines the accuracy to at least 1 pixel. For application such as formation of 3D images, a registration error of 0.014 resolution cells for a cross track baseline of 100m corresponds to elevation error of 100m [1].

1.2 Image registration methods

Image registration can be broadly divided into area-based and featured based methods [2]. In area-based methods emphasis is put on the feature matching step rather than the detection of features. Pixel values are used for the matching. Feature-based methods also estimate transformation by matching salient features such as lines or points. Both methods include the following steps:

1) Feature detection. This involves the extraction of features to be used for the matching process. Some distinctive features include edges, contours and lines of intersection.

2) Feature matching. In this section the correspondence between the features detected in the input image and features detected in the reference image. Similarity metrics such as mutual information and feature descriptors are used for this purpose.

3) Transformation model estimation. The mapping functions between the input and reference images are estimated by matching the corresponding features.

4) Image resampling and transformation. The input image is transformed according to the estimated mapping function

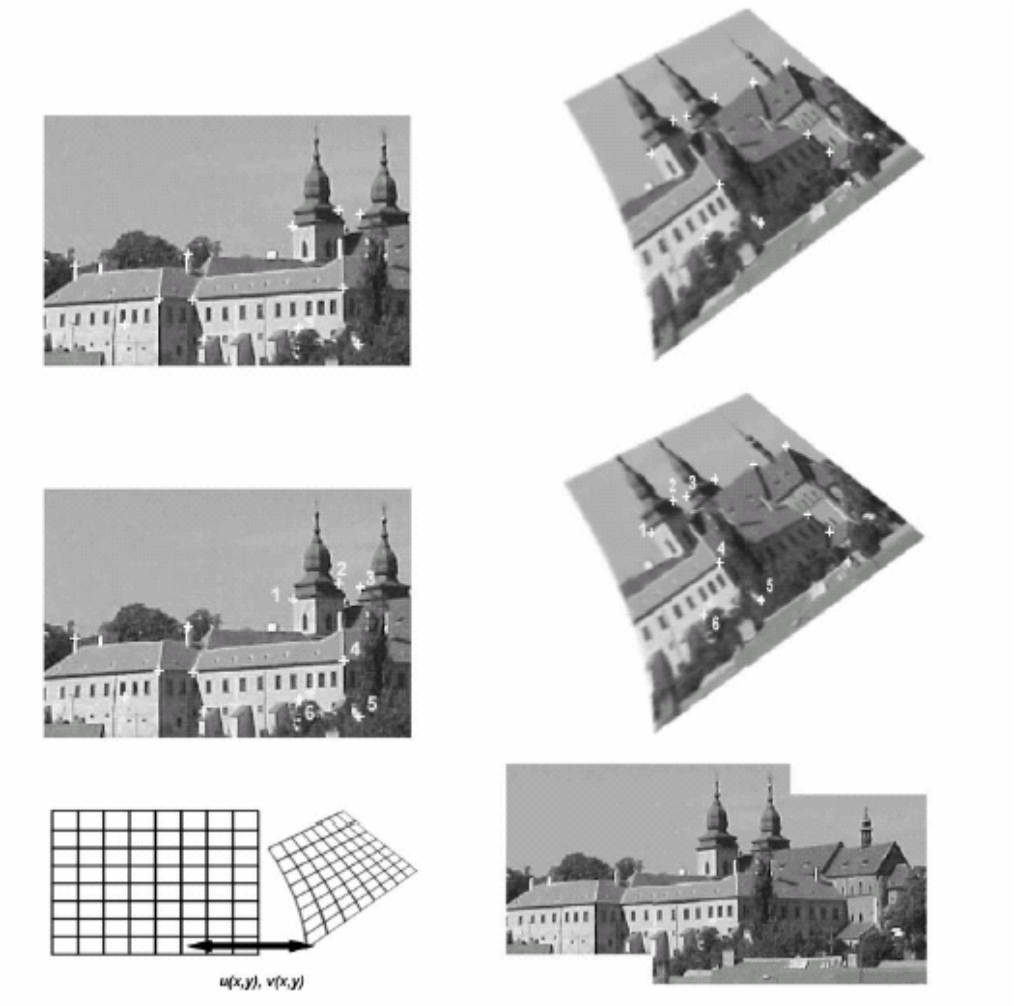


Figure 1.1 The four stages in image registration source [2]

1.3 Transformations

Distortions are sources of misregistration. There are two main sources of distortions in remote sensing imagery [3]. These are radiometric distortions and geometric distortions. Radiometric distortions are caused by the effects of atmosphere on remote sensing imagery as well as instrumentation errors. Sources of geometric distortions include change in platform altitude, velocity and

aspect ratio. The geometric distortions can be classified into translation, rotation, scaling and shearing [4].

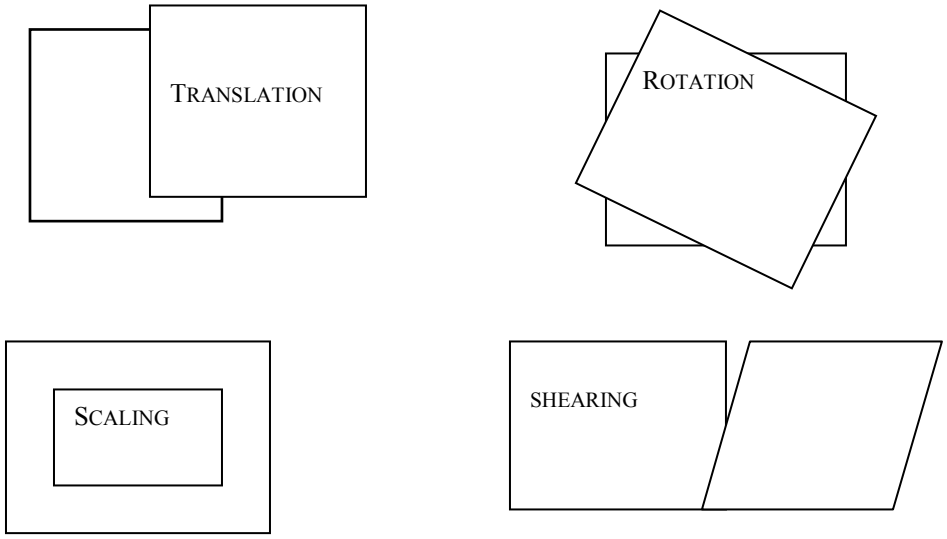


Figure 1.2 Examples of geometric transformations

Sensors viewing from a different position may cause stretching and shearing and sensors on different altitude may cause scaling. Rigid transformations are among the most common general transformations. They account for sensor or object movement in which objects in images retain their relative shape and size [4]. Rigid transformation consists of scaling, translation and rotation. This can be written as

$$T(x,y)=(s)\cdot\begin{bmatrix}\cos\theta & \sin\theta & T_x\\ -\sin\theta & \cos\theta & T_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \qquad 1.1$$

where *T* is the transformation, θ is the rotation angle, (*x, y*) is the coordinate of pixels, s is the scaling factor, and *Tx* and *Ty* are translations in the x-direction and y-direction respectively.

Affine transformations are linear transformations. They map straight lines into straight lines. An affine transformation consists of all the transformations in figure 1.2. The general affine transformation can be expressed as

$$T(x,y)=\begin{bmatrix} a & b & T_x \\ c & d & T_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} . \qquad 1.2$$

The transformation does not preserve angles and lengths. The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ can be rotation, shear, or scaling.

The shear component consist of affine transformation can be represented by

$$Shear_x=\begin{pmatrix} sx & 1 \\ 0 & 1 \end{pmatrix} \quad , \quad Shear_y=\begin{pmatrix} 1 & sy \\ 0 & 1 \end{pmatrix} . \qquad 1.3$$

The scale can also be expressed as

$$Scale=\begin{pmatrix} s1 & 0 \\ 0 & s2 \end{pmatrix} . \qquad 1.4$$

1.4 Robust and efficient image registration methods

For reliable and computationally efficient image registration, a robust similarity metric, search data strategy and a search space strategy is required.

The similarity measure chosen is an essential part of image registration. The choice of similarity measures depends on whether image pairs being registered are unimodal or multimodal. Some of the proposed similarity measures are mutual information, correlation coefficient, absolute mean difference and partial Hausdorf distance [16].

Search data strategies are methods used to reduce the size of data use image registration. They are employed to reduce computational cost and efficiency. Some of the methods include subimage selection, Fourier transform and wavelet transform.

Search space strategies are optimization algorithms used in finding the maximum (minimum) of similarity measures. Some of these methods include simulated annealing, Spall algorithm and Robbins-Monro stochastic approximation.

1.5 Problem definition and goals

We model the problem as follows: let the reference image and input image be *R* and *I* respectively.

- Find the geometric transformation *T*, such that the similarity *S* between the images is maximum

$$Arg \max_{T} \{ S(R, T(I)) \} \quad . \qquad 1.5$$

- Develop fast, robust and accurate algorithms for registering *I* and *R* by
  - using a robust and reliable similarity metric(s)
  - exploiting search data and search space reduction techniques

- Design experiments to validate the efficiency and applicability of the image registration algorithms.

1.6 Overview of thesis

The second chapter gives an overview of related works and other methods used in image registration. The third chapter discusses automatic subimage selection and experimental results. The fourth chapter gives an overview of probability theory and talks about the maximization of mutual information in image registration. The fifth chapter discusses the incorporation of spatial information into mutual information. The sixth chapter presents the results of the registration test images. The seventh chapter summarizes the findings and suggests future research directions.

CHAPTER 2

RELATED WORK

The importance of image registration in many image processing applications as well as its complicated nature has made it the topic of much recent research. One of the first survey papers [5] focuses on the image registration methods based on image correlation. Registration techniques applied in remote sensing are summarized in [6, 7]. An exhaustive review of the techniques in image registration is presented in [2] and [4]. This chapter focuses on similarity metrics, search space and search data strategies, and methodologies in image registration.

2.1 Elements of image registration

Image registration searches for the best match between image pairs using similarity measures such as mutual information, correlation coefficient, Hausdorf distance and sum of absolute differences. The choice of the similarity metric depends on the type of images being registered. Search data strategies are used to reduce the amount of data for registration without compromising the reliability of registration. Examples of search data strategies are subimage selection, wavelet transform and Fourier transform. Search space strategies are optimization algorithms used to enhance computational efficiency. Some examples are steepest gradient algorithm, Spall algorithm and simulated annealing.

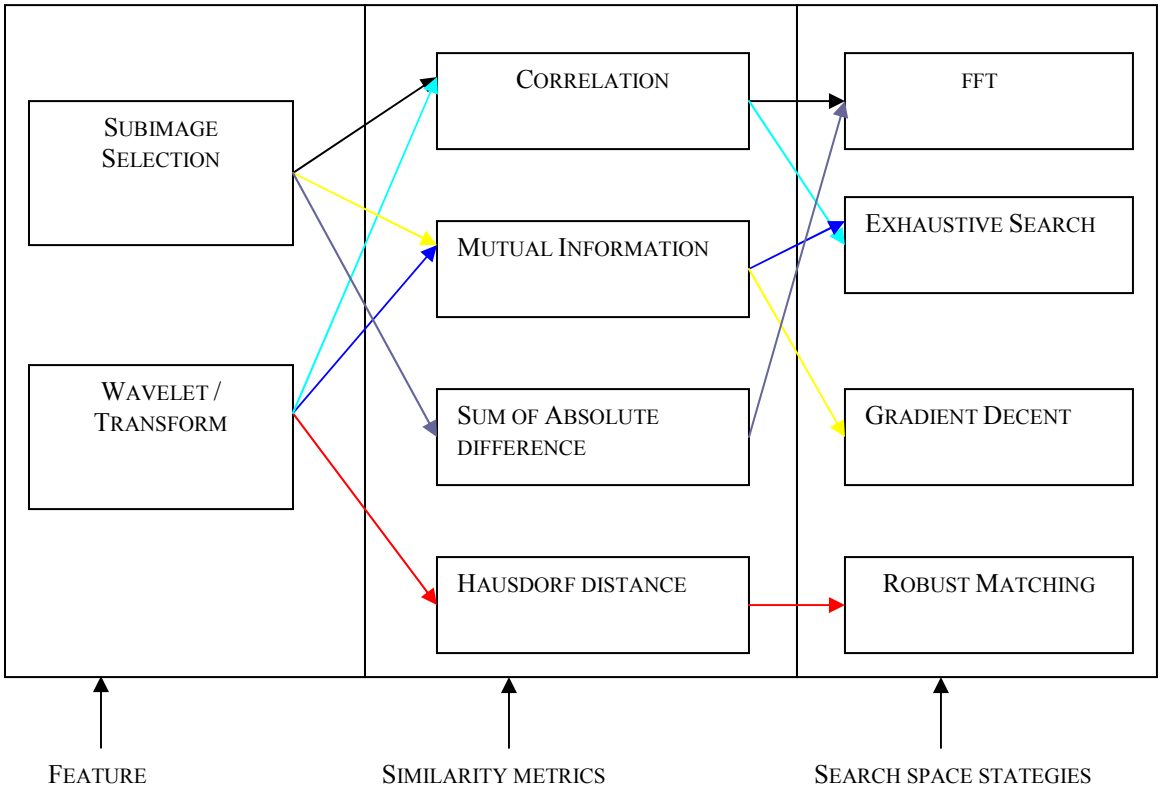


Figure 3. Diagram showing possible methods of implementing image registration

2.2 Similarity metrics

Similarity metrics are used in image registration to find, how similar or the accurate match between the input and reference image is. The similarity functions are often chosen depending on the type of image and the combination of modalities. Unimodal images are images taken by the same sensor. Multimodal images are images taken by different sensors. In multimodal images the same scene may have different intensity values. A similarity measure for

unimodal images would not necessarily be applicable for multimodal images. In this section, we discuss some popular similarity metrics.

2.2.1 Sum of absolute differences

The sum of absolute intensity differences (*SD*) is of order one and is defined by

$$SD = \sum_i \sum_j \left| \frac{a_{ij} - b_{ij}}{N} \right| \qquad 2.1$$

where a_{ij} and b_{ij} are the pixel values of master image and slave image respectively. *N* is the number of pixels. Since the similarity measure finds the sum of absolute measure between the images it is required that the reference image and input image be of the same modality. SD is very fast to calculate and has a time complexity of order one *O(N).* A study by Penny et al [8] shows that using gradients instead of raw intensity values produce more accurate matches.

2.2.2 Correlation coefficient

The correlation coefficient is one of the most popular similarity metrics. It is expensive to estimate and has time complexity of order two $O(N^2)$. Correlation coefficient is defined by

$$CC = \frac{\sum_i \sum_j (a_{ij} - mean(a)) * (b_{ij} - mean(b))}{\left(\sum_i \sum_j (a_{ij} - mean(a))^2 * \sum_i \sum_j (b_{ij} - mean(b))^2 \right)^{1/2}} \qquad 2.2$$

where a_{ij} and b_{ij} are the pixel values of images A and B respectively. The denominator is the normalizing factor. Positive values will show dependency

(similarity) of image pairs. *CC* increases as the similarity between image pair increases. *CC* is between -1 and 1. When the image pair being registered completely depends on each other, *CC* equals 1 or -1. CC is resistant to some intensity difference between images. Therefore, even when reference image and input image have intensity differences, as long as they decrease and increase together, a high correlation coefficient could be obtained. *CC* can produce more accurate results than *SD* when the images being registered show no rotational differences [9]. Fourier methods can also be used to estimate correlation coefficient. They exploit the Fourier representation of images in the frequency domain (Section 2.3.2).

Consider the numerator of equation 2.2, and assume that we have images $f_1(x,y) \equiv a_{ij} - mean(a)$, $f_2(x,y) \equiv b_{ij} - mean(b)$ in which the mean value has been removed,

$$CC^{num} = \sum_i \sum_j f_1(x,y) f_2(x,y). \qquad 2.3$$

Equation 2.3 is a convolution of the image with the reversed feature $f_2(-x,-y)$ and can be computed by

$$\Gamma^{-1}(\Gamma(f_1)\Gamma^*(f_2)) \qquad 2.4$$

where Γ is the Fourier transform. The complex conjugate accomplishes the reversal of feature via the Fourier transform property $\Gamma(f^*(-x)) = F(\omega)$.

2.2.3 Mutual information

Mutual information is a basic concept in information theory. It is also referred to as relative entropy. Collignon and Voila and Wells independently proposed the use of mutual information as a similarity metric in image registration [10], [11]. This similarity metric measures the statistical dependence between two data sets. It is also suitable for the registration of image pairs from different sensors. The Mutual information (*MI*) between two random variables *X* and *Y*, is given by

$$MI(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad , \qquad 2.5$$

where *p(x)* and *p(y)* are the probabilities of variable x and y respectively and *p(x, y)* is their joint probability. It is applied in the context of image registration to measure the amount of information that one image contains about the other. In our work, we use entropy to define the amount of information. Entropy (H(X), H(Y)) measures the dispersion of probability distributions. Figure 2.1 shows that, maximizing mutual information is equivalent to minimizing the joint entropy (H(X, Y)).

The maximization of mutual information criterion postulates that mutual information is maximal when images are correctly registered. Mutual information has demonstrated to be a very general and powerful similarity metric that can be applied automatically and very reliably, without prior pre-processing, on a large variety of applications [12].

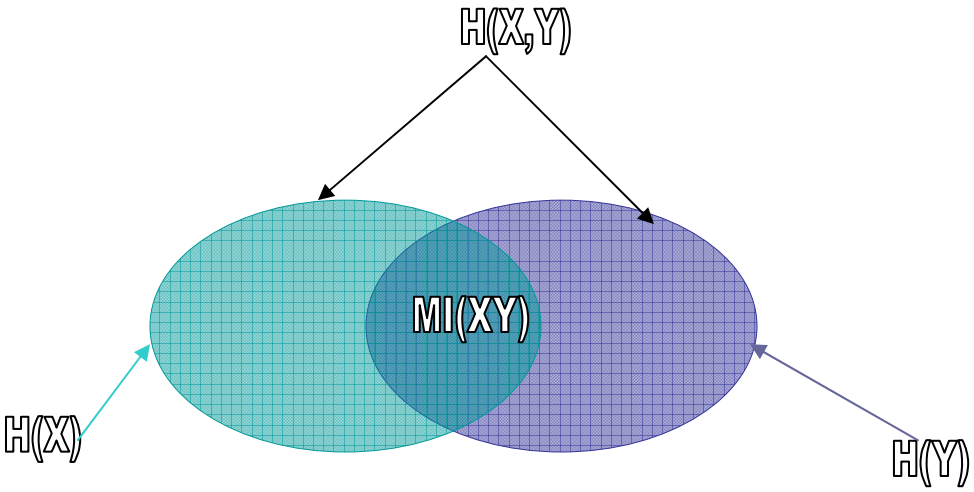


Figure 2.1 Relationship between entropy (H(X), H(Y)), joint entropy (H(X,Y)) and mutual information

We will further discuss mutual information in the following chapters.

2.3 Search Data Strategies

Search data strategies are used to reduce the amount of data used in image registration. This helps to increase the computation efficiency. Some of the methods used in the reduction of search data include automatic subimage selection and using two dimensional affine transformations such as wavelet transform (Coarse-to-fine) and Fourier transform.

2.3.1 Automatic subimage selection

Instead of using the whole image pairs being registered, subimages or templates can be used to improve computational efficiency [13]. Once corresponding subimages are found in image pairs, their centres are used as corresponding control points to determine the registration parameters. Subimages can be selected manually [14] or automatically [15].

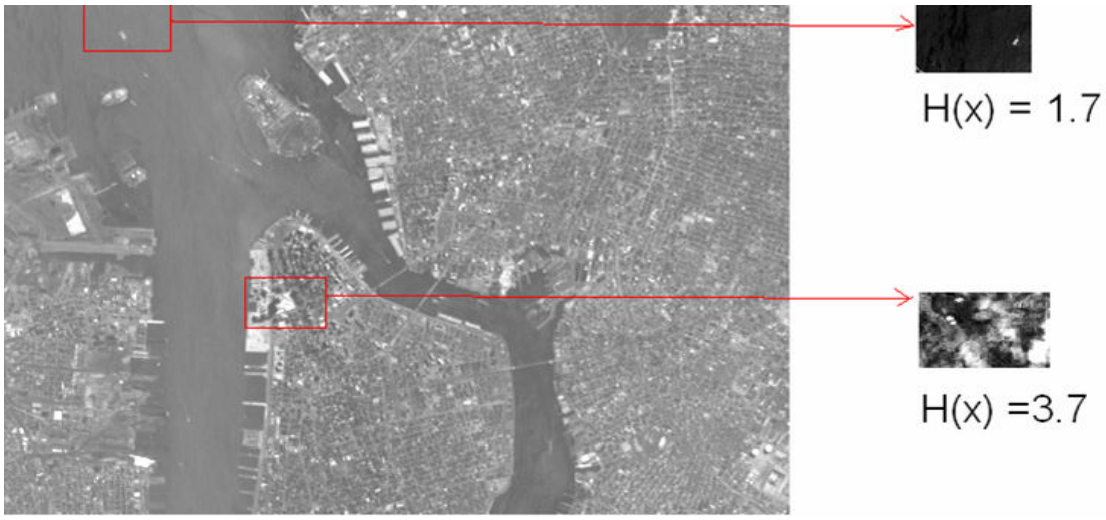


Figure 2.2 Image showing subimage with highest and lowest entropy respectively. The image was divided into 64 parts. The entropies of each of the 64 subimages were calculated. The subimages with the highest and the lowest entropies are selected

To enhance robustness in registration, the choice of subimage should be unique. Some of the features used to determine uniqueness include high edge content, high information content and high variance.

2.3.2 Fourier transformation

The Fourier transform decomposes images in the frequency domain. This can accelerate computational speed if correlation is used as the similarity metric [2]. For a square image of size *N x N* the discrete two dimensional transform is given by

$$F(\omega_1,\omega_2)=\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(\frac{x\omega_1}{N}+\frac{y\omega_2}{N})} \qquad 2.6$$

where *f (x,y)* is the image in the spatial domain and the exponential term is the corresponding basis function of $F(\omega_1, \omega_2)$ in the Fourier space. The cross power spectrum of reference and input images and the location of the peaks in its inverse is determined. The cross power is the cross correlation of the Fourier transform onto the frequency domain. Let $f_1(x,y)$ and $f_2(x,y)$ be to images to be registered. Let also:

$$f_2(x,y) = f_1(x+x_o, y+y_o) \qquad 2.7$$

Then according to the Fourier shift theorem:

$$F_2(\omega_1,\omega_2) = F_1(\omega_1,\omega_2)e^{j(\omega_1 x_o+\omega_2 y_o)} \qquad 2.8$$

Or equivalently:

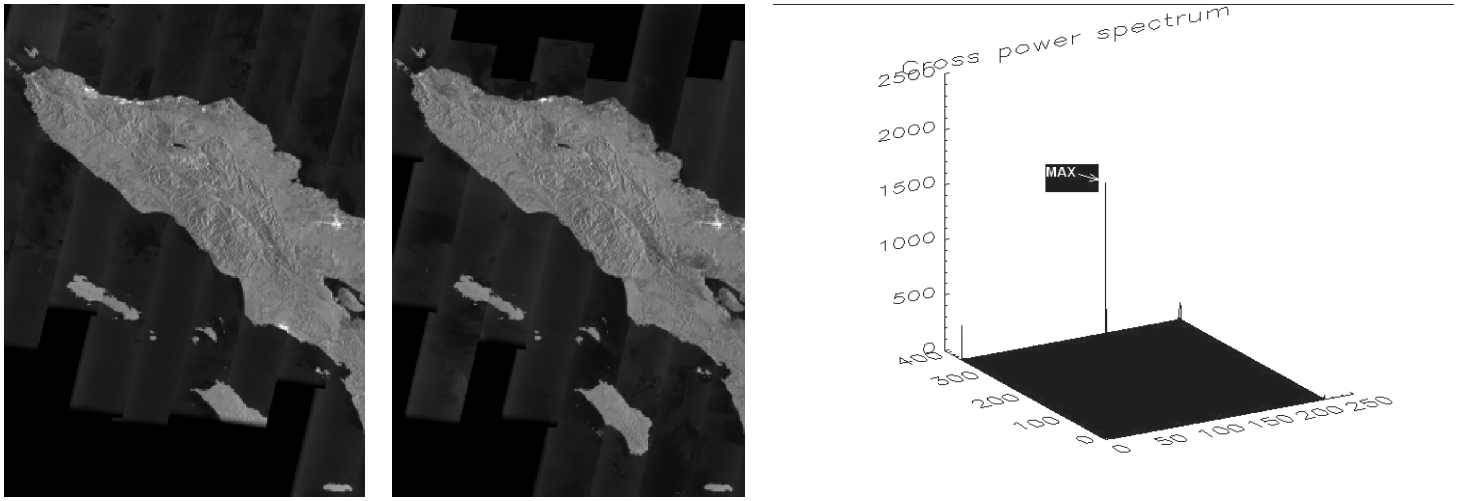
$$\frac{F_2(\omega_1,\omega_2)F_1^{*}(\omega_1,\omega_2)}{\left|F_2(\omega_1,\omega_2)F_1^{*}(\omega_1,\omega_2)\right|} = e^{j(\omega_1 x_o+\omega_2 y_o)} \qquad 2.9$$

where * denotes the complex conjugate and the left hand side is referred to as the cross power spectrum of the two images.

It is now simple to determine the shift x_o and y_o, since the inverse Fourier transform of the right hand side is a Dirac delta function centred at (x_o, y_o):

$$\Gamma^{-1}\left(\frac{F_2(\omega_1,\omega_2)F_1^{*}(\omega_1,\omega_2)}{\left|F_2(\omega_1,\omega_2)F_1^{*}(\omega_1,\omega_2)\right|}\right)=\Gamma^{-1}\left(e^{j(\omega_1 x_o+\omega_2 y_o)}\right)=\delta(x_o,y_o) \qquad 2.10$$

Replacing the Fourier transform by its finite size discrete version, while assuming periodic extension of image and also replacing the Dirac delta function by the unit impulse, it has been shown in [19] that the result still holds. Figure 2.3 shows an example of multitemporal images with no displacement (0, 0) along the two axes. The computational time savings are more significant if the images being registered are large. The method shows robustness against frequency dependent and correlated noise and non-uniform varying illumination disturbances.



(a) Reference Image (320x260)

(b) Input image (320x260)

(c) The cross power spectrum (transformed in spatial domain) showing correct registration point. The maximum at (0,0) identifies the matching position

Figure 2.3 Image pairs from JERS1 and phase diagram

Phase correlation was originally introduced for the registration of translated images. De Castro *et al* [18] introduced an extension for additional rotation transform. A further extension of phase correlation to subpixel registration by means of the analytic phase correlation on down sampled images was introduced by Foroosh et al [19].

2.3.3 Wavelet transform

In using wavelet transforms the amount of data used for image registration can be reduced by initially searching at the lowest resolution and then proceeding to higher resolutions where results are refined. Wavelet transforms generates multi-resolution representation of image data [16]. Wavelet-based multiresolution preserves most of the features even a lower resolutions. It also highlights strong features at high resolutions [17].

In two dimensions, a two-dimensional scaling function $\Phi(x,y)$, and three two-dimensional wavelets, $\Psi^H(x,y)$, $\Psi^V(x,y)$, and $\Psi^D(x,y)$ are required. Each is the product of one-dimensional scaling function Φ and corresponding wavelet Ψ. Excluding products that produce one-dimensional results, like $\Phi(x)\Psi(x)$, the four remaining products produce the separable scaling function $\Phi(x)=\Phi(x)\Phi(y)$ and separable directional wavelets $\Psi^H(x,y)=\Psi(x)\Phi(y)$, $\Psi^V(x,y)=\Phi(x)\Psi(y)$, $\Psi^D(x,y)=\Psi(x)\Psi(y)$.

Given separable two-dimensional scaling and wavelet functions, the scaled and translated basis are as follows:

$$\Phi_{j,m,n}(x,y)=2^{j/2}\Phi(2^j x-m,2^j y-n) \qquad 2.11$$

$$\Psi^i{}_{j,m,n}(x,y)=2^{j/2}\Psi^i(2^j x-m,2^j y-n) \qquad i=\{H,V,D\} \qquad 2.12$$

where index *i* identifies the directional wavelets. The discrete wavelet transform of function $f(x,y)$ of size $M\times N$ is then

$$W_\Phi(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\Phi_{j_0,m,n}(x,y), \qquad 2.13$$

$$W_{\Psi}^{i}(j_0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\Psi_{j_0,m,n}^{i}(x,y)\qquad i=\{H,V,D\}, \qquad 2.14$$

where j_0 is an arbitrary starting scale. The $W_{\Phi}(j_0,m,n)$ coefficients define an approximation of *f(x, y)* at scale j_0. The $W_{\Psi}^{i}(j_0,m,n)$ coefficients add horizontal, vertical, and diagonal details for scales $j \geq j_0$. Normally, $j_0=0$ and $N=M=2^J$ so that $j=0,1,2,...,J-1$ and $m,n=0,1,2,...,2^{j-1}$.

2.4 Search space strategies

The search for the optimum transformation by exhaustive search is computationally expensive. The computational cost increases exponentially with both dimension of parameter space and dimension of data set. Exhaustive search become even more expensive when the goal is sub-pixel accuracy. The choice of search space strategy depends on the type of problem under consideration. Traditional nonlinear programming methods, such as the constrained conjugate gradient and back propagation in neural networks are well suited to deterministic optimization problems where the exact knowledge of the gradient of the objective function is known. However, in image registration, the exact knowledge and behavior of the objective function, which is the similarity metric during the registration process, is not known.

In this section we briefly review some search space strategies suitable for image registration.

2.4.1 Simulated annealing

Simulated annealing was first introduced in 1983 [20]. It is based on the analogy of heated metals. Simulated annealing performs a random search, though local, through the parameter space. At each step random modification to the parameters is proposed and the new objective function is evaluated. If the new

value of the objective function is higher than the previous one the parameter is updated. If the new objective function is lower than the older one, the parameters are accepted probabilistically. In order to generate annealing behaviour, the probability of acceptance is proportional to the exponential difference

$$p_{accept}(d) = \frac{\exp(-d/t)}{t} \qquad 2.15$$

where *d* is the temperature is the difference between the new objective function and the old objective function, t is the temperature that controls the likelihood that a good modification is accepted. This probability can be compared with a randomly generated number over the range [0, 1]. If $p_{accept}(d) \geq random[0...1]$ then the trial is accepted. The dependence on random numbers make simulated annealing a stochastic method.

The application of simulated annealing to finding the registration point in image registration requires the definition of four components (i) (ii) (iii) (iv):

(i) Problem Configuration: a definition of the suitable domain over which the optimum can be sought. This knowledge is often expressed in the form of constraint equations.

2. Neighborhood Configuration: a method of iteratively perturbing the design vector to create new trial points.

3. Objective function: a scalar equation (similarity metric) that weighs the entire design variable to provide a measure of the goodness for each trial point.

4. Cooling / Annealing Schedule: a methodology for specifying the maximum number of inner loop iterations and the manner in which the control parameter will be decremented in each iteration of the outer loop.

2.4.2 Random search

Random search is based on searching randomly in the domain of interest. It is perhaps the simplest among the stochastic approximation methods. There are several random search methods. Consider the problem to find the optimal of $\theta \in \Theta$ based on measurements of the objective function *L(θ)*. One of the simplest random search methods is the "blind random search" method. The approach can be implemented in batch (non-recursive) form simply by laying down a number of points in Θ and taking the value of *θ* yielding the lowest L(*θ)* value as our estimate of the optimum. The approach can be conveniently implemented in recursive form as illustrated below:

Step 0

(*Initialization*) Choose an initial value of *θ*, say $\theta_0 \in \Theta$ either randomly or deterministically. (If random, usually a uniform distribution is used.) Calculate $L(\theta_o)$ and set $i = 0$.

Step 1

Generate a new independent value $\theta_{new}(i+1) \in \Theta$, according to the chosen probability distribution. If $L(\theta_{new}(i+1)) < L(\theta_i)$, set $\theta_{i+1} = \theta_{new}(i+1)$. Else, take $\theta_{i+1} = \theta_i$.

Step 2

Stop if the maximum number of *L(θ)* evaluations has been reached or the user is otherwise satisfied with the current estimate for *θ* via appropriate stopping criteria; else, return to Step 1 with the new *i* set to the former *i+1.*

The simplest setting for conducting the random sampling of new (candidate) values of *θ* is when Θ *i*s a hypercube as well as using uniformly generated values of *θ*. The uniform distribution is continuous or discrete for the elements of *θ* depending on the definitions for these elements [21].

2.4.3 Robbins-Monro stochastic approximations

The Robbins-Monro stochastic approximation algorithm (RMSA) is a gradient based stochastic optimization algorithm for a wide variety of non linear problems. It was introduced in 1951 [22].The prototype stochastic application for RMSA algorithm is the problem of finding the root of the equation $g(\theta) = 0$ based on a noisy measurement of *g(θ)*, which is represented by *Y(θ).* The algorithm has the form

$$\theta_{k+1} = \theta_k - a_k Y(\theta_k) \quad , \qquad 2.16$$

where a_k is a non negative gain sequence that must satisfy certain conditions. θ_k represent the estimate of θ at the k^{th} iteration. Since the deterministic term, $\partial L / \partial \theta$ does not equal that of the stochastic term *Y,* the RMSA algorithm is fundamentally different from the steepest decent algorithm. However, there is an intuitive connection, since $E(Y) = \partial L / \partial \theta$ under standard RMSA conditions (typically under relatively modest regularity conditions justifying the interchange of a derivative and an integral or expectation).

A variation of the basic equation for in equation 2.16, is to include a projection operator, say Π_C, that maps solutions outside the constraint set C back to the nearest point within C. In such a case equation 2.16 becomes

$$\theta_{k+1} = \prod_C \left[\theta_k - a_k Y(\theta_k)\right] . \qquad 2.17$$

One of the richest aspects of RMSA is the convergence theory that has been developed over the years. Probably the most famous of these convergence conditions are those of the gain sequence a_k. The condition provides the careful balancing between the desired damping out of the noise effects as we get near to the solution θ* ($a_k \rightarrow 0$) and avoiding premature (false) convergence of the algorithm ($\sum_{k=0}^{\infty} a_k = \infty$ i.e. a_k is too large). The scaled harmonic sequence

(*a/(k+1)*, a>0) is the best-known example of gain sequence that satisfies the gain condition. Usually some numerical experimentation is required to choose the best value of the scale factor that appears in the gain sequence.

2.4.4 Steepest gradient decent

The steepest gradient descent method is the most straightforward method for incorporating gradient information in the minimization process. The minimum of the function is found by a number of consecutive one-dimensional line minimization steps using for instance Brent's algorithm [22], each step starting at the minimum found in the previous step and going in the direction of the gradient at that point, i.e. the direction of steepest descent. Steepest gradient descent is in general not a very good algorithm. As the gradient generally does not point to the optimum directly and because consecutive steps towards the optimum are necessarily at orthogonal angles, many tiny steps are usually required before reaching the optimum, especially when going down a long and narrow valley.

2.5 Feature-based methods

In feature-based image registration, the main aim is to find pair-wise correspondences between control points in reference and input images respectively, using their spatial relations or various descriptors of features. In this method, it is assumed that the two sets of features represented by control points have been detected. This method is typically applied when local structural information is more significant than information carried by image intensities. They allow for the registration of multimodal images. The main drawback of the feature-based method is that respective features might be difficult to find.

2.5.1 Spatial relations methods

Spatial relations methods are normally applied if the detected features are ambiguous or the neighbors are deformed. The spatial information about the distance between control points is normally exploited to correct the anomaly.
Stockman et al [49] proposed a method, which matches points connected by abstract edges. The assumed geometric model is the similarity transform. For every pair of control points from image pairs being registered, the parameters of the transformation which maps the points on each other are computed and represented as a point in the space of transform parameters. The parameters of the registration points that tend to match the highest number of features tend to cluster. On the other hand mismatches fill the parameter space randomly. The centroid of the cluster is used as the most probable vector for the registration parameters. The overall registration process is not affected by local errors.

Gosthaby et al also proposed a method based graph matching algorithm. In [50], they evaluated the number of features in the sensed image that, after the particular transformation, fall within a given range next to the features in the reference image. The transformation parameters with the highest score were then set as a valid estimate.
The distance values can be determined in two passes through the image feature array by a process known as “chamfering”. The feature array (F[i, j] i, j =1...N) is initially two valued : 0 for feature points and infinity otherwise. Two passes are made over the image, first “forward” from left to right and from top to bottom; and “backward” from bottom to top and from right to left. The forward pass modifies the features as follows :

```
FOR i =2 STEP 1 UNTIL N DO
        FOR j=2 STEP 1 UNTIL N DO
                F[i, j]  = minimum( F[i-1, j-1]+4, F[i-1, j-1]+3,
                        F[i-1, j+1]+4, F[i, j-1]+3, F[i, j]).
```

The backward pass is as follows:

```
FOR i =N-1 STEP 1 UNTIL 1 DO
    FOR j=N-1 STEP 1 UNTIL 1 DO
        F[i, j]  = minimum( F[i, j], F[i, j+1]+3,
            F[i+1, j-1]+4, F[i+1, j]+3, F[i+1, j+1]+4).
```

2.5.2 Invariant descriptors methods

Invariant descriptors exploit the correspondences between features that can be estimated using their description. It helps when descriptors are robust and invariant in the presence of image distortions. Descriptors should be unique, independent, robust and invariant. The simplest description is the image intensity function, limited to a close value of the feature [3]. However, direct measurement of image geometry becomes very complicated under prospective projection. One of the most popular transformations is the Fourier- descriptor transform.
Consider the *M* contour points of an image as a discrete complex function

$$f(m) = x(m) + jy(m) \qquad 2.18$$

f(m) can be transformed into the frequency, a(k), domain using the *discrete Fourier transform* (DFT). The result can be transformed into the spatial domain using the *Inverse discrete Fourier transform (IDFT)* without any loss.

$$a(k) = \frac{1}{M}\sum_{m=0}^{m=M-1} f(m)e^{-j2\pi km/M} \qquad \text{k = 0,…, M-1.} \quad 2.19$$

$$f(m) = \frac{1}{M}\sum_{k=0}^{k=M-1} a(k)e^{j2\pi km/M} \qquad \text{m =0…, M-1.} \quad 2.20$$

The coefficients *a(k)* represent the Fourier descriptors. They represent the contour in the Fourier domain. Certain geometric transformations can be related to simple transformations in the Fourier domain. Translation affects only the first Fourier descriptor. The other Fourier descriptors remain constant. Scaling the contour with *S* leads to scaling the Fourier descriptors with *S.* Rotating the contour by θ, yields a phase shift of θ in Fourier descriptors. Changing the contour points by m_o results in a linear phase shift of $2\pi m_0 k / M$.

The contour functions are made invariant against translation by setting the Fourier descriptor a(0) to zero. This moves the centroid of the contour to 0. Scaling invariance could be achieved by dividing the Fourier descriptors by the second Fourier descriptor: $a(k) = a(k)/|a(1)|$. To achieve rotation invariant descriptors that still represent the original shape of contour, the orientation of the basic ellipse which is described by the Fourier descriptor $a(1) = r_1 e^{j\Phi_1}$ $a(-1) = r_1 e^{-j\Phi_{-1}}$, is represented as

$$f(m)_e = a(-1)e^{-j2\pi m/M} + a(1)e^{j2\pi m/M} \quad . \qquad 2.21$$

With a few transformations and with the abbreviation $\overline{\Phi} = (\Phi_{-1} + \Phi_1)/2$,

We get $f(m)_e = e^{j(\Phi_{-1}+\Phi_1)/2}\left(r_{-1}e^{-j(\overline{\Phi}+2\pi m/M)} + r_1 e^{j(\overline{\Phi}+2\pi m/M)}\right)$ 2.22

which shows the rotation of the basic ellipse as $\Phi_e = (\Phi_{-1} + \Phi_1)/2$.

Using the orientation of the basic ellipse leads to an ambiguity of *π* radians. Therefore, the Fourier descriptors are only rotation invariant modulo by *π* radians. A large group of methods also use moment-based invariants for description of closed-boundary region features. Flusser et al. developed this approach in [57] by introducing combined blur rotation invariants. An example of this algorithm is shown in figure 2.4.

The correlation coefficient is often used to determine the correspondence between features [51]. Mutual information can also be used to improve the correspondence between features [52].

In [53], Ventura et al described image features by various descriptors such as, ellipticity, angle, and thinness, and represented relations among them by a multivalue logical tree. Then they compared the multivalue logical trees of the reference and sensed images to find the feature correspondence.

2.5.3 Relaxation methods

Relaxation methods involve the labelling of each feature from a sensed image with the label of a feature from the reference image, so it is consistent with the labelling given to the other feature pairs [54]. It is considered as a solution to the consistent labelling problem. The process of recalculating the pair figures of merit, considering the match quality of the feature pairs and of matching their neighbors, is iteratively repeated until a stable situation is reached. Backtracking can also be used a solution to the consistent labelling problem, where consistent labelling is generated in a recursive manner [54].

2.6 Estimation of the mapping function

The transformation model estimation involves choosing the mapping function that transforms the input image to overlay the reference image. The type of mapping function used depends on the assumed deformation of the input image, the method of acquisition of the input image, and the required accuracy of the image registration. Mapping function models can be broadly divided into global mapping functions and local mapping functions.

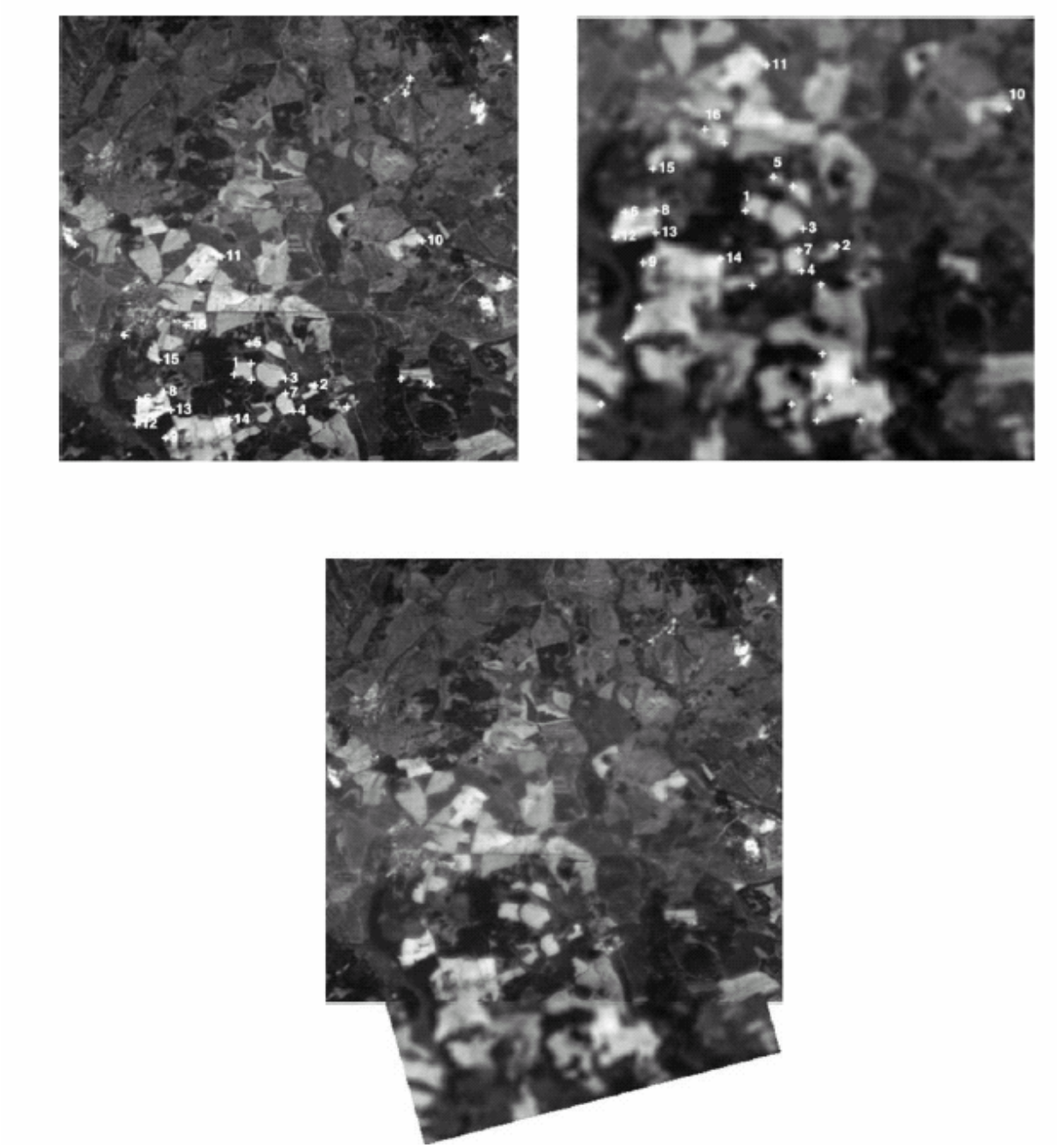


Figure 2.4 Image registration using feature-based method (invariant descriptors). In the image pairs above, control points are corners which were registered using complex moments [57]. The numbers show the control points (corners).The result is displayed in the bottom image source [2]

2.6.1 Global mapping functions

Global models use all control points to estimate the mapping function parameters valid for the entire image. The general mapping models are chosen as polynomial functions of first, second and third order. A second order polynomial can be written as

$$u = a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 \qquad 2.23$$

$$v = b_0 + b_1 x + b_2 y + b_3 xy + b_4 x^2 + b_5 y^2 \qquad 2.24$$

where a_i and b_i are polynomial coefficients. A simpler model consists of scaling, rotation, and translation.

$$u = s(x\cos\theta - y\sin\theta) + T_x \qquad 2.25$$

$$v = s(x\cos\theta + y\sin\theta) + T_y \qquad 2.26$$

The coefficients or transformation parameters can be estimated using least squares estimation. Assuming the coordinates of corresponding points in the images after correction for scale are: {$\mathbf{p}i = (x_i, y_i)$, $\mathbf{P}_i = (u_i, v_i)$: i =1, …, N}, the relations between corresponding control points in the images can be written as

$$\boldsymbol{P}_i = \boldsymbol{R}\boldsymbol{p}_i + \boldsymbol{T}, \quad i = 1, \ldots, N. \qquad 2.27$$

The Vector **T** and the matrix **R** are determined by minimizing

$$E^2 = \sum_{i=1}^{N} \left\| \mathbf{P}_i - \left(\mathbf{R}\mathbf{p}_i + \mathbf{T}\right) \right\|^2 . \qquad 2.28$$

Thus the polynomial coefficients are also determined. In general, the number of control points is higher than the minimum number required in the determination the mapping function. This approach has proved to be very effective and accurate for satellite images.

2.6.2 Local mapping models

Local mapping functions treat the image as a composition of patches and the function parameters depend on the location of their support in the image. This leads to defining mapping functions for each patch separately. Unlike global mapping model functions, local mapping model functions can handle locally deformed images. Least squares techniques average out local geometric distortions over the entire image which might not be wanted. In [56], Goshtasby showed the superiority of locally sensitive image registration methods. The algorithm is as follows:

1. Triangulate the control points in reference image
2. Form the corresponding triangles with control points in input image and reference image.
3. Map triangular regions in sensed image one by one to corresponding regions in the sensed image using linear or cubic functions.

2.6.3 Using Radial Basis functions for mapping

Radial basis functions were developed for the interpolation of irregular images. It is an important property of the function that its value at each point depends just on the distance of the point from the CPs, not on its particular position. The mapping function has a form of a linear combination of translated radially symmetric function plus a low-degree polynomial

$$v = \mathrm{b}_0 + b_1 x + b_2 y + \sum_{i=0}^{N} c_i g(\mathbf{x}\mathbf{x}_i), \qquad 2.29$$

where the radial basis function has the form

$$g(\mathbf{x}\mathbf{x}_i) = \left\|\mathbf{x} - \mathbf{x_i}\right\|^2 \ln(\left\|\mathbf{x} - \mathbf{x_i}\right\|) \qquad 2.30$$

This form is called a thin-plate spline (TPS). The form corresponds to the 1st order of the general TPS form. $\ln(\|\mathbf{x} - \mathbf{x}_i\|)$ is the natural logarithm of the Euclidian norm. (see [60] for prospective mathematical background). TPS registration gives good results but the computations can be very time consuming, namely if the number of CPs is high. Considerable attention has been paid to the methods decreasing the complexity of the TPS evaluation while preserving reasonable accuracy.

2.7 Image resampling

Image resampling is a process that involves the extraction and interpolation of grey levels from pixel locations in the original input image and thus to register images. Image interpolation methods are as old as image processing. In the early years, simple algorithms, such as nearest neighbor or linear interpolation, were used for resampling. As a result of information theory introduced by Shannon in the late 1940's, the sinc function was accepted as the interpolation function of choice. However, this ideal interpolator has an infinite impulse response (IIR) and is not suitable for local interpolation with finite impulse response (FIR). From the mathematical point of view, Taylor or Lagrange polynomials have been suggested to approximate the sinc function [58]. This is documented in most textbooks on numerical analysis. Thereafter, due to their numerical efficiency, different families of spline functions have been used instead. B-splines sometimes are referred to as cubic splines, while cubic interpolation is also known as cubic convolution, and bi-cubic spline interpolation. For image resampling, the interpolation step must reconstruct a two-dimensional continuous signal f(x,y) from its discrete signal samples f(k,l). Thus the amplitude at the (x,y) must be estimated from its discrete neighbors. This can be described formally as the convolution of discrete image samples with a two dimensional impulse response *h(x ,y)* of a two dimensional reconstruction filter

$$f(x,y) = \sum_k \sum_l f(k,l) \cdot h(x-k, y-l)\,. \qquad 2.31$$

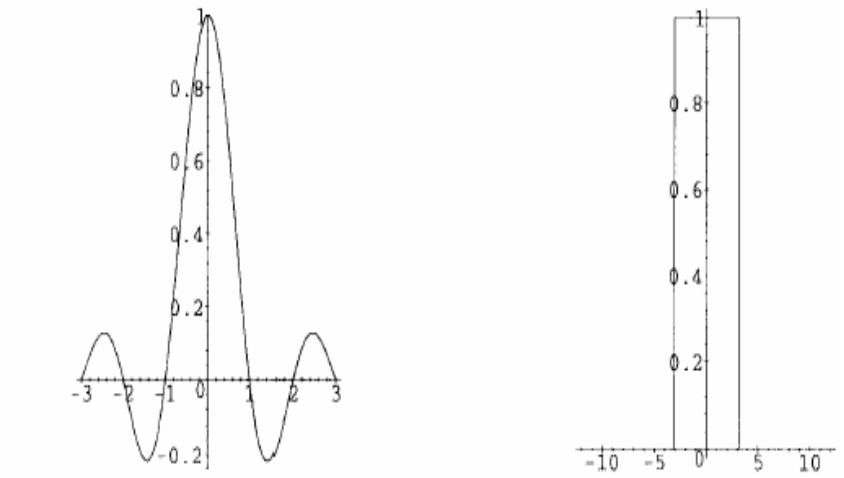
Usually, symmetrical and separable interpolation kernels are used to reduce the computational complexity,

$$h(x,y) = h(x) \cdot h(y)\,. \qquad 2.32$$

2.7.1 Ideal interpolation

Sampling theory explains that, the scanning of a continuous image f(x,y) yields infinite repetitions of its continuous spectrum F(u,v) in the Fourier domain, which do not overlap if the Nyquist criterion is satisfied. If and only if this is so, can the original image *f(x,y)* be perfectly reconstructed from its samples *f(k,l)* by the multiplication of its spectrum with of a rectangular function in the Fourier domain. The 1-D ideal interpolation equals the multiplication with the rect function in the Fourier domain and can be realized in the spatial domain by a convolution with the sinc function.

$$h(x)_{ideal} = \frac{\sin(\pi x)}{\pi x} \qquad 2.33$$



(a) (b)

Figure 2.5 Ideal interpolation (a) kernel is plotted for -3<x<3 ,(b) Magnitude of Fourier transform

Figure 2.5(a) above shows the ideal IIR-interpolation $h(x)_{ideal.}$

The plot was truncated within the interval -3>x>3. The magnitude $|H(f)_{ideal}|$ of the Fourier transform $H(f)_{ideal}$ of infinite kernel $h(x)_{ideal}$ is plotted within the interval $-4\pi \le \omega = 2\pi f \le 4\pi$ is shown in figure 2.5(b). The interval $-\pi < \omega < \pi$ is called the passband and f=1/2 Nyquist frequency.

Some fundamental properties of any interpolator can be derived from this ideal interpolation function. $h(x)_{ideal}$ is positive from zero to one, negative from one to two, positive from two to three, and so on. For $H(0) \equiv 1$ these zero crossings guarantee that the image is not modified if it is resampled on the same grid. Therefore, the kernels satisfying $h(0) \equiv 1 \quad h(x) \equiv 0$, |x| = 1,2,….

avoid smoothing and preserve high frequencies. They are called interpolators.

2.7.2 Nearest neighbour

The easiest way to approximate the sinc function by a spatially limited kernel is given by the nearest neighbour method. The value f(x) at the location is chosen as the next known value f(k). Therefore, only N=1 finite kernel's supporting point is required for the nearest neighbor interpolation. This is equal to convolution with a rect function.

$$h(x) = rect(x) = \begin{cases} 1, if \left|x\right| \le 0.5 \\ 0, elsewhere \end{cases} \qquad 2.34$$

Nearest neighbour interpolation determines the grey level from the closest pixel to the specified input coordinates, and assigns that value to the output coordinates. This method is considered the most efficient in terms of computation time. Because it does not alter the grey level value, a nearest neighbour interpolation is preferred if subtle variations in the grey levels need to be retained, if classification will follow the registration, or if a classified image is to be resampled. Nearest neighbour interpolation introduces a small error into the newly registered image. The image may be offset spatially by up to 1/2 a pixel, causing a jagged or blocky appearance if there is much rotation or scale change.

2.7.3 Bilinear interpolation

For separated bi-linear interpolation, the values of both direct neighbors are weighted by their distance to the opposite point of interpolation. Therefore, the linear approximation of the sinc function follows the triangular function

$$h(x) = \Lambda(x) = \begin{cases} 1 - |x|, if\ |x| \le 1 \\ 0, elsewhere \end{cases}. \qquad 2.35$$

The linear kernel is also a DC constant interpolator. Bilinear interpolation determines the grey level from the weighted average of the four closest pixels to the specified input coordinates, and assigns that value to the output coordinates. This method generates an image of smoother appearance than nearest neighbour, but the grey level values are altered in the process, resulting in blurring or loss of image resolution. Because of these changes in the grey level values, any image classification processes should be performed before the interpolation. Bilinear interpolation requires three to four times the computation time of the nearest neighbour method.

2.7.4 Cubic interpolation

Cubic polynomials can be used to approximate the sinc function. In case of cubic interpolation with two points, a symmetric kernel can be defined as

$$h(x) = \begin{cases} A|x|^3 + B|x|^2 + C|x| + D, if\ |x| \le 1 \\ 0, elsewhere \end{cases}. \qquad 2.36$$

The parameters A, B, C, and D can be determined by applying the boundary conditions h(k) =1 for k=0, h(k)=0 for k≠0, and continuity boundary conditions. For N=2 (N is the interpolating kernel size), the boundary conditions yield four equations and four parameters resulting in;

$$h(x) = \begin{cases} 2|x|^3 - 3|x|^2 + 1, if\,|x| \le 1 \\ 0, elsewhere \end{cases} \quad . \qquad 2.37$$

By definition, this is a DC-constant interpolator.

Cubic convolution determines the grey level from the weighted average of the 16 closest pixels to the specified input coordinates, and assigns that value to the output coordinates. This method is closer to the perfect sin(x)/x resampler than nearest neighbour or bilinear interpolation. The image is slightly sharper than that produced by bilinear interpolation, and it does not have the disjointed appearance produced by nearest neighbour interpolation. Because the grey level values are altered by this method, any image classification processes should be performed before the interpolation. Cubic convolution requires about 10 times the computation time required by the nearest neighbour method.

(a)

(b)

(c)

(d)

Figure 2.9 Interpolation techniques: The original image (a), was enlarged 6 times using three different interpolation methods – (b) nearest neighbor, (c) sinc interpolation and, (d) bilinear interpolation

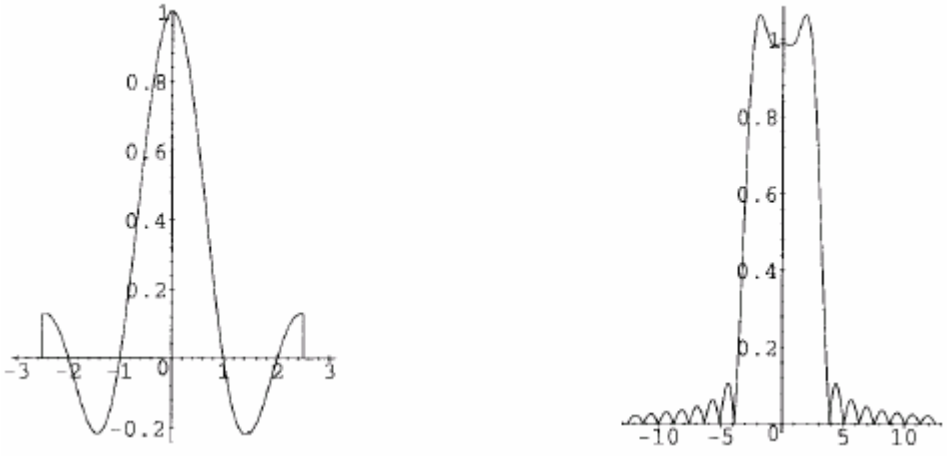
2.7.5 Truncated sinc interpolation

Although the sinc function provides an exact reconstruction of *f(x,y),* it is spatially unlimited. Two common approaches for overcoming this drawback, include truncation and windowing with the window function w(x).

$$h(x) = \begin{cases} h_{ideal}(x) \cdot w(x), 0 \le |x| \prec N/2 \\ 0, elsewhere \end{cases}, \qquad 2.38$$

where w(x)=w(-x), w(x)=w(x)·rect(x), w(x)=1, and *N* is the number of finite kernel's supporting points. All windowed or truncated sinc kernels are real interpolators.

Truncation is equivalent to the multiplication of with a rectangular function in the spatial domain, which is tantamount to a convolution with a sinc function in the frequency domain. Therefore, truncations of the ideal interpolator produce ringing effects in the frequency domain because a considerable amount of energy is discarded.



(a) (b)

Figure 2.9 Truncated sinc interpolation (a) Kernel (b) Magnitude of Fourier Transform

Sinc interpolation 8pt determines the grey level from the weighted average of the 64 closest pixels to the specified input coordinates and assigns the value to the output coordinates. Sinc interpolation 16pt does the same, using the 256 closest pixels. The image is sharper than that produced by bilinear interpolation, and it does not have the disjointed appearance produced by nearest neighbour interpolation. Because the grey level values are altered by this method, any image classification processes should be performed before the interpolation. Sin(x)/x with an 8 x 8 window require about 20 to 40 times the computation time required by the nearest neighbour method. Sin(x)/x with a 16 x 16 window requires 40 to 80 times the computation time required by the nearest neighbor method.

2.8 Registration quality estimation

Estimation of accuracy of registration algorithms is a substantial part of registration process. Without quantitative evaluation, no registration method can be accepted for practical utilization. A lot of work has been done on validation of rigid-body registration while validation of non-linear, local and elastic registration methods is still at the beginning [2].

The evaluation of the quality of image registration is a non trivial problem. In general, there is no true solution available. This is partially because the errors can be dragged into the registration process in each of the stages and also it is hard to distinguish between the registration inaccuracies and actual physical differences in the image contents. Some of the methods used in measuring registration accuracy include root mean square error, matching error, localization error and alignment error.

2.8.1 Root Mean Square Error

The root mean square error (RMS) is calculated in pixels. It is defined as

$$RMS = \sqrt{\left(\frac{1}{N}\left(\sum_i \sum_j \left(a_{ij} - b_{ij}\right)^2\right)\right)} \qquad 2.39$$

where a_{ij} is the pixels values of the reference image, a, and b_{ij} is the pixel values of image with estimated transformation , b, and N is the total number of pixels in image. In order to compare the relation between RMS and correlation, let the mean square error be defined as

$$MSE = \sum (a-b)^2 = \sum a^2 - 2\sum ab + \sum b^2 \,, \qquad 2.40$$

Opposed to normalized correlation

$$CC = \frac{\sum a \cdot b}{\left(\sum a^2 \sum b^2\right)^{1/2}} \,. \qquad 2.41$$

Over a large sample size, it is safe to assume that $v = \sum a^2$ and $u = \sum b^2$ are constant. Under this assumption,

$$MSE \approx c_1 - 2\sum a \cdot b \,, \qquad 2.42$$

$$CC \approx c_2 . \sum a \cdot b \qquad 2.43$$

where $c_1 = u + v$ and $c_2 = (uv)^{-1/2}$, represent constant values. The minimization of mean square error is equivalent to the maximization of correlation coefficient. Furthermore, if the c_1=c_2 =c, the following linear dependence is found between correlation and square error.

$$CC = 1 - \frac{MSE}{2c} \,. \qquad 2.44$$

2.8.2 Matching error

Matching error is measured by the number of correct or false matches when matching the correspondence between known transformations and the estimated

transformation. This can used to measure the robustness of image registration algorithms. Correct match can be identified by consistency check, where the different registration algorithms are applied to the same set of transformations. Higher number of correct matches indicates high robustness. Instead of known transformations, control points (CP) can also be used to measure matching error [2].

2.8.3 Localization error

The localization error is the displacement due to control point coordinates due to their inaccurate detection. It cannot be measured directly on a given image because it is an intrinsic error of a detection method. However, the mean precision of most control point's detection methods is known for various image types from computer simulation studies and ground truth comparisons. Localization error can be reduced by selecting efficient feature detection algorithms for a given data. There is often a trade-off between number of selected candidates and the mean localization error.

2.8.4 Alignment error.

Alignment error is defined as the difference between the mapping model used for the registration and the actual between-image distortions. It can be estimated using the consistency check using multiple cues or approaches. Here the image registered by the method under investigation is compared with the same image by another comparative method. The gold standard method [23] is used in medical imaging as the comparative method. In applications areas where gold standard does not exist, like in remote sensing and industrial inspection, any method of different nature can be taken as a comparative method. A small difference between the different methods indicates high registration accuracy.
Alignment error can be evaluated in several ways. The simplest measure is a mean square error at the control points (CPE). Although commonly used, it is not

good alignment error measure. In fact, it only quantifies how well the CP (Control points) coordinates can be fitted by the chosen mapping model. For any CP set, zero CPE can be reached just by selection of a mapping model with enough degrees of freedom (this well known phenomenon is in numerical analysis called ‘over-fitting’). On the other hand, large CPE can be caused by CP localization errors and does not necessarily reflect poor registration accuracy.

Finally, the oldest method of registration accuracy estimation—visual assessment by a domain expert—should be mentioned. It is still in use at least as a complement of the above mentioned objective error measures.

CHAPTER 3

AUTOMATIC SUBIMAGE MATCHING

Subimage matching involves the process of identifying unique templates in an image and comparing these templates with windows of the same size in another image and identifying the windows that are most similar to the template. The process can be manual or automatic. Once the corresponding templates are found, their centres are used as control points to determine the registration parameters. To achieve highly accurate matches, the templates selected should be unique. Some of the features used to determine uniqueness include high entropy, high variance (high bandwidth) and high gradient magnitude [24]. Using subimages for image registration is not only computationally efficient but can produce more reliable results. In this chapter, a measure is proposed, called alignability, which shows the ability of a subimage to produce reliable and accurate results. A comparison with other selection methods such as using entropy, gradient and variance is also presented. Alignabilty measures the effectiveness of a subimage for registration by registering the subimage with itself using mutual information. Using the alignablity measure shows more reliable and accurate results than using entropy, gradient, and variance. The efficiency and reliability using mutual information, correlation coefficient and absolute differences as similarity metrics in template matching is also presented.

3.1 Advantages of template matching

Using the entire image for registration does not necessarily produce optimal results when the images have partial intensities and/or geometric differences. For images with partial differences using templates should produce more accurate results than using the entire image. When using the template matching method, we would like to select information from areas in image pairs that are similar and

avoid from selecting from dissimilar areas. Some of the methods used to discard outliers are histogram matching and standard deviation. If the centres of two images correspond to each other and have small rotational differences, the average distance between corresponding pixels in the images will be larger than the average distance between corresponding pixels in two subimages of the image when the centres of the subimages coincide. Thus, matching subimages is less sensitive to small rotational differences.

Another main reason for using subimages is the large reduction of computations in image registration:

Let *N x N* be the size of the original image and *N/i x N/i the* size of subimages where i^2 is the number of subimages. If the computational complexity for mutual information is *O (N) + O* (K^2) where *K* is number of histogram entries, since O (K^2) is almost a constant then, the speed up using a subimage for registration is

$$S = N^2 / \frac{N^2}{i^2} = i^2 \qquad 3.1$$

3.2 Automatic subimage selection

For reliable and accurate subimage registration, subimages selected should be highly detailed and unique. Normally a small percentage of unique subimages are selected in an ordered list. The unique subimages are used for the registration of the image pairs.

In this section, automatic subimage selection methods are presented. The alignalibilty based method is also discussed.

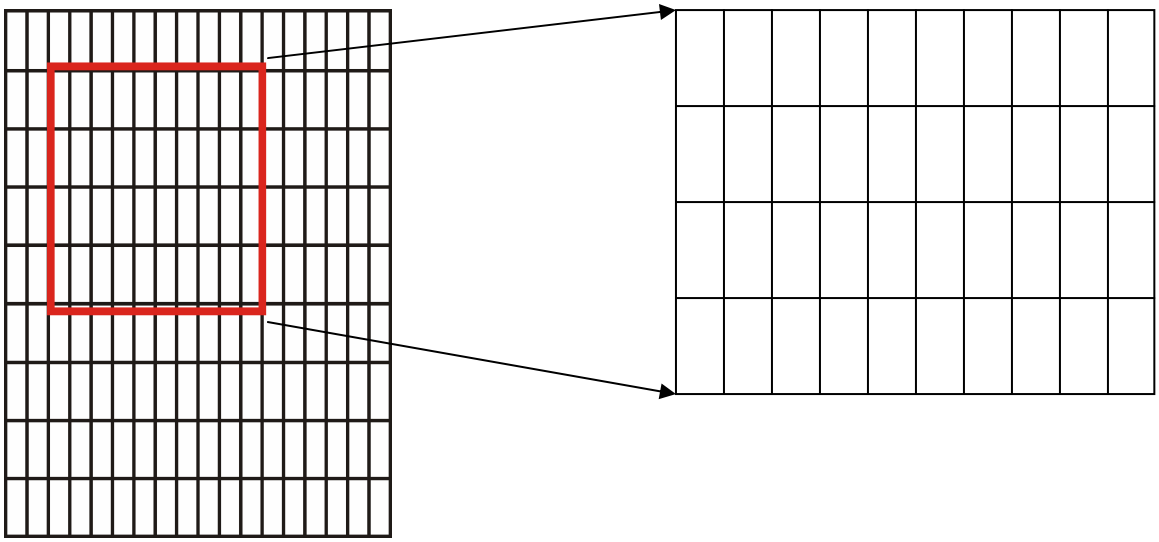


Figure 3.1 Template or subimage selected from image

3.2.1 Gradient-based selection method

A method of determining highly detailed subimage is to search for high-gradient edges. Gradient magnitude can be used to determine the amount of edges in an image [25]. An edge is a jump in intensity.

The gradient of an in image, f*(x, y)* at a location (x, y) is defined by

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}. \qquad 3.2$$

The gradient magnitude is also shown in the equation below. For the modulus we have then:

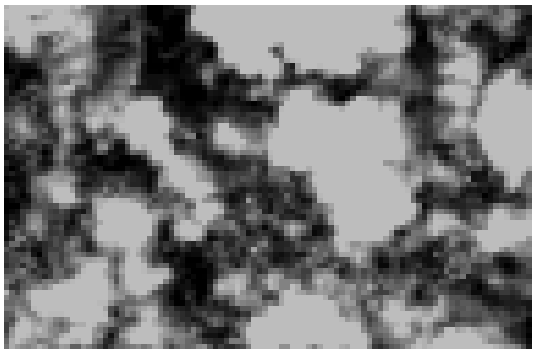
$$|\nabla f| = \left[G_x^2 + G_y^2\right]^{1/2}. \qquad 3.3$$

The sum of gradient magnitude is computed for each subimage, as a measure of edge content. The subimages are selected based on accumulated gradient magnitudes. The Sobel operator was used as our gradient operator in our experiment. The filter kernels in figure 3.2 (i) and figure 3.2 (ii) are used to determine the row gradient and column gradient respectively.

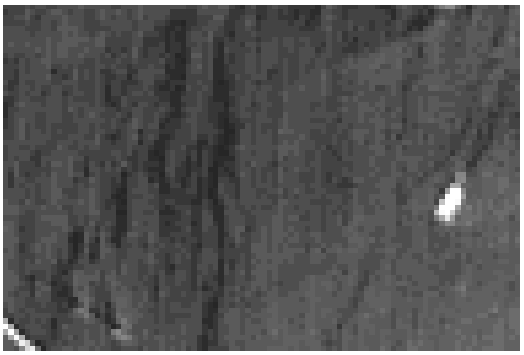
$$\frac{1}{4}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix} \qquad \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 2 \\ -1 & 2 & 1 \end{bmatrix}$$

(i) (ii)

Figure 3.2 Kernels constituting the Sobel operator



(i) Template from figure 2.2 with gradient magnitude of 4136



(ii)Template from figure 2.2 with gradient magnitude of 564

Figure 3.3 The subimages with highest gradient magnitude (i) and lowest gradient magnitude (ii) of image in figure 2.2

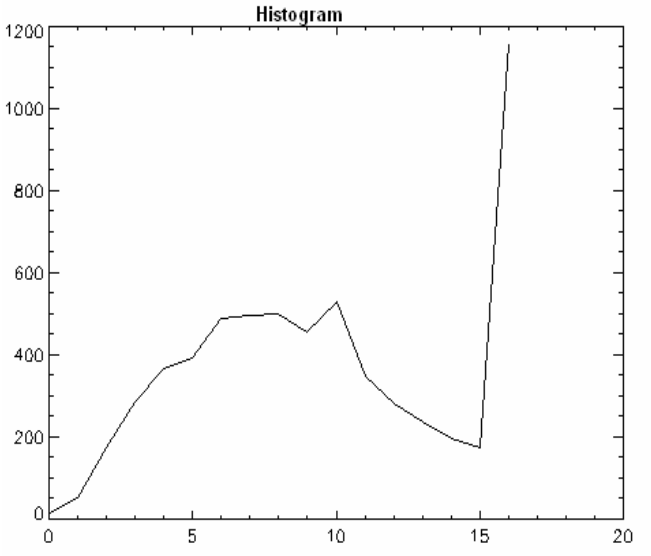
3.2.2 Entropy-based selection method

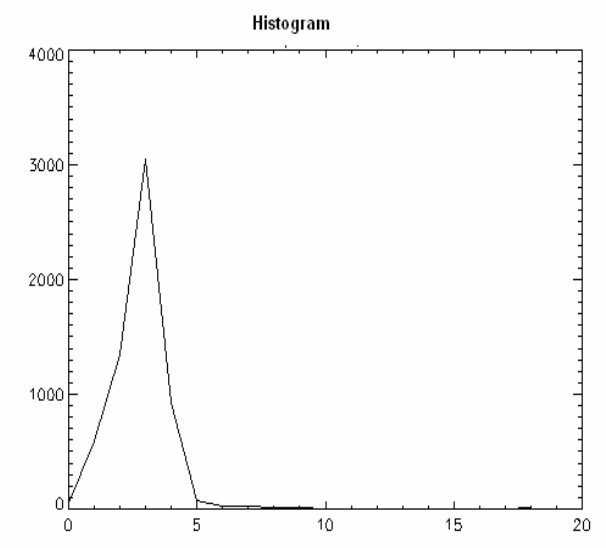
The entropy of a subimage is the information content in a subimage. There are different approaches to estimate entropy. We used Shannon's entropy definition

for our experiment [26]. If S_i is the *ith* pixel in a subimage, and P(S_i) is the probability of S_i, then entropy is defined as

$$H(S) = -\sum_{i=1}^{N} P(S_i)\log(P(S_i)) \qquad 3.4$$

where *N* is total number of pixels in the subimage.





Histogram for the template in figure 3.3 (i) (i)

Histogram for template in figure 3.3(ii) (ii)

Figure 3.4 Histograms for the subimage with high gradient in figure 3.3(i) and low gradient in figure 3.3(ii)

A submage with a sharp probability density distribution correspond to low entropy whereas a dispersed distribution yields a high entropy value. Using the same images in figure 3.3, the entropy of image 3.3 (i) and image 3.3 (ii) are 3.7 and 1.7 respectively.

3.2.3 Variance-based selection method

The variance of a subimage which is equivalent to its bandwidth is defined by

$$V = \frac{1}{N}\sum_{i=0}^{N-1}\left(x_i - \bar{x}\right)^2 \quad , \qquad 3.5$$

where x_i are the pixel values and $\bar{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i$, is the sample mean or the mean of pixel values of the subimage. Experiments show that a high variance of a subimage enhances accurate registration. Using the same images in figure 3.3, the variance of image 3.3 (i) and image 3.3 (ii) are 6060 and 188, respectively.

3.2.4 Alignability-based selection method

Alignability is the ability of an image to provide reliable image registration results by showing the correct transformations. Alignability (Ay) is computed by mutual information values of a subimage against the transformed version of itself. A set of mutual information values, MI_{S_i} are obtained by comparing the subimage $S_i(0)$ by a rotated version of itself, $S_i(\theta_i)$,

$$MI_{S_i} = MI(S_i(0), S_i(\theta_j)) \qquad 3.6$$

where θ_j is the angle of rotation. In our work, θ_j is varied between -10° to 10° with an increment of 1°.

The scaled difference of highest and the second highest MI values is defined as alignability. This can be expressed as

$$Ay = MI_1 - MI_2 \qquad 3.7$$

where MI_1 and MI_2 are the highest and second highest mutual information values respectively. Subimages with weak alignabilty are ignored. Although we used rotation to estimate alignabilty, the results show that subimages with good alignabilty are robust to translational transformation. Using the same images in figure 3.3, the alignability for figure 3.3 (i) and figure 3.3 (ii) are 1.6 and 2.6 respectively.

Even though figure 3.3 (i) has a higher gradient magnitude, entropy and variance than figure 3.3 (ii), it has lower alignability than figure 3.3(i). The desired qualities for image pairs being aligned are high entropy, high gradient magnitude, high variance and high alignability.

3.2.5 Avoiding outliers

Using outliers in subimage matching can affect reliable registration results. In remote sensing this might occur as a result of atmospheric changes or noise. Thus, to refine automatic selection of subimage, statistical methods are applied to ignore subimages which are anomalies. The methods employed include the following:

Standard Deviation (S_d), which is expressed as

$$S_d = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\left(x_i - \overline{x}\right)^2} \qquad 3.8$$

where x_i and $\bar{x}$ are the pixel values and mean of pixel values respectively and N is the number of pixels in both subimages.

The matched sample t-test is defined as [27],

$$t = \frac{M}{\frac{S_d}{\sqrt{N}}} \qquad 3.9$$

where M is the mean differences between histograms of subimages ($M = \frac{1}{N}\sum_{i=0}^{N-1} x_i - \frac{1}{N}\sum_{i}^{N} p_i$ where x_i and p_i are the subimage pixels). S_d is the standard deviation of the difference. The top part of the ratio is just the difference between the two means or averages. The bottom part is a measure of the variability or dispersion of the scores. This formula is essentially another example of the signal-to-noise metaphor in research: the difference between the means is the signal that, in this case, the bottom part of the formula is a measure of variability of subimages.

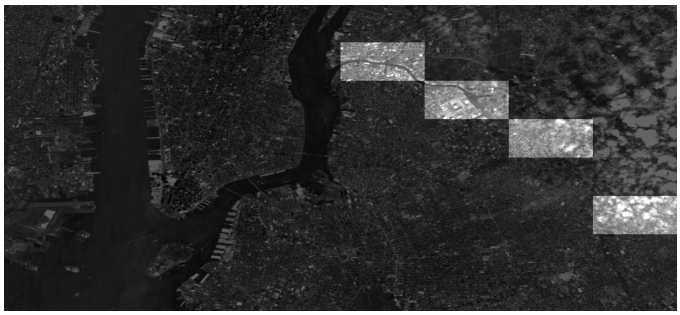
3.3 Experiment and results

The reference images used in the experiment are of size 768 x 512. Gaussian noise is added to image to input (slave) images with a SNR of 20 dB.The noise is added to mimic real circumstances. Sensor noise can make reference and input image vary in random fashion. Without noise, subimages selected will give a perfect match with their corresponding subimages. The input images are translated between -2 to 2 pixels for both x-direction and y-direction. The images are divided into 64 subimages of size 96 x 64. All 64 subimages are registered by maximizing MI. The alignability, entropy and gradient magnitude of each subimage reference images are computed and sorted in ascending order. The top 4 subimages out of the 64 subimages are selected. The equivalent percentage of total number of correctly registered subimages for translation in x-direction (Tx), y-direction (Ty), and rotation angle (θ) are and their root mean square error are computed for each method. The overall percentage of correctly registered subimages is the total number of subimages among the 64 subimages which are correctly registered divided by 64 and expressed as a percentage.

SAR 1

(i) Subimages selected using entropy

(ii)Subimages selected using gradient



(iii) Subimages selected using alignalbilty

(iv)Subimages selected using variance

Figure 3.4 Top 4 subimages of each feature selected for SAR1

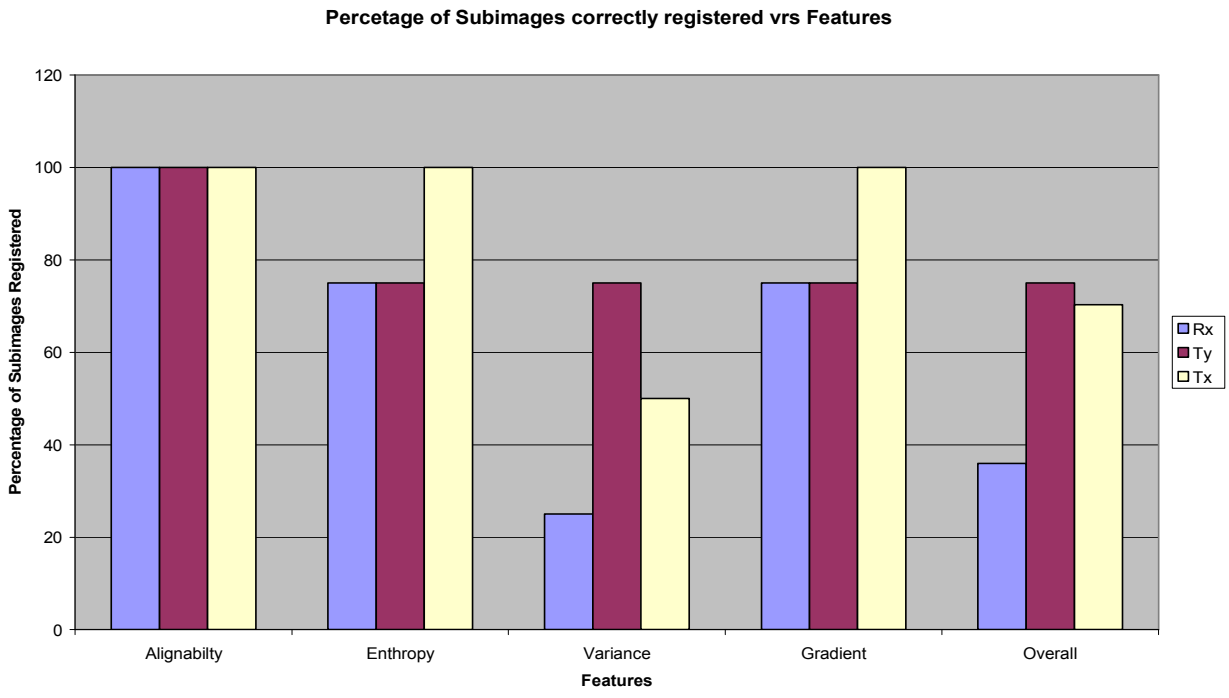


Figure 3.5 Percentage of correctly registered subimages versus different selection methods for SAR1

Lena

Subimages selected using entropy

Subimages selected using gradient

Subimages selected using alignalbility

Subimages selected using variance

Figure 3.6 Top 4 subimages of each feature selected for Lena

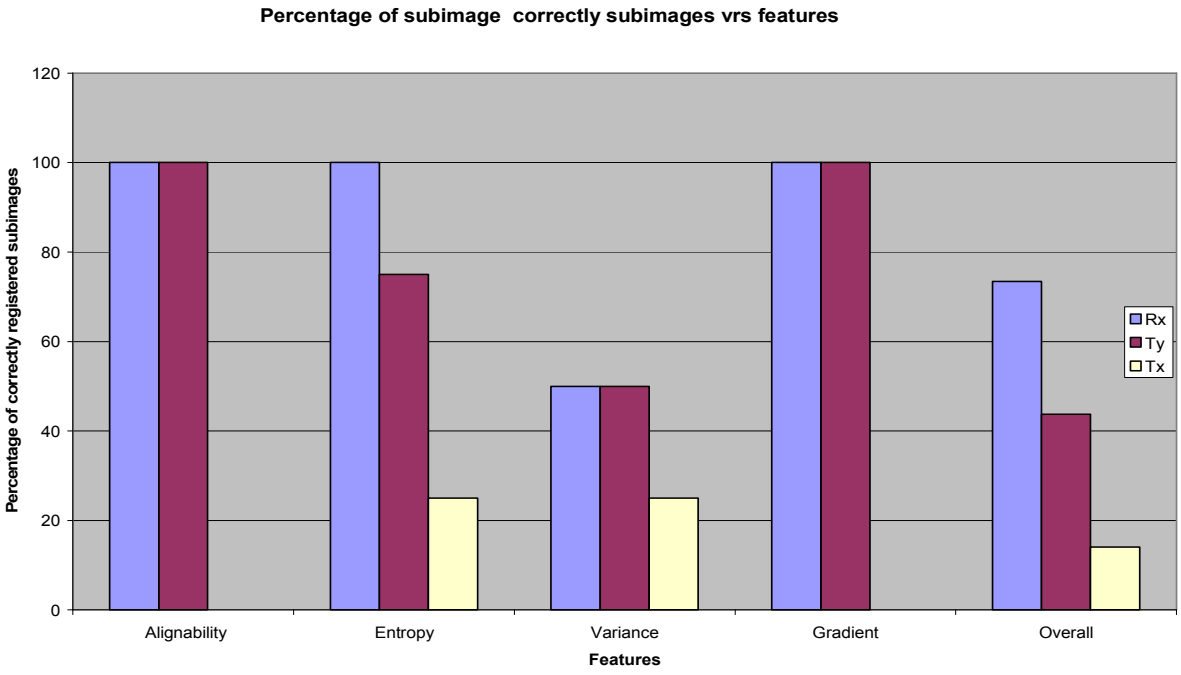
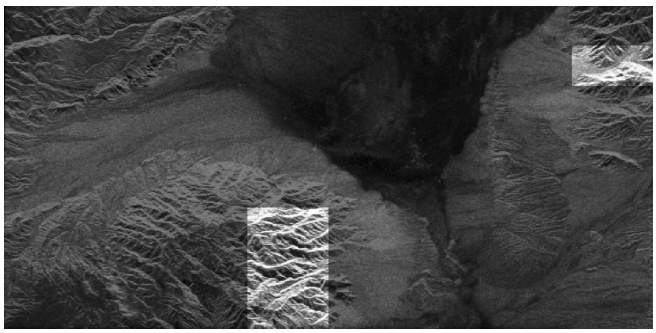
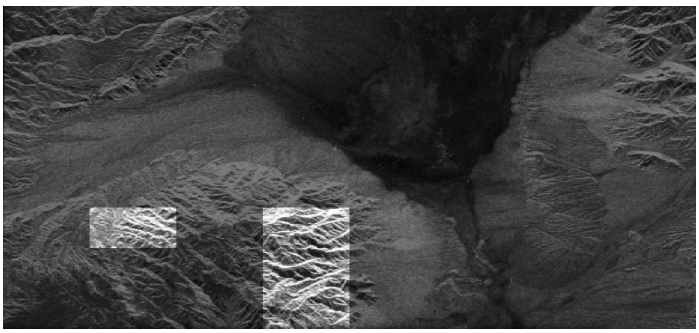


Figure 3.7 Percentage of correctly registered subimages versus different selection methods for Lena. None of the subimages selected for gradient or alignability is correctly registered in the x-direction (Tx)

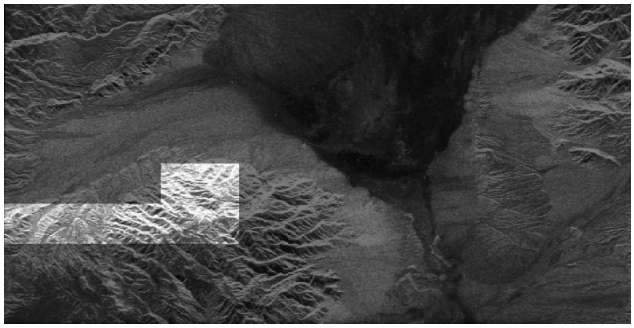
SAR2



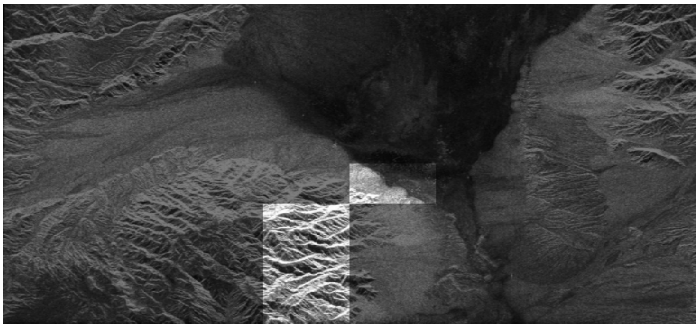
Subimages selected using entropy



Subimages selected using gradient



Subimages selected using alignability



Subimages selected using variance

Figure 3.8 Top 4 subimages of each feature selected for SAR2

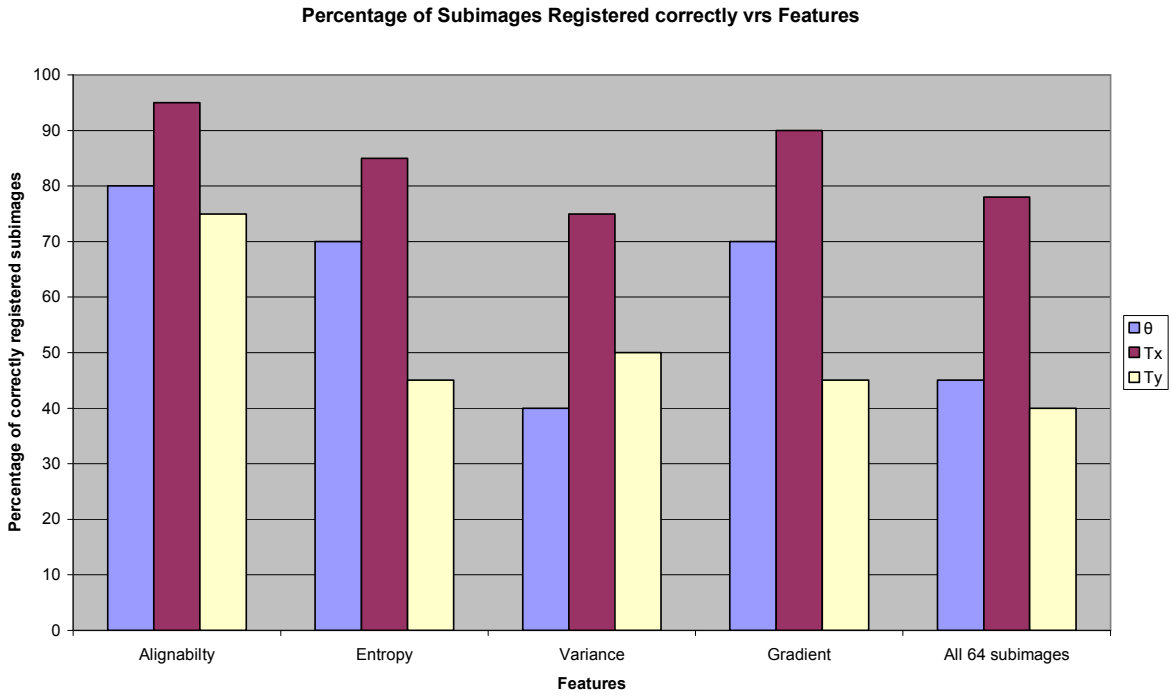
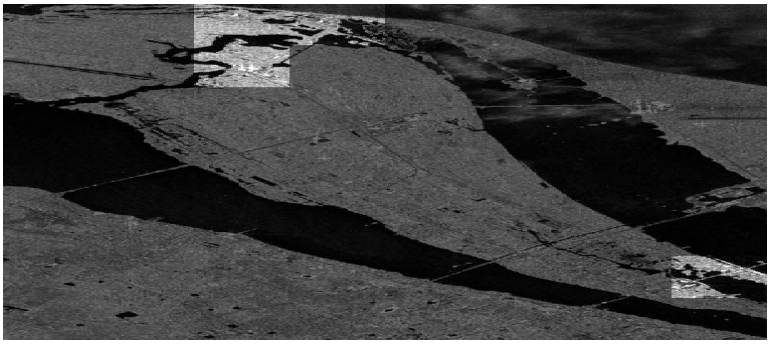


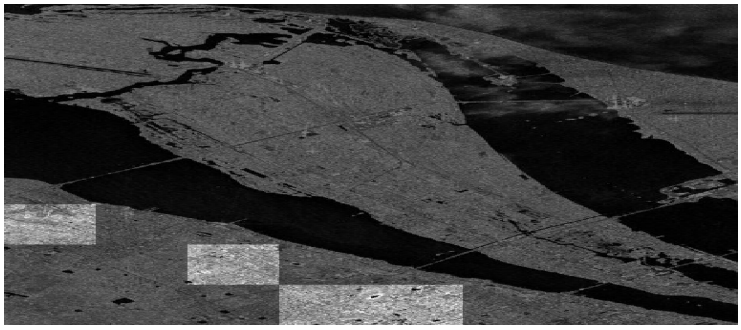
Figure 3.9 Percentage of correctly registered subimages versus different selection methods for SAR 2

SAR3

Subimages selected using entropy



Subimages selected using gradient



Subimages selected using alignabilty

Subimages selected using variance

Figure 3.10 Top 4 subimages of each feature selected for SAR3

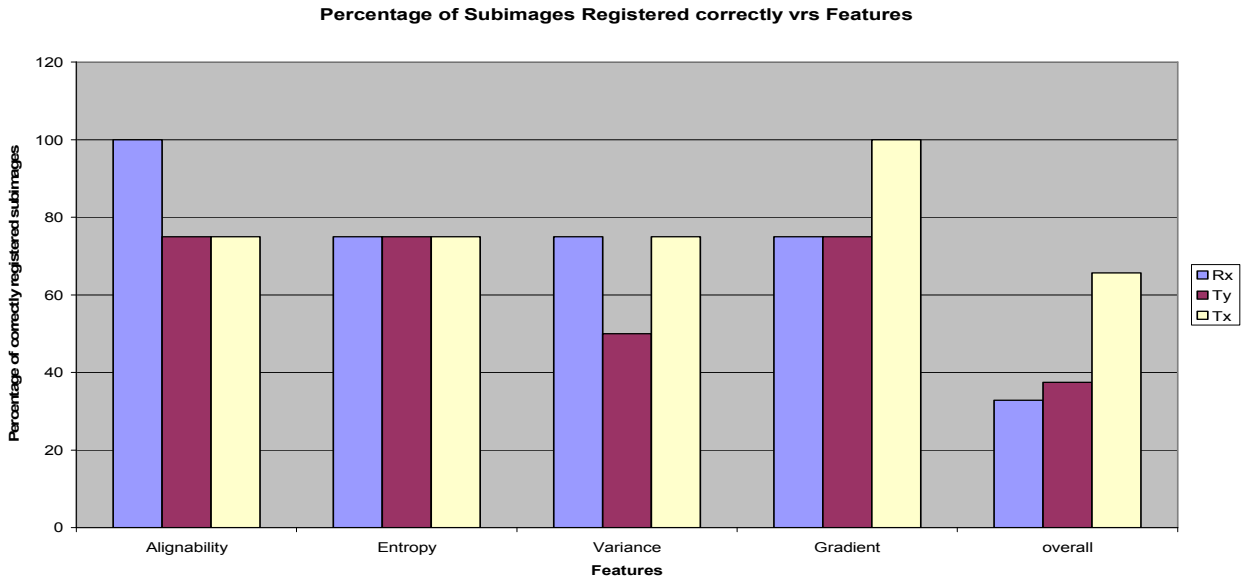
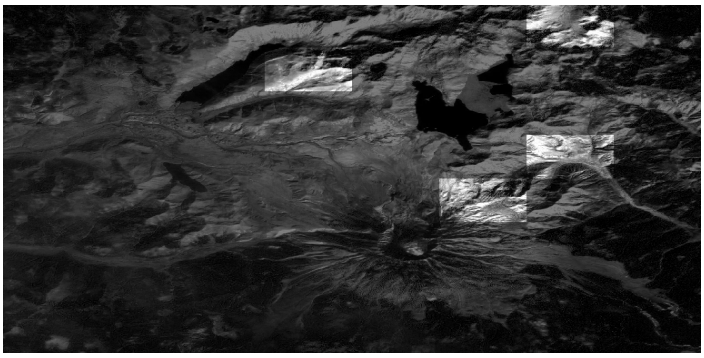


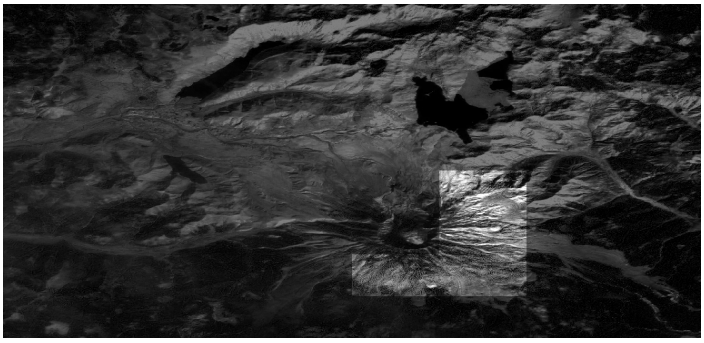
Figure 3.11 Percentage of correctly registered subimages versus different selection methods for SAR3

SAR4

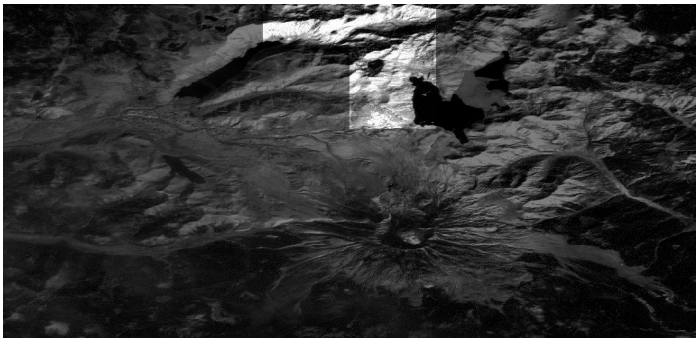


Subimages selected using entropy

Subimages selected using gradient



Subimages selected using alignabilty



Subimages selected using variance

Figure 3.12 Top 4 subimages of each feature selected for SAR4

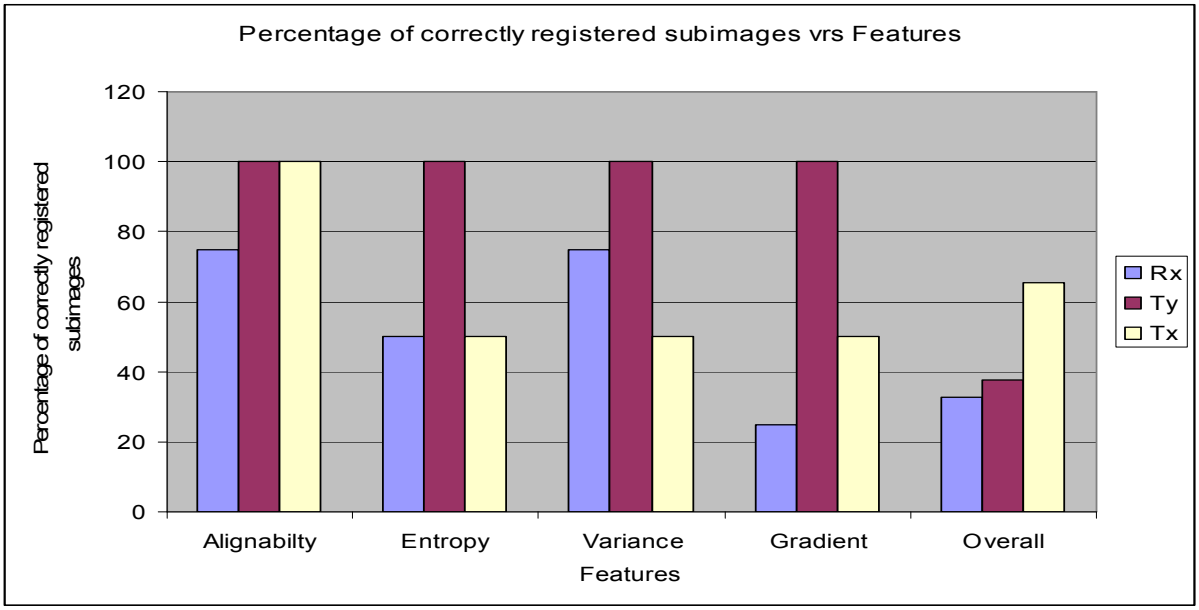
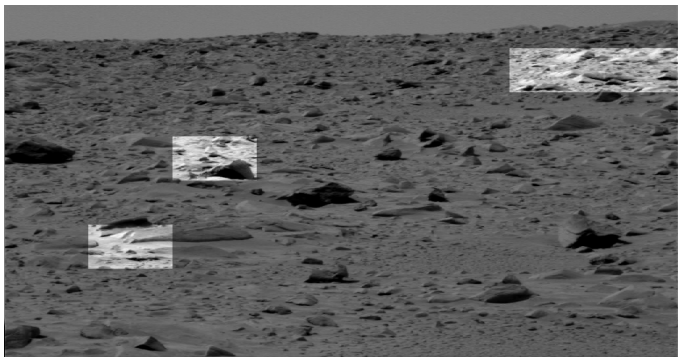
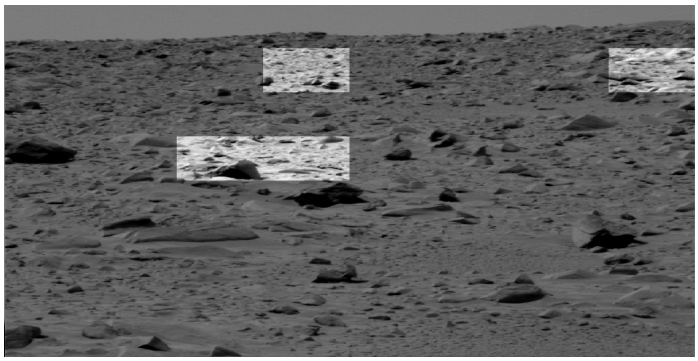


Figure 3.13 Percentage of correctly registered subimages versus different selection methods for SAR4

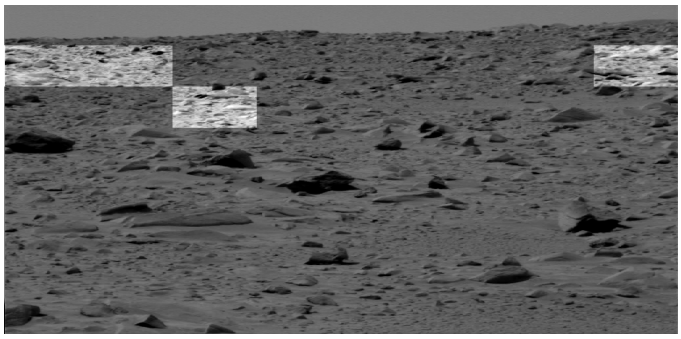
SAR5



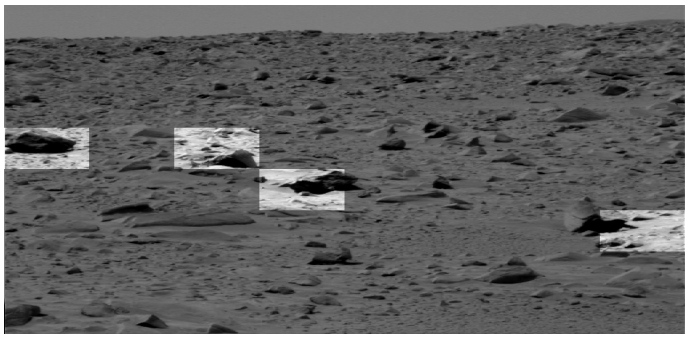
Subimages selected using entropy



Subimages selected using gradient



Subimages selected using alignabilty



Subimages selected using variance

Figure 3.14 Top 4 subimages of each feature selected for SAR 5

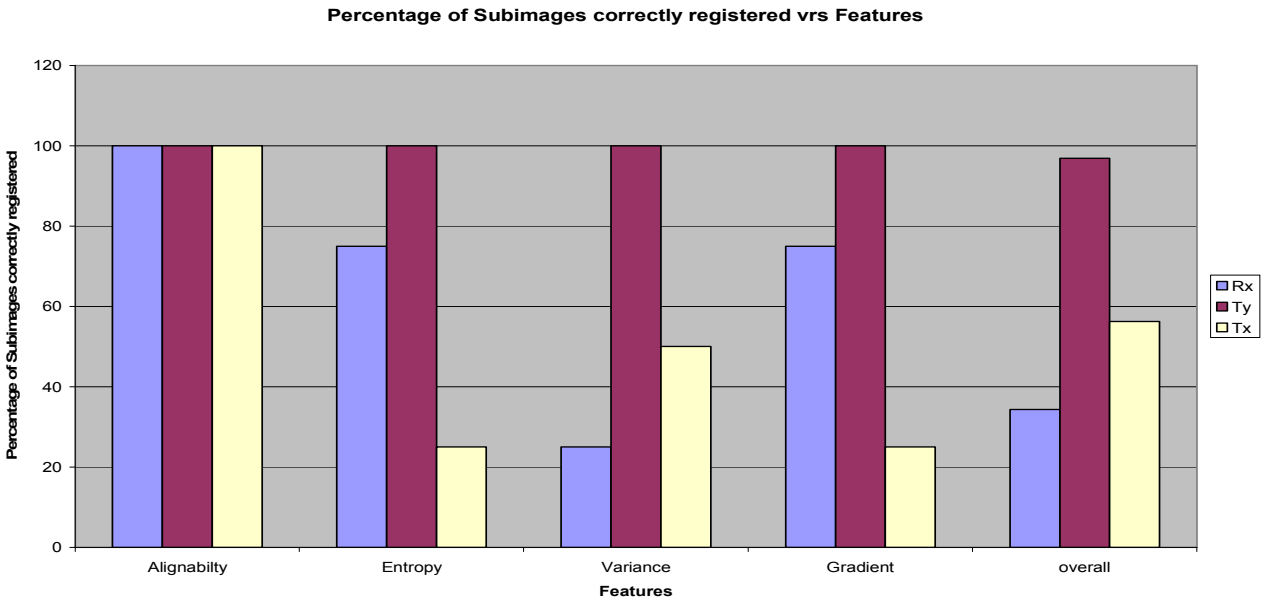
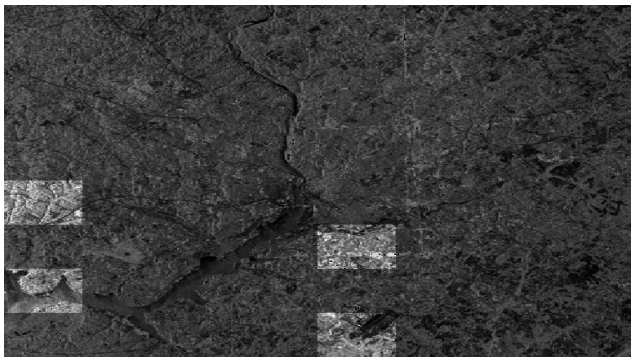
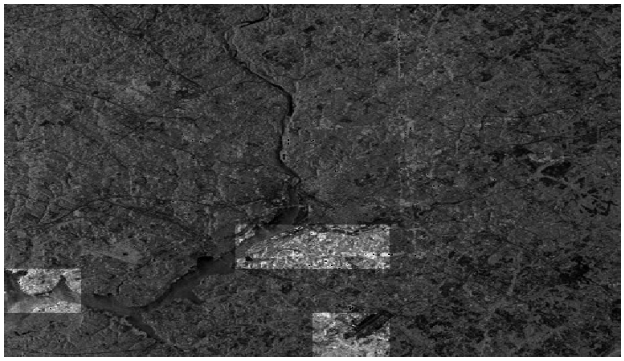


Figure 3.15 Percentage of correctly registered subimages versus different selection methods for SAR5

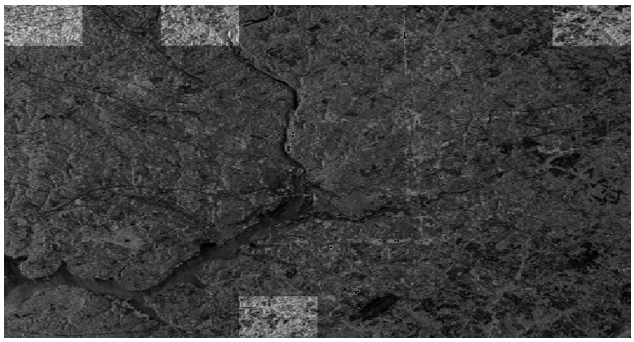
SAR6



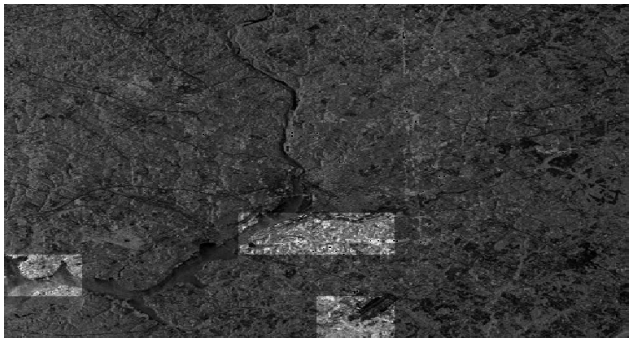
Subimages selected using entropy



Subimages selected using gradient



Subimages selected using alignabilty



Subimages selected using variance

Figure 3.16 Top 4 subimages of each feature selected for SAR 6

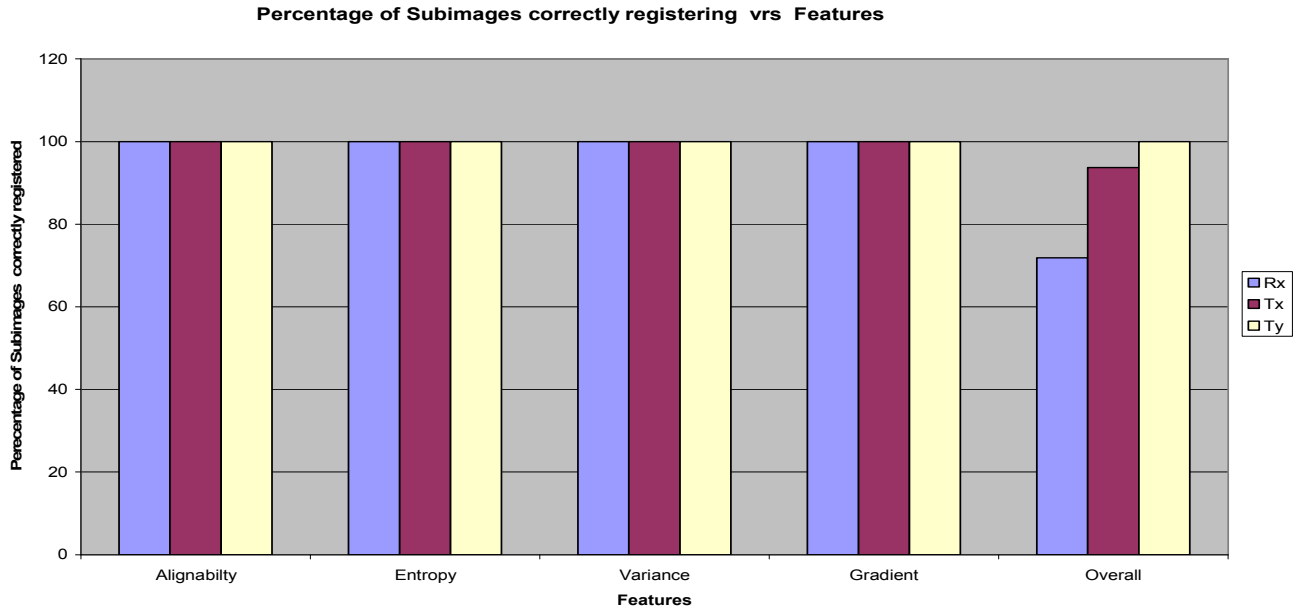


Figure 3.17 Percentage of correctly registered subimages versus different selection methods for SAR 6

The Overall percentage of selected subimages correctly registered and corresponding root mean square error

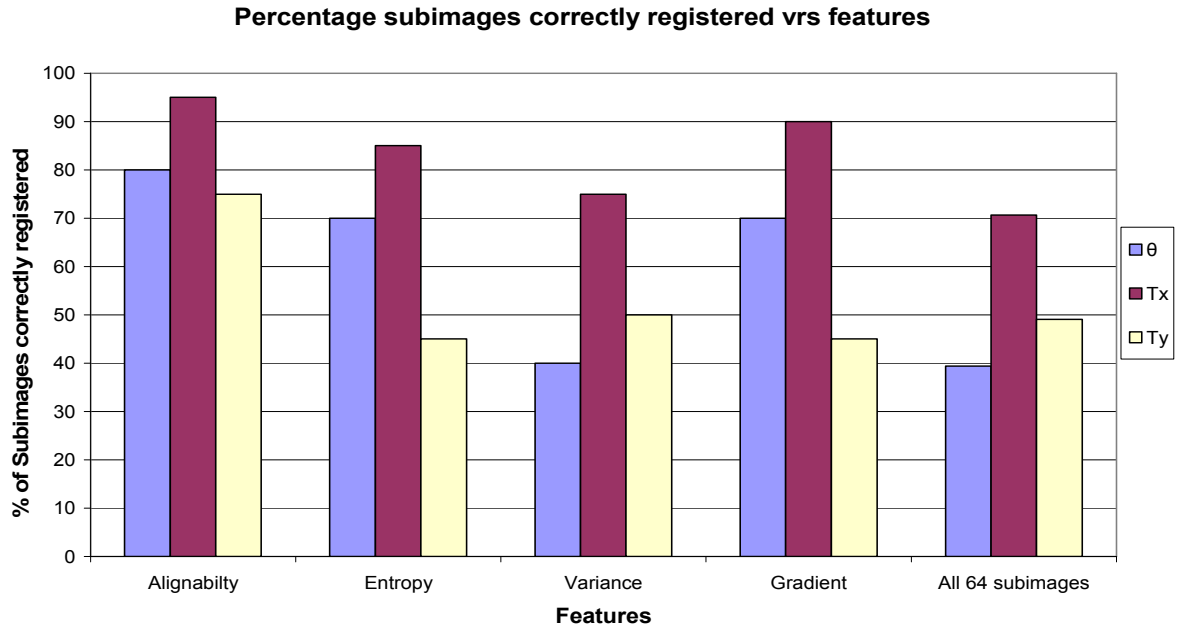


Figure 3.18 Percentage of correctly registered subimages verses different selection methods for the all the images

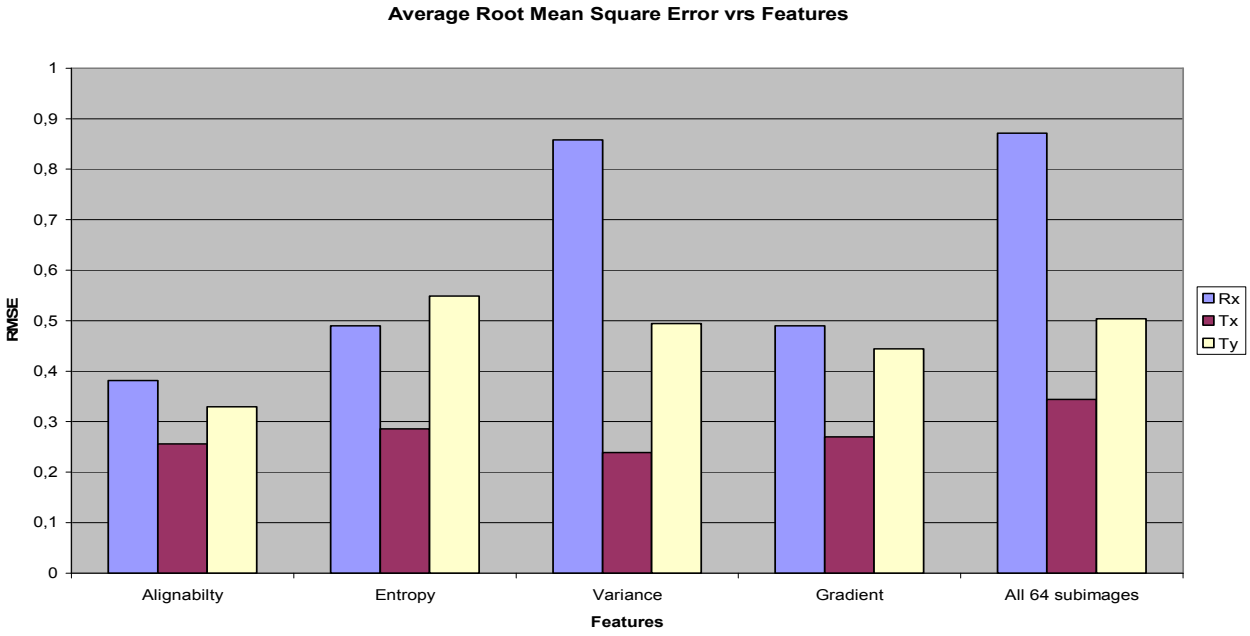


Figure 3.19 Root mean square error versus different selection methods for all images

3.4 Discussion and Analysis

The above results show that using alignability as a selection method is not only reliable but accurate as well. The overall percentage of correctly registered (as shown in figure 3.18) subimages for alignalbility-based selection method is 20%, 19% and 32% more than entropy-based method, gradient-based method and variance-based method respectively. The overall root mean square error for the alignabilty-based method is also the smallest among the selection methods (figure 3.19). The analysis of the individual images used for the experiment is as follows:

(i) In SAR1, 75% of the subimages selected by the gradient-based method and variance-based method are the same. 25% of the subimages selected by the entropy-based method are also by both gradient-based method and variance-based method or the alignability based method. None of subimages selected by the alignabilty-based method is selected the variance based method.

All the subimages selected by alignabity-based method are correctly for translation and rotation. Only 25% of the subimages failed to register the correct registration point for translation in x-direction (Tx) and y-direction (Ty) (figure 3.5).

(ii) None of the subimages selected by the variance-based method are selected by the other methods in Lena. 25% of subimages selected by alignability-based method are selected by either the gradient-based method or entropy-based method. The average performance of the selection methods are about the same. None of the subimages selected by the alignability based method or gradient-based method is correctly registered in the x-direction (Tx) (figure 3.7).

(iii) In SAR2, 75% of subimage selected by the variance-based method, entropy-based method and gradient-method are the same. The alignability-based method completely selected different subimages. The alignability-based method performed considerably better than the other selection method in finding the correct registration point.

(iv) In SAR3, none of the subimages selected by the alignablity-based selection method were selected by the other methods. This is also true for the entropy-based method. 25% of subimages selected by the gradient-based method and variance-based method are the same. The average performance of the alignablity-based method and gradient are the same and better than the other methods (on the average). 80% of the subimages selected by both were correctly registered (figure 3.11).

(v) In SAR4, none of the subimages selected by the gradient-based method is selected by the other methods. This is the same for the variance-based method. 25% of subimages selected by the alignability based method and entropy based method is the same. The average performace of the alignabilty base method is better than the other methods (figure 3.3). Only 7% of subimages (on the average) chosen by the alignabilty-method failed to determine the correct registration point (figure 3.13).

(vi) In SAR5, 25% of the subimages selected by the alignability-based method, entropy-based method, and gradient-based method are the same. 25% of subimages selected by the gradient-based method and variance-based method are the same. The alignablity-based method performed much better than the other selection methods. All the subimages selected by the alignability-based method correctly determined the correct registration point (figure 3.17).

(vii) In SAR6, all the subimages selected by the variance-based method and gradient-based method are the same. The alignabilty-based method as well as the entropy-based method selected completely different subimages from the other selection methods. All the subimages selected by the selection methods were correctly registered (figure 3.19)

In summary, the histograms in figures 3.5, 3.7, 3.9, 3.11, 3.13, 3.15, and 3.17 show that the alignabilty-based method is consistently reliable. Figures 3.5, 3.15 and 3.17 show that all the subimages selected by the alignabilty-based method correctly registered image pairs been registered. This is not the case for other experiments as shown in figures 3.9, 3.7 and 3.11. Thus the performance of the

selection method is a function of the reference and input image. However, the alignability-based method is among the best selection method or the best selection method in the experiments performed (figures 3.5, 3.7, 3.9, 3.11, 3.13, 3.15, 3.17).

The computational cost of alignabilty-based method increases linearly with *O(N).* This is comparable to the other selection methods as the size of the subimage increases.

Table 3.1 The overall evaluation results for the different selection methods

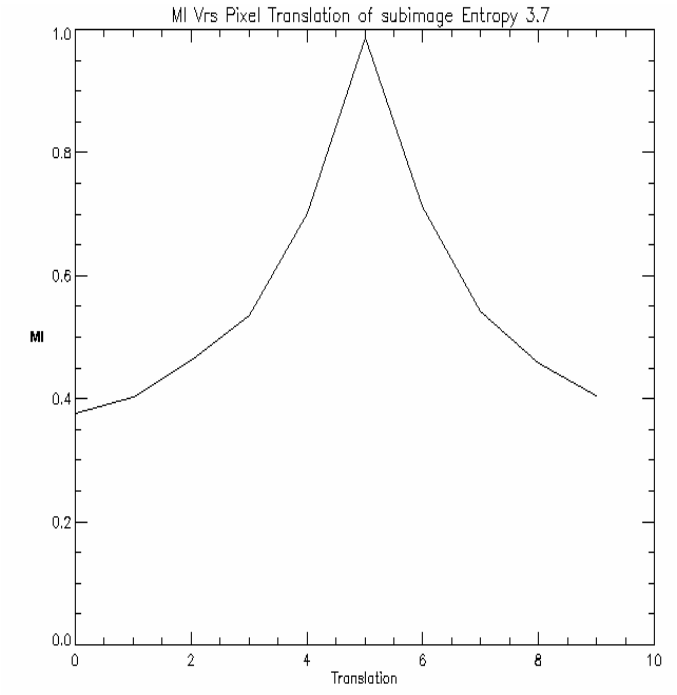
Feature	Alignalbility	Entropy	Gradient	Variance	All 64 Subimages
Total Percentage of Subimages correctly registered	87%	67%	68%	55%	53%
Average RMSE	0,322078	0,44159	0,401481	0,530056	0,572874

3.5 The Sharpness of MI curves for different features

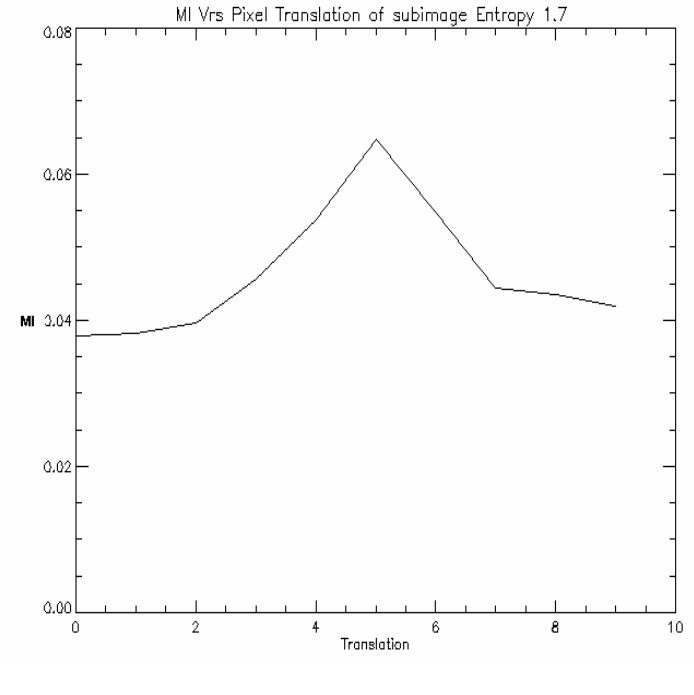
Results show that generally there is a correlation between sharpness of mutual information curves and the features considered. Figure 3.20 below shows that the subimages with higher alignability, entropy, gradient and variance (figure 3.3 (iii)) have a much sharper mutual information curve. In order to measure the sharpness of the MI curve for the features we normalized the MI curve to values between [0, 1]. The area under the curve is also computed, centred on the maximal point and bounded by points where MI curves intersect. Hence, the following assumptions are noted to be true:

1) The MI curves are defined and continuous in the range of finding the registration point.

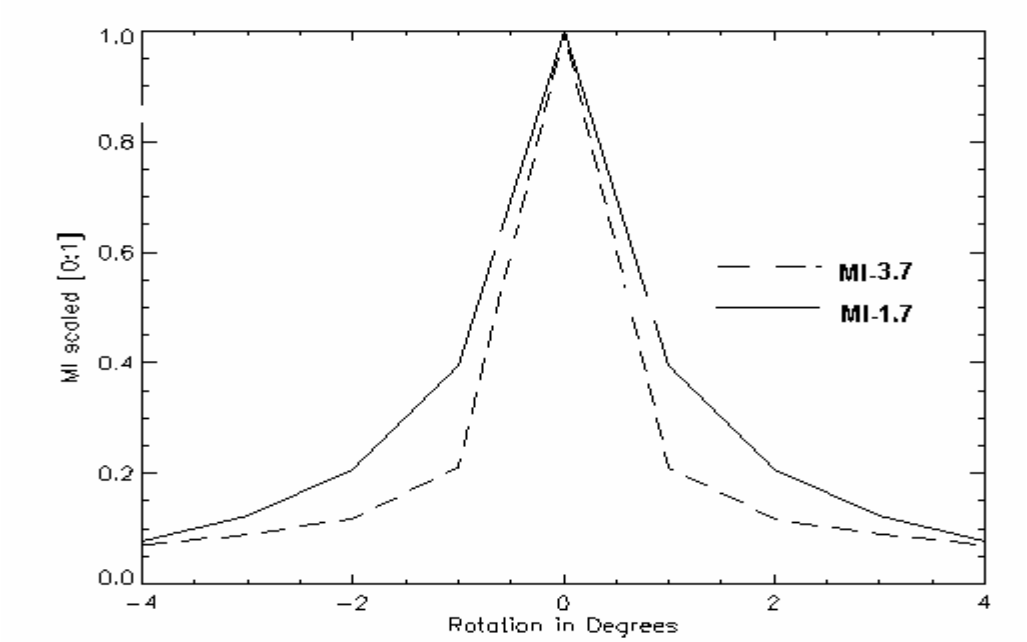
2) The MI curves are positive in the range.



(i) MI curve for registering translated version of subimage in figure 3.3 (i)



(i) MI curve for registering translated version of figure 3.3 (ii)



(iii) Scaled MI curve for registering rotated versions of figures 3.3(i) and 3.3 (ii) respectively

Figure 3.20 MI curves for templates with different entropies

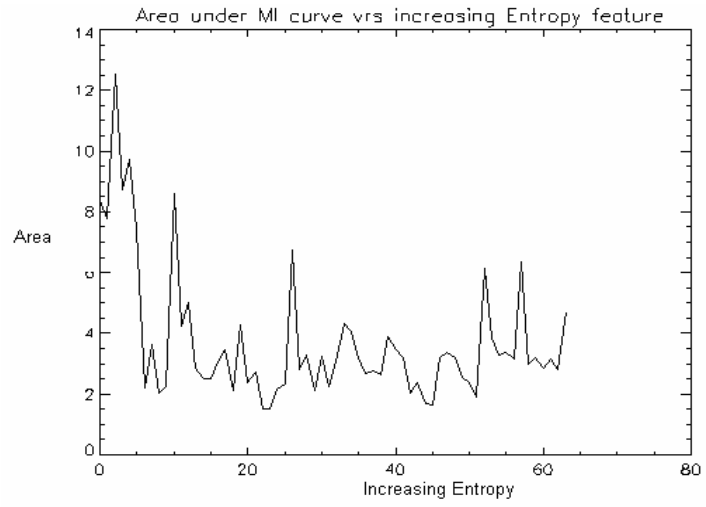
3) The only point of intersection (Rotation=0 in figure 3.20 (iii)) for the MI curves is the maximal point.

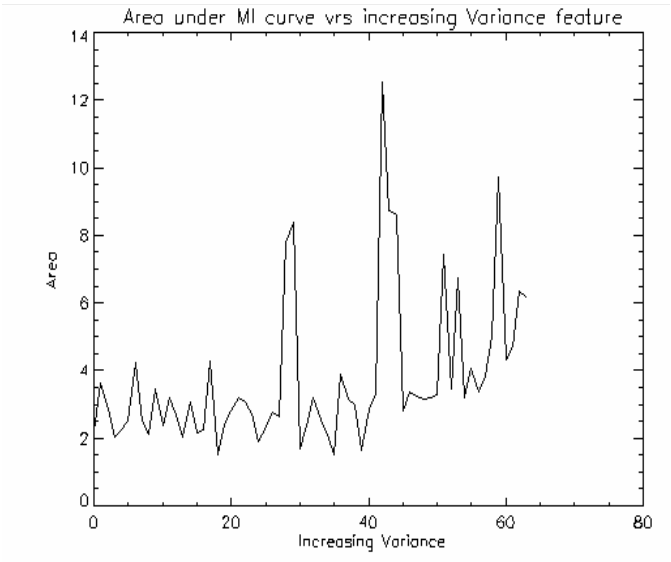
Based on the assumptions we say that a MI curve is sharper if the area under the curve is smaller. Since the only point the MI curves intersect is the maximal point, one can also use the slope of the curves to measure sharpness. The former definition is also valid for curves with points of inflexion.

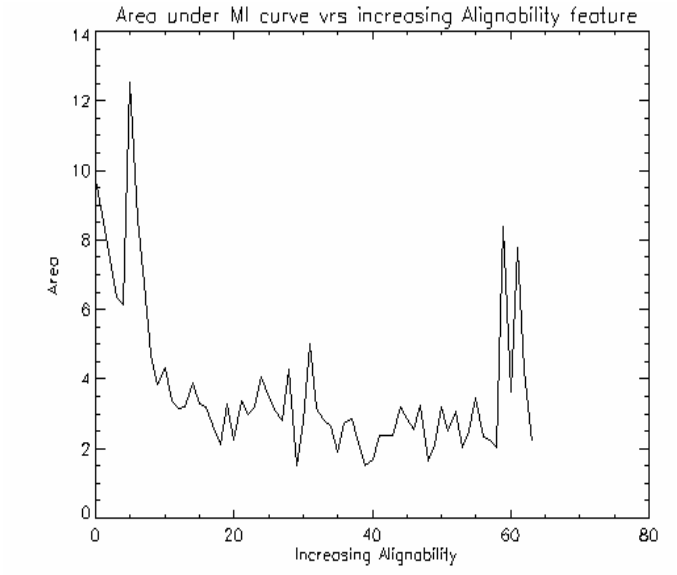
Table 3.2 Ground truth transformations and estimated registration points

Images	Ground truth (Rx,Ty,Tx)	Registration point (Rx,Ty,Tx)			
		Alignability	Entropy	Gradient	Variance
SAR1	(2/1/2)	(2/1/2)	(1/1/2)	(2/1/3)	(2/2/3)
Lena	(0/2/2)	(0/2/5)	(0/2/5)	(0/2/6)	(0/4/4)
SAR2	(2/2/1)	(2/2/1)	(2/2/2)	(2/2/3)	(1/2/2)
SAR3	(-2/2/0)	(-2/2/0)	(-2/2/0)	(-2/2/0)	(-2/1/0)
SAR4	(1/-2/0)	(1/-2/0)	(2/-2/0)	(3/-2/1)	(1/-2/1)
SAR5	(0/-2/-2)	(0/-2/-2)	(0/2/6)	(0/-2/2)	(1/-2/0)
SAR6	(1/-1/-1)	(1/-1/-1)	(1/-1/-1)	(1/-1/-1)	(1/-1/-1)

We also measured the correlation between area under the MI curve and the features mentioned. The correlation was estimated by using correlation coefficient (equation 2.2). The reference image used for this experiment is SAR1. The correlation between the alignability-based selection method, gradient-based selection method, entropy based selection method are -0.73, -0.48, -0.62, 0.28 respectively. Smaller areas correspond to sharper curves. Apart from the variance-based selection, there is a positive correlation between the sharpness of MI curves and increasing magnitude of the features (see figure 3.21 below). The alignability-based selection method shows the highest correlation with sharpness of MI curves.







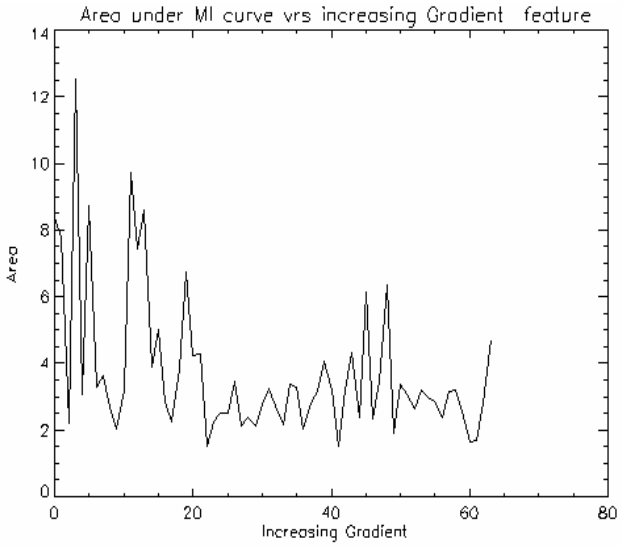


Figure 3.21 Graphs showing the area under curves and respective feature. The noisy nature of the curves indicate that the areas under the curve does not only depend on a feature. The correlation between the area and the alignability-based selection method, gradient-based selection method, entropy based selection method and variance-based method are -0.73, -0.48, -0.62, 0.28 respectively

3.6 Reliability of similarity measures

We also used the reference images to compare the reliability of mutual information (MI), correlation coefficient (CC) and sum of absolute differences (SD). Again we added Gaussian noise to input (slave) images with a SNR of 20 dB. Five images (SAR1, Lena, SAR2, SAR3, SAR4) were used for the experiment. The input images are translated to the reference images between -3 to 3 pixels for both x-direction and y-direction. The images are divided into 64 subimages of size 96 x 64. All 64 subimages are registered by maximizing *MI, CC* and *SD.*

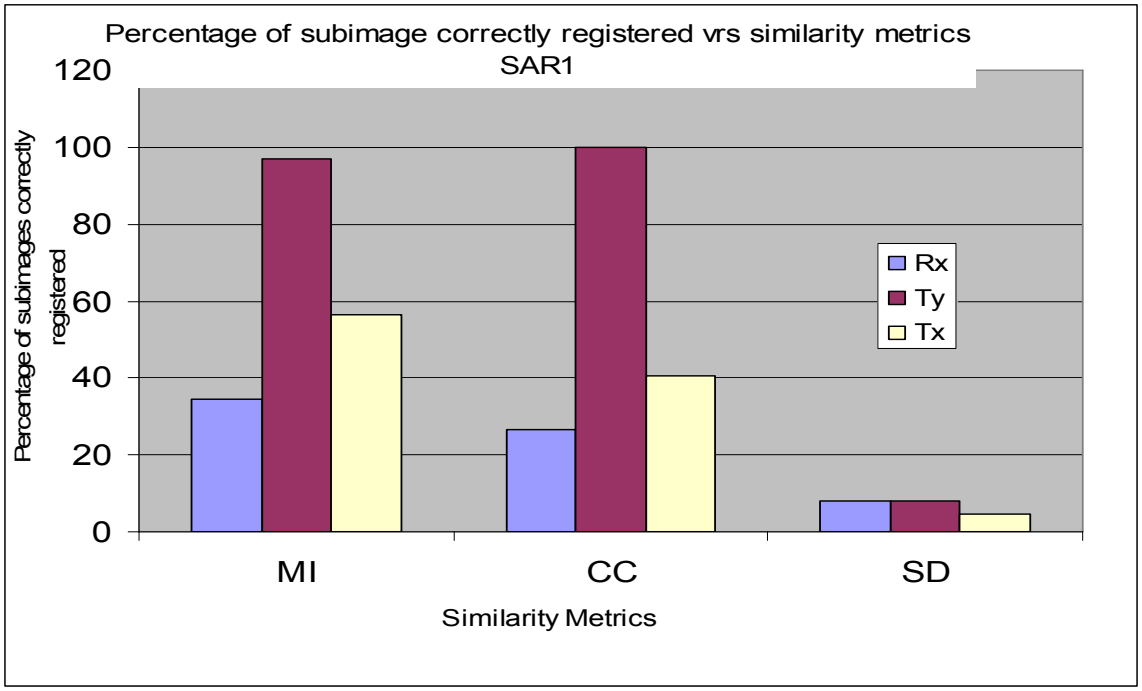
Figure 3.22 (i) shows that the robustness of MI is higher than CC when SAR1 is used as the reference image. The percentage of correctly registered subimages for *MI* and *CC* are 63%and 55% respectively. Only 8% of the subimages are correctly registered when *SD* is used as similarity metric.

CC shows slightly more robustness than MI when Lena is used as reference image (figure 3.22(ii)). The average percentages of correctly registered subimages for CC, MI and SD are 63%, 57% and 9% respectively.

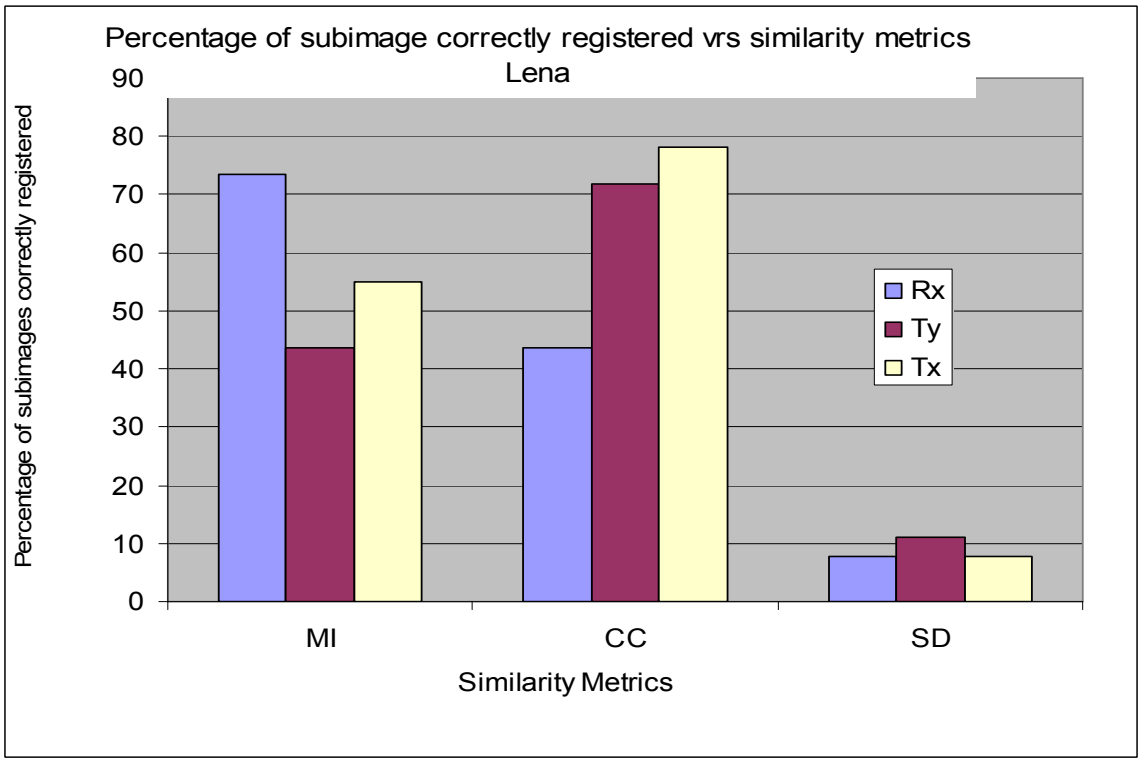
The robustness of CC and MI is about the same when SAR2 is used as reference image. Only 9% of subimages are correctly registered when SD is used as similarity metric (figure 3.23(i)).

MI shows more robustness than CC when SAR3 is used as reference image. The mean percentages of the correctly registered subimages for MI, CC, and SD are 49%, 40 and 10% respectively (figure 3.23(ii)).

Figure 3.24(i) shows that MI is more robust than CC when SAR4 is used as reference image. The average percentages of the correctly registered subimages for MI, CC, and SD are 70% ,50%, and 9% respectively.

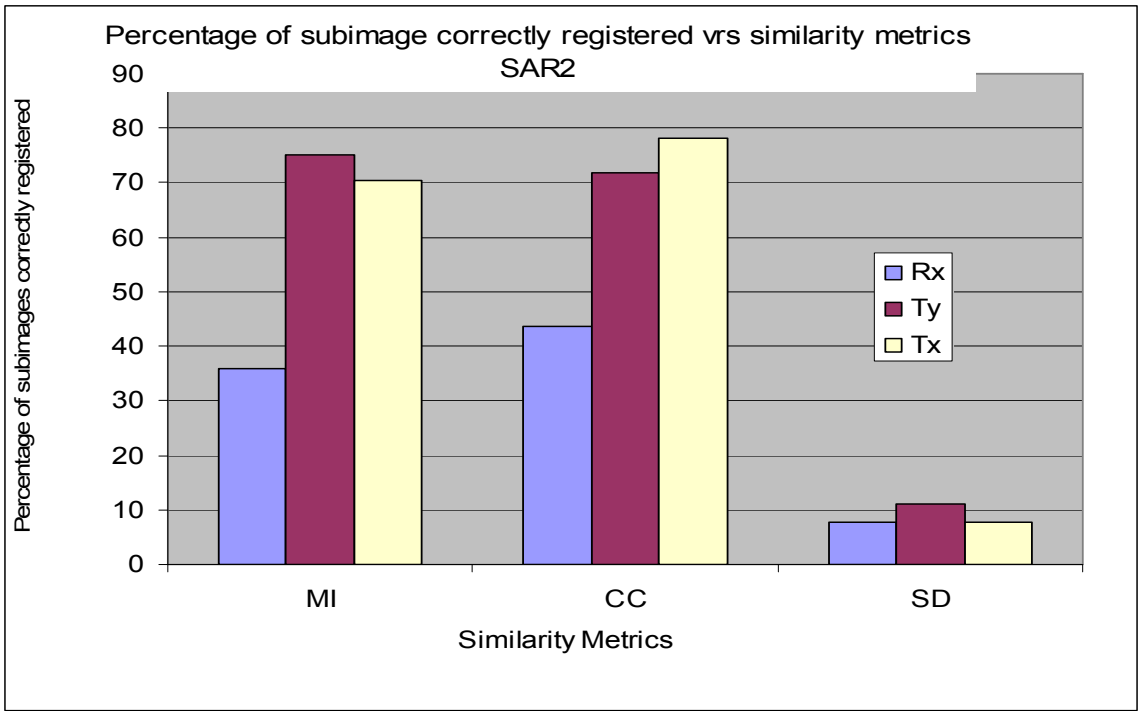


(i)

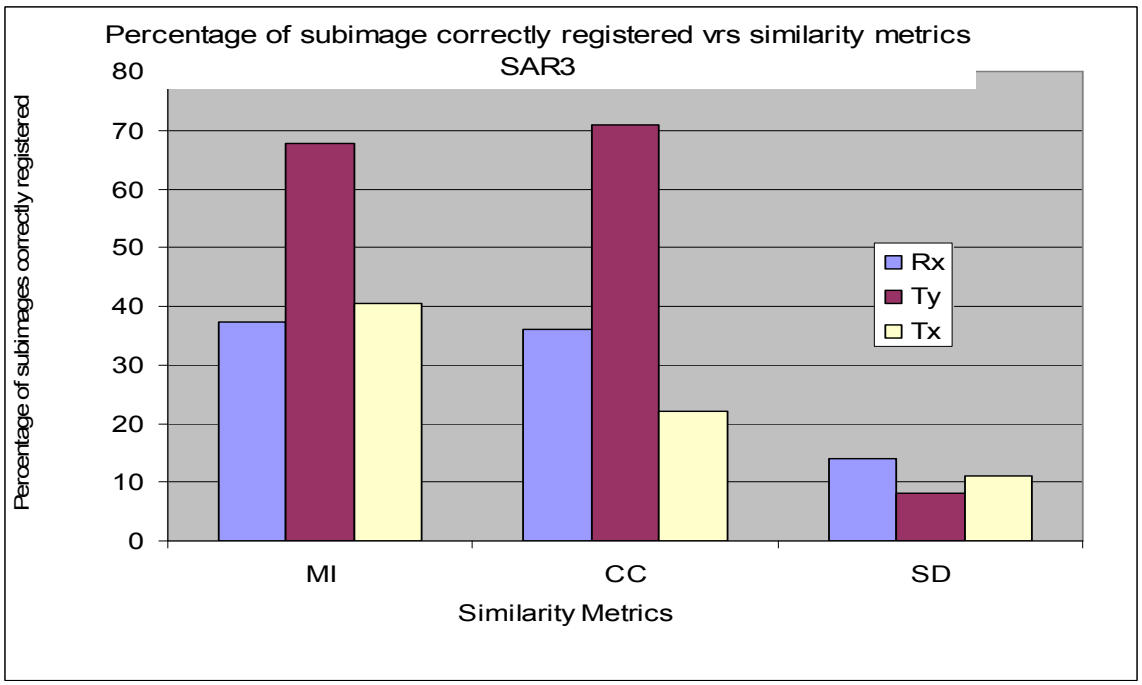


(ii)

Figure 3.22 Similarity metrics and their corresponding reliability for rotation and translation on x-axis and y-axis respectively for SAR1 (i) and Lena (ii).

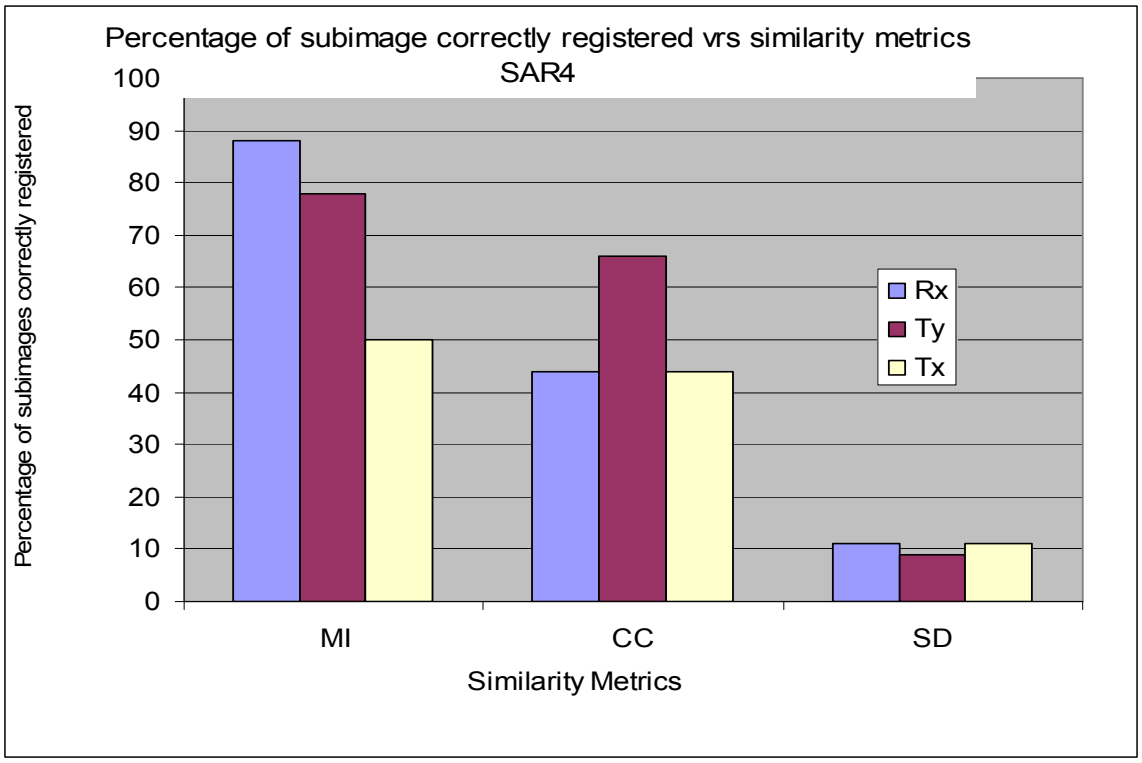


(i)

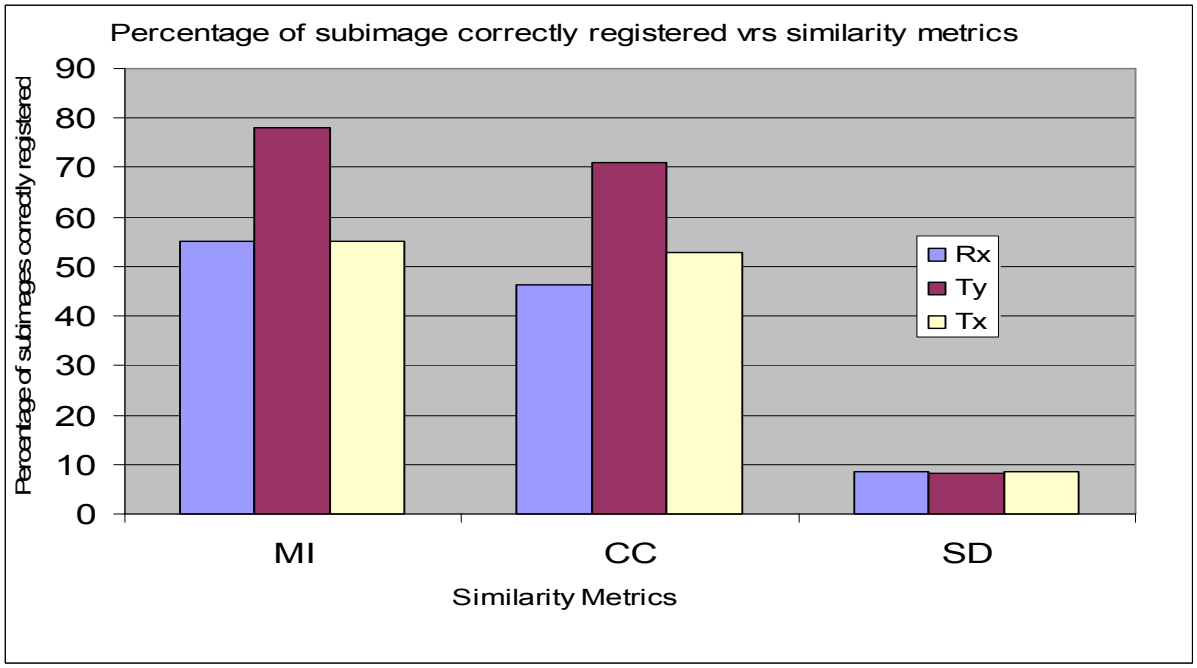


(ii)

Figure 3.23 Similarity metrics and their corresponding reliability for rotation and translation on x-axis and y-axis respectively for SAR2 and SAR3.



(i)



(ii)

Figure 3.24 Similarity metrics and their corresponding reliability for rotation and translation on x-axis and y-axis respectively for SAR4 and the overall average for the five images.

Table 3.3 RMSE for the similarity metrics

Similarity Metrics	MI	CC	SD
RMSE	0.32345	0.322078	1,322078

In summary, the results in figures 3.22, 3.33, 3.24 show that mutual information and correlation coefficient are considerably more reliable than the sum of absolute differences. Mutual information produced not only the most reliable results, but the most accurate as well. The overall percentage of subimages correctly registered for *CC, MI* and *SD* are 56%, 62%, and 8% respectively. The results of the root mean square error is presented in Table 3.3. There is no much difference between RMSE for MI and CC. The RMSE for SD is much higher. The results show that SD is relatively unreliable. Figures 3.22, 3.23, and 3.24 show that using SD is not suitable for registering image pairs with a SNR of 20dB. SD consistently performed poorly in finding the registration points.

3.8 Summary

In this chapter, we discussed methods used in the selection of subimages. We showed that, using subimages can reduce the computational cost in image registration. We demonstrated that alginability can produce more reliable results comparing with entropy, variance, and gradient. We also compared the reliability and accuracy of mutual information, correlation coefficient and sum of absolute difference respectively. Mutual information and correlation coefficient are considerably more reliably and accurate than the sum of absolute differences.

In the experiment, 20dB Gaussian noise is added only to the slave image which mimics true situations, since the reference is assumed to contain the noise from sensors. If no noise is added all the image registration experiments would give

the correct registration regardless of the translations in the input image which would not be realistic.

CHAPTER 4

ENTROPY-BASED SIMILARITY METRICS FOR IMAGE REGISTRATION: MUTUAL INFORMATION

Mutual information is a basic concept in information theory. It is applied in the context of image registration to measure the amount information that one image contains about the other. The maximization of mutual information criterion postulates that mutual information is maximal if images are correctly registered. Mutual information has demonstrated to be a very general and powerful similarity metric; it can be applied automatically and very reliably, without prior pre-processing, on a large variety of applications. The research that eventually lead to the introduction of mutual information dates back to the 1990's and is found in [28]. An image registration measure was introduced by Woods et al [28] which is based on the assumption that regions of similar grey values in one image would correspond to regions in the other image that contain similar grey values. Since the ratio of these grey values of all corresponding points in a certain region does not vary much, the average variance of this ratio for all regions is minimized to achieve registration. Joint histograms of images being registered show dispersion with misregistration. Studholme et al suggested the use of entropy to measure dispersion [29]. Shortly after that, Collignon et al [30] and Voila end Wells [31] introduced mutual information.

In this chapter, we introduce the basic theory of that underlies entropy and mutual information and investigate mutual information as a similarity metric. We show the effect of noise on mutual information and compare the result with that obtained using correlation as a similarity metric. We also propose two methods for the selection of bin size in the estimation of mutual information. We compare the bin size selections method with other proposed methods. The robustness of different I-information measures are compared with *MI*.

4.1 Basic Statistics

4.1.1 Mean

A random variable is a variable whose value is not predictable. The modelling of a process as random can be helpful in predicting the behaviour of the process. A random process (stochastic process) is every process which progresses in time according to stochastic laws. It is a family of time functions, *Y(t),* one function for each outcome of a random experiment, which is represented by the random variable Y. The voltage across a resistor can be modelled as a random process. Though the voltage across a resistor is unpredictable, its expected value can be determined. The expected value of the random variable is also the mean $E_Y[Y]$, which is defined as

$$E[Y] = \sum_{y_i \in Y} y_i P(Y = y_i) \qquad 4.1$$

where y_i are the realizations and Y the random variable. The expected value of a random variable is a deterministic function of its distribution. Intuitively *E[Y]* is the mean of the random variable value over a large sample *N.*

4.1.2 Variance

The expected value alone does tell much about the random variable. In order to know, on average, how close samples of Y are to the mean, the variance can be calculated. It is defined as

$$V(Y) = E[(Y - E[Y])^2] = E[Y^2] - E[Y]^2 \quad , or$$

$$V[Y] = \sum_{y_i \in Y} {y_i}^2 P(Y = y_i) - E[Y]^2 \qquad 4.3$$

The mean and variance are the first and second elements of a series of an infinite class of number moments in statistics.

4.1.3 Joint probability

Two random variables X and Y are dependent if *Y= F(X)*. The knowledge of X allows for the prediction of Y. It is also possible to have two random variables which are correlated but not perfectly predictable from each other. An example is measuring noise in a city and the number of industries in the city. Knowing the amount of industries helps but it does not tell us everything. Finally, it also possible, that the random variables are totally independent. An example is the tossing of two dies.

The joint probability *P(X,Y),* can formalize the joint dependency of two random variables. It shows the co-occurrence of events from X and Y. Marginal probabilities can be computed from the joint probability.

$$P(Y) = \sum_{x \in X} P(Y, X = x) \qquad 4.3$$

In general, joint probabilities can be decomposed into a marginal probability and a conditional marginal probability,

$$P(X,Y) = P(X) \cdot P(Y/X) \qquad 4.4$$

where *P(Y/X)* is the conditional probability. Equivalently the conditional probability can be expressed as:

$$P(Y / X) = \frac{P(Y, X)}{P(X)} \qquad 4.5$$

Complete dependence can be determined for the case where for all $x \in X$,

$$P(Y= F(x)/X=x) =1 \quad . \qquad 4.6$$

Two random variables are independent if

$$P(X, Y) = P(X)\, P(Y) \quad or$$

$$P(Y/X) = P(Y). \qquad 4.7$$

4.2 Entropy and maximum likelihood

The likelihood of a sample s under X is actually the conditional probability, *P(s/X).* From Bayes' rule (equations 4.5 and 4.4),

$$P(X/s) = P(s/X)\frac{P(X)}{P(s)} \qquad 4.8$$

In order to estimate the maximum likelihood, one must multiply the sample likelihood with a correcting factor *P(X)/P(s*). *P(s*) can be arbitrary. The maximum likelihood model assumes that all models that are to be evaluated are equally likely to occur. Maximum likelihood assumes, thus, that P(X) is constant.

Let the probability of the co-occurrence of the trials in s (s is a sample under X of size N_s), $P_X(s) = P_X(x_1 = x_{s_1}, x_2 = x_{s_2} \ldots x_{s_{N_s}})$. In general, evaluating joint probabilities over large number of random variables is difficult. Maximum likelihood also assumes that different trials of X are independent. Thus, the

probability of the co-occurrence is the product of the product of the independent X,

$$P_X(x_{s_1}, x_{s_2} \ldots x_{s_{N_s}}) = \prod_{i=1}^{i=N_s} P_X(x_i) \quad , \tag{4.9}$$

where $P_X(x_{s_1}, x_{s_2} \ldots x_{s_{N_s}})$ is the likelihood probability and $P_X(x_i)$ is the probability of the variable samples, *x*. The goal is to select the most probable model given large sample measurements.

Further simplification can be achieved by finding the logarithm of $P_l(s)$

$$\log P_X(x_{s_1}, x_{s_2} \ldots x_{s_{N_s}}) = \sum_{i=1}^{i=N_s} \log P_X(x_i) \tag{4.10}$$

Since the entropy,

$$H_s(x) \equiv -\sum_{i=1}^{i=N_s} P_X(x_i) \log P_X(x_i) = -\frac{1}{N_s} \sum_{i=1}^{i=N_s} \log P_X(x_i), \tag{4.11}$$

where N_s is the sample size of s. We can conclude that

$$H_s(x) = -\frac{1}{N_s} \log(P_X(x_{s_1}, x_{s_2} \ldots x_{s_{Ns}})) \tag{4.12}$$

This shows that instead of finding the model that makes the data most likely, we could instead find the model that has the lowest empirical entropy, or we could present a new interpretation of entropy as a distribution, which has low entropy if the probability of a sample drawn from that distribution is high. A distribution has high entropy if the sample has low probability. As shown in figure 3.21, a density with a very narrow peak has low entropy because most of the samples will fall in

the region where the density is large. A very broad density has high entropy because the samples are spread out and fall where the density is low.

4.3 F-divergence

F-divergence measures the expectation of the likelihood ratio between two distributions $P = \{p(x_i \mid i = 1,...,n)\}$ and $Q = \{q(x_i \mid i = 1,...,n)\}$, defined as

$$f(P \,||\, Q) = \sum_i q(x_i) f\left(\frac{p(x_i)}{q(x_i)}\right) \qquad 4.13$$

where *f* is a convex function. An interesting use of divergence concepts is similarity measures. The divergence measure reaches a minimum when two probability distributions are identical and increases with increasing disparity between the two distributions.

A function *f* is convex over an interval [a, b], if for every $x_1, x_2 \in [a,b]$ and $0 \le k \le 1$,

$$f\left(kx_1 + (1-k)x_2\right) \le kf(x_1) + (1-k) f(x_2). \qquad 4.14$$

A function is strictly convex if equality sign holds only if *k=0* or *k=1*. A function is concave if - *f(x)* is convex. Examples of strictly convex functions are x^2 and *x log x* (for $x \ge 0$) and *log x* (for $x \ge 0$) is strictly concave.

4.3.1 Cross entropy and maximum likelihood

The asymmetric divergence or cross entropy *D(P||Q)* is the measure of difference between two distributions:

$$D(P \parallel Q) = E\left[\log(\frac{p(X)}{q(X)})\right] \quad 4.15$$

$$D(P \parallel Q) = \sum_i \log(\frac{p(x_i)}{q(x_i)}) \cdot p(x_i) \quad 4.16$$

$$D(P \parallel Q) = \sum_i p(x_i) \log(p(x_i) - \sum_i p(x_i) \log(q(x_i)) \quad 4.17$$

$$= -H(X) - E_X[\log q(X)] \quad 4.18$$

If Q is the probability distribution of sample S, then

$$= -H(X) + H_S(X) \quad 4.19$$

The first term H(X) is constant, the second term (cross entropy) is – N_S (size of sample S) times the log likelihood of sample drawn from *X*. The asymmetric divergence is greater or equal to zero (Gibb's inequality). It is zero if and only if p*(X)* and *q(X)* are the same. Whereas the maximum likelihood looks for the model that makes sample most likely, asymmetric divergence looks for the model that is closest, in the cross entropy sense to the true distribution.

4.4 F-information

F-information is a special case of f-divergence. F-information measures apply only to specific probability distributions; namely, the dependence between the joint probability and their joint probability when they are independent [33].

$$I(P(X,Y) \parallel P(X) \times P(Y)) = \sum_x \sum_y p(X,Y) f\left(\frac{p(X,Y}{p(X) \cdot p(Y)}\right) \quad 4.20$$

where P*(X, Y)* is the joint probability distribution and *P(X) x P(Y)* is the set of joint probability distributions *P(X)·P(Y*) assuming *X* and *Y* are independent. *f* is a convex function.

In this section, we will limit ourselves to the f-information measures frequently encountered in literature.

4.4.1 Examples of F-information

(i) The Kolomogorov-Smirnov measure is one of the simplest measures of dependence. It is obtained by using the function $I_{KS} = |x-1|$, and gives the equation

$$I_{KS}(P(X,Y) \,||\, P(X) \times P(Y)) = \sum_x \sum_y \left| p(X,Y) - p(X) \cdot p(Y) \right| \quad . \tag{4.21}$$

(ii) Another popular measure is the I-information measure [48]

$$I_\alpha(P(X,Y) \,||\, P(X) \times P(Y)) = \frac{-1}{\alpha(\alpha-1)} \left(\sum_x \sum_y \frac{p(X,Y)^\alpha}{(p(X)p(Y))^{\alpha-1}} - 1 \right) \tag{4.22}$$

for $\alpha \neq 0, \alpha \neq 1$. The limit as α approaches 0 the I-information equals mutual information. It can proved that the limit that the limit as α approach 0 Iα(P(X)||P(Y) also equals Kullback-Leibler distance.

$$\lim_{\alpha \to 1} I_\alpha(P(X) \,||\, P(Y)) = \lim_{\alpha \to 1} \frac{-1}{\alpha(\alpha-1)} \left(\sum_x \sum_y \frac{p(X)^\alpha}{(p(Y)^{\alpha-1}} - 1 \right) \tag{4.23}$$

$$= \lim_{\alpha \to 1} \frac{(p(X)^\alpha \, p(Y)^{1-\alpha} \log(p(X)) - p(X)^\alpha \, p(Y)^{1-\alpha} \log(p(Y)))}{(2\alpha - 1)} \tag{4.24}$$

$$= \sum p(X) \log\left(\frac{p(X)}{p(Y)} \right) \tag{4.25}$$

Using l'Hopital's rule in the equation 4.24 : if both $\lim_{\alpha \to 1} g(x)$ and $\lim_{\alpha \to 1} h(x)$ equal zero, $\lim_{\alpha \to 1} \left(g(x) / h(x) \right) = \lim_{\alpha \to 1} \left(g'(x) / h'(x) \right)$.

(iii) χ^2 measure or Pearson statistic is obtained if the function $I_{\chi}^2 = (x-1)^2$. Thus

$$I_{\chi^2}(P(X,Y) \,||\, P(X) \times P(Y)) = \sum_x \sum_y \left(p(X,Y) - p(X) \cdot p(Y)\right)^2 . \qquad 4.26$$

It is often used to measure the statistical dependence of images.

4.4.2 Sharpness of F-information similarity measures

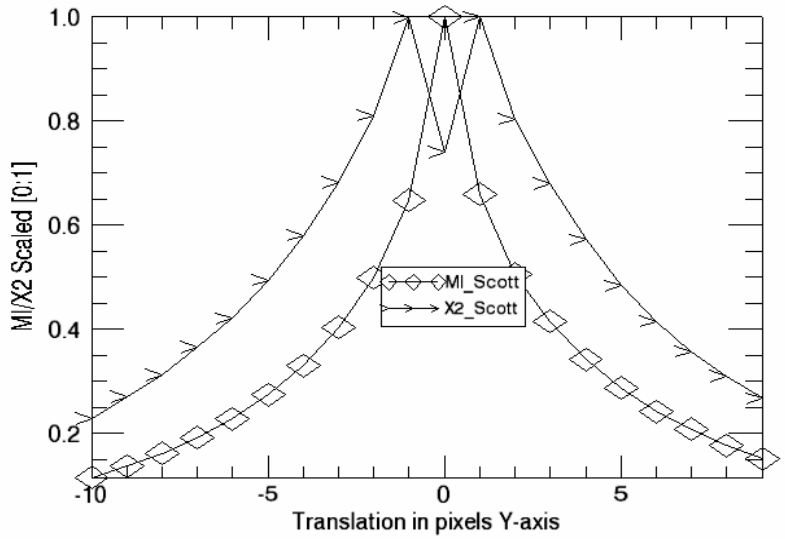
In order to compare the robustness and sharpness of I_{KS} (Kolmogorov-Smirnov) and *MI* curves, we registered reference images against their transformed version. We used the histogram method to estimate the Kolomogorov-Smirnov measure. We used Scott's rule [41] to select the bin size for the estimation of mutual information (MI_Scott) and Kolomogorov-Smirnov (KS_Scott) measure respectively. If *X* and *Y* are the reference and input images, M is the number of entries, $Hist_x(X)$ and $Hist_y(Y)$ are defined as their histograms, and $Hist_{ab}(X,Y)$ is the joint histogram, KS measure and χ^2 are defined as,

$$I_{KS}(X,Y) = \frac{1}{M} \sum_x \sum_y \left| Hist_{xy}(X,Y) - Hist_y(Y) \cdot Hist_x(X) \right|, \qquad 4.27$$

$$I_{\chi^2}(X,Y) = \frac{1}{M} \sum_x \sum_y \left(Hist_{xy}(X,Y) - Hist_y(Y) \cdot Hist_x(X) \right)^2 \quad , \qquad 4.28$$

respectively. The results are presented in figures 4.1, 4.2 and 4.3 respectively. The results show that mutual information is not only consistently sharper but also more robust.

I_X^2 and MI Curves vs. Translation on the Y-axis for first image



I_{KS} and MI Curves vs. Translation on the y-axis for second image

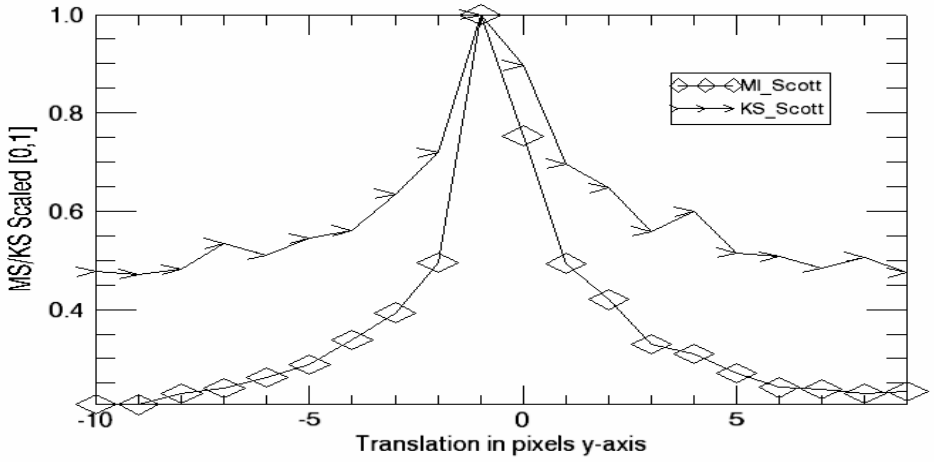
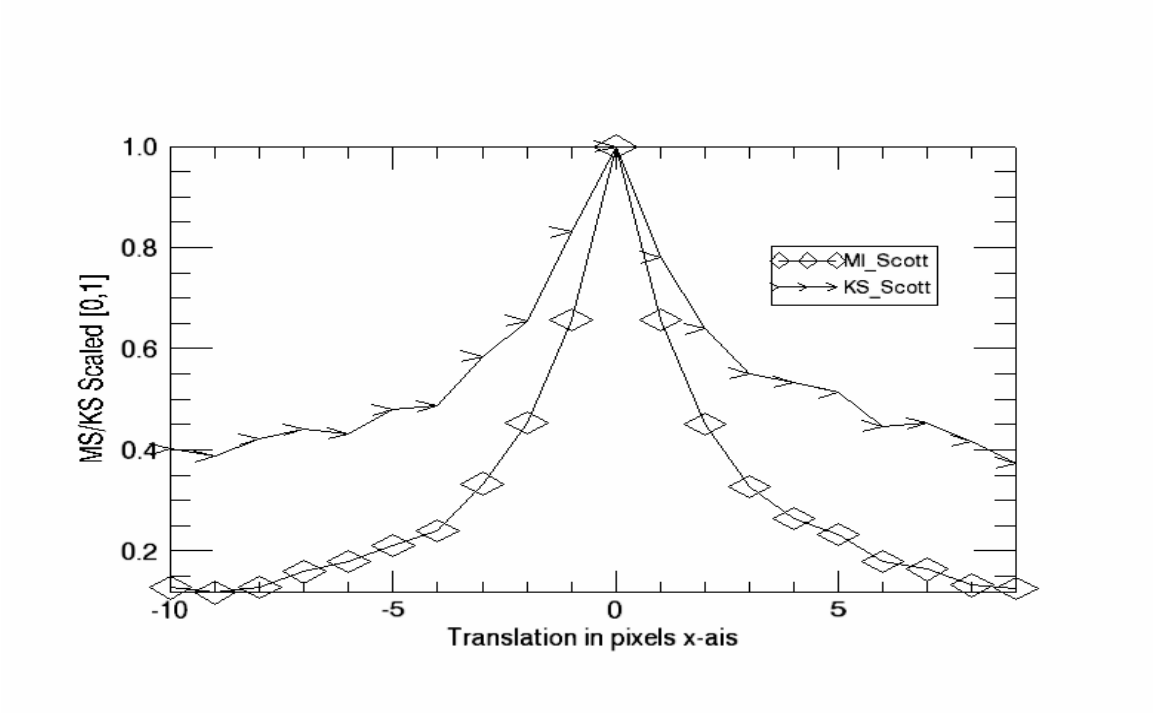


Figure 4.1 Graph showing sharpness similarity metrics with respect to y-axis. I_X^2 failed to determine the registration point in the first image. I_{KS} is not smooth in the second image. The roughness and non unimodality of I_{KS} and I_X^2 indicate that they are less robust than MI.

I_{KS} and MI Curves vs. Translation on the x-axis for second image



I_X^2 and MI Curves vs. Translation on the x-axis for first image

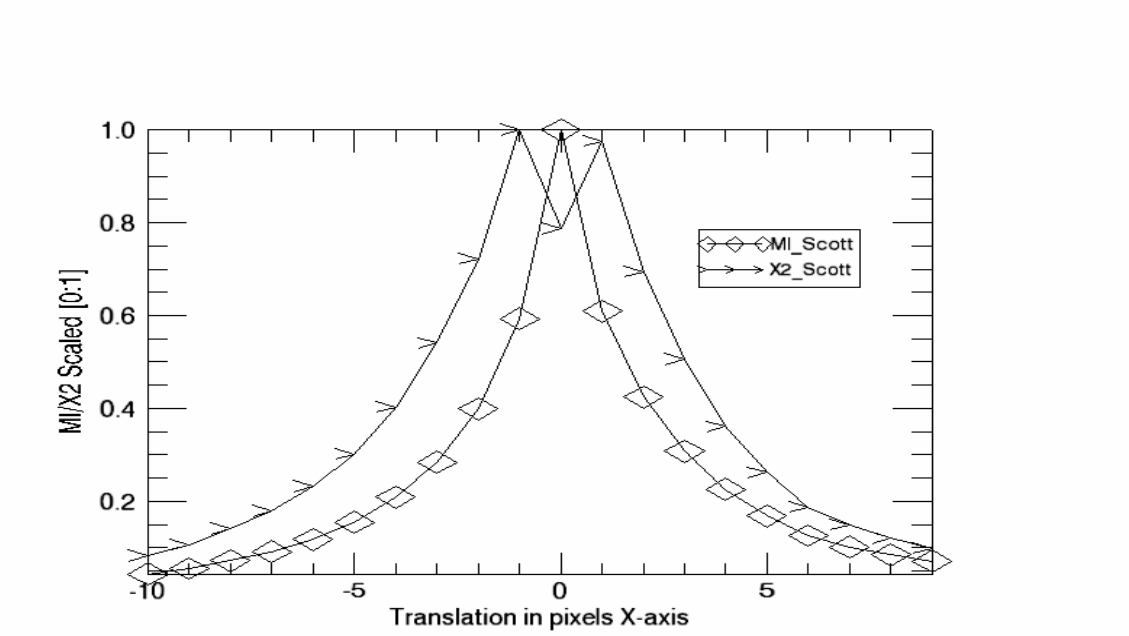
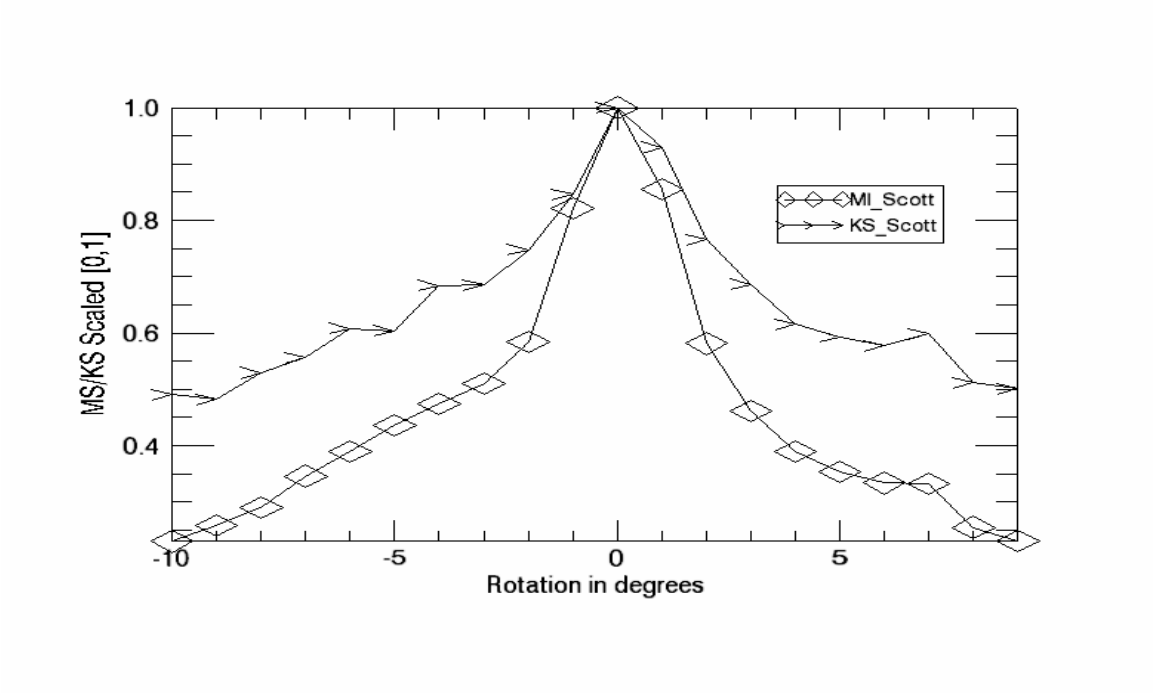


Figure 4.2 Graph showing sharpness similarity metrics with respect to x-axis. I_X^2 failed to determine the registration point in the first image. I_{KS} is not smooth in the second image

I_{KS} and MI Curves vs. Rotation for first image



I_{χ}^2 and MI Curves vs. Rotation for second image

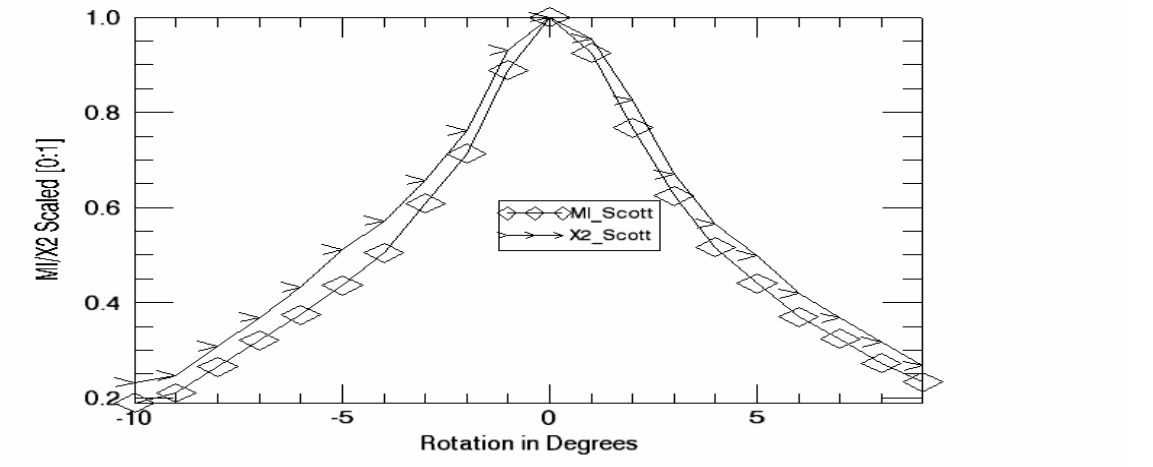


Figure 4.3 Graph showing sharpness similarity metrics with respect to rotation. Again, MI is smoother than I_{KS} .

4.5 Concept of mutual information

In section 4.1 we noted, that two random variables are independent if and only if their joint density is the product of their marginal densities. In order to determine how much a random variable depends on the other, entropy is employed. Entropy can be used to determine the predictability of two random variables. Joint entropy can be defined as:

$$H(X,Y) = -\sum_{a,b} p_{XY}(x,y) \log p_{XY}(x,y) = E_{XY}\left[\log\left(P(X,Y)\right)\right] \qquad 4.29$$

where E_X and E_Y are the means with random variables X and Y respectively.
If X and Y are independent, from equation 4.7

$$H(X, Y) = H(X) + H(Y). \qquad 4.30$$

Conditional entropy measures the randomness of a variable Y given X, and is given by

$$H(Y/X) = -\sum_{a,b} p_{XY}(x,y) \log p_{X/Y}(y/x) = E_{XY}\left[\log\left(P(Y/X)\right)\right] \qquad 4.31$$

From equation 4.30

$$H(Y/X) = H(X,Y) - H(X) \qquad 4.32$$

If Y becomes more independent of *X*, H*(Y/X)* becomes smaller. H(X/Y) gets larger as the dependence between the variable increases. H(Y/X) is equal to H(Y) if Y and X are independent.
Mutual information which uses entropy to quantify dependence between two variables is defined as:

$$MI(X,Y) = H(Y) - H(Y/X). \quad (4.33)$$

From equation 4.32

$$MI(X,Y) = H(X) - H(X,Y) + H(Y) \quad (4.34)$$

$$MI(X,Y) = H(X) - H(X/Y). \quad (4.35)$$

$$MI(X,Y) = MI(Y,X) \quad (4.36)$$

Equation 4.36 shows that mutual information is symmetric. This is a much desired property when using mutual information as a similarity metric.

4.6 Properties of mutual information

(i) Equation 4.36 above proves the symmetry of mutual information. In practice however, mutual information might not be symmetric. In image registration the various stages such as interpolation can affect the symmetry of mutual information [32], hence, MI(X,Y) might not necessarily be equal to MI(Y,X).

(ii) The mutual information of a random variable and itself is the same as the entropy the variable. From equation 4.34,

$$MI(X,X) = H(X) - H(X,X) + H(X) \quad (4.37)$$

Since $$H(X,X) = H(X) \quad (4.38)$$

$$MI(X,X) = H(X). \quad (4.39)$$

(iii) The information, variables contain about each other cannot be greater than the information in the variables themselves

$$MI(X,Y) \leq H(X) \ , \tag{4.40}$$

$$MI(X,Y) \leq H(Y). \tag{4.41}$$

(iv) *MI (X, Y)* is equal to 0 if and only if *Y* and *X* are independent. This means that knowledge of *X* does not give you any information about *Y*. Thus *X* and Y are not related in any way.

(v) The uncertainty of *X* cannot be increased by learning from Y or Z (random variables). In other words, the mutual information between random variables is at least zero.

$$MI(X,Y) \geq 0 \tag{4.42}$$

$$MI(X,Z) \geq 0 \tag{4.43}$$

(vi) If $U = F(X)$ and $V = F(Y)$ *are* individually measurable one-to-one functions then

$$MI(U,V) = MI(X,Y) \tag{4.44}$$

Properties in (iii) and (v) can be derived using Jensen's equality [34], and using the fact that logarithmic functions are concave functions. For a concave function H(X), it can be proven that from Jensen's inequality [34],

$$E[H(X)] \leq H(E[X]) \ . \tag{4.45}$$

where E[H(X)] is the expectation of the concave function (entropy).
This equation allows us to prove the following inequalities:

$$H(X) \geq 0 \quad (4.46)$$

$$H(X) \geq H(X/Y) \quad (4.47)$$

$$MI(X,Y) \geq 0 \quad (4.48)$$

(vii) If X and *Y* are two Gaussian random variables, then

$$MI(X,Y) = -0.5 \log(1- CC^2) \quad (4.49)$$

where *CC* is the correlation coefficient between *X* and *Y* [41].

4.7 Estimation of probability density function

The main difficulty in the estimation of mutual information from empirical data lies in the fact that the underlying probability density function is unknown. The three methods used to estimate mutual information are histogram-based estimators, kernel-based estimators and parametric methods.

4.7.1 Histogram-based estimator

Histogram based methods are the most straight forward way to estimate probability density functions or the joint probability density of two images [35][36]. Each entry in the histogram determines the number of times x in one image coincides with b in another image. The probability distribution is estimated by dividing the entries by the total number of entries,

$$P(X,Y) = \frac{Hist(X,Y)}{M} \quad (4.50)$$

where *Hist(X,Y)* is the histogram entries and M is the number of entries. If A and B are images to be registered, $Hist_a(A)$ and $Hist_b(B)$ are defined as their histograms, and $Hist_{ab}(A,B)$ joint histogram, then MI is defined by

$$MI = \frac{1}{M}\sum_{a}\sum_{b} Hist_{ab}(A,B) * \log\left(\frac{M * Hist_{ab}(A,B)}{Hist_{a}(A) * Hist_{b}(B)}\right) \quad 4.51$$

In our work we used this method to estimate mutual information.

4.7.2 Kernel-based estimator

The kernel-based method is also often used. Unlike the parameter model which uses parameters to define models, this method uses samples to directly define the model. The general form of the model is defined as:

$$P(x,s) = \frac{1}{N_s}\sum_{x_s \in S} G(x - x_s) \quad 4.52$$

where *s* is a sample and *G* is a valid density function [37]. Whereas the simple histogram method places a spike function which corresponds to x_s and updates only a single bin, the kernel-based method places a kernel at the bin x_s and updates all bins falling under the kernel corresponding to the kernel value. The quality of approximation is dependent on the width of *G* and its functional form. Whereas parametric methods make strong assumptions about the functional form to be approximated, the kernel-based method requires that the kernel be smooth. The Gaussian density is the most common selection of *G*, making the probability estimates a sum of Gaussians. Intuitively, this method computes local or windowed average.

4.7.3 Parameter-based method

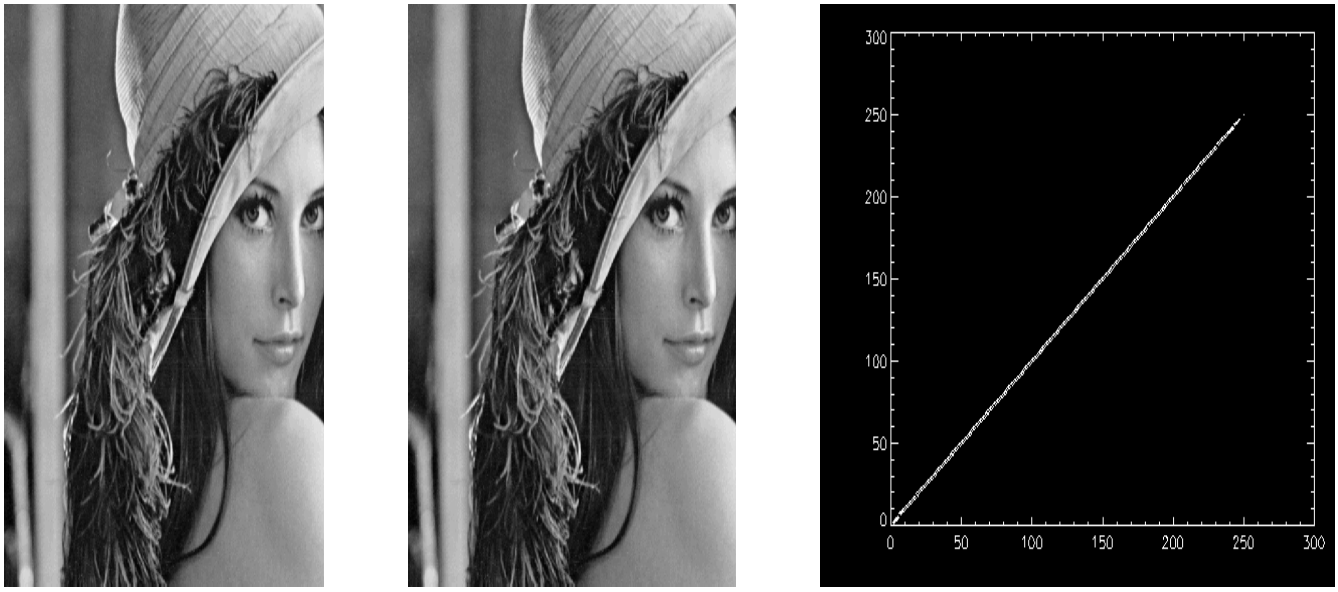
In the estimation of the histogram using the parameter-based method, the histogram is assumed to have a known distribution. Thus, in the registration of 2D images, a standard bivariate Gaussian would be used. The standard multivariate Gaussian has the following form

$$G_{\Psi}(x-\mu) = \frac{1}{(2\pi)^{\frac{m}{2}}|\psi|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)\psi^{-1}(x-\mu)). \quad 4.53$$

In an *m* dimensional space, the mean μ is also a m-vector. The variance is replaced by a covariance matrix ψ and $|\psi|$ is the determinant. The mean and covariance matrix are determined using a Bayesian or maximum likelihood estimation. The estimation is carried out by randomly selecting samples from the data.

4.8 Concept of feature space

In order to examine the effects of misregistration in image registration, Hill et al [38] introduced the concept of feature space. A feature space is the display of the 2D histogram of image pixel values by plotting the value of the pixel in the master image against corresponding value of slave image pixel by pixel. An example is illustrated in the figure 4.6 below. Figure 4.6 shows a perfectly registered image with corresponding 2D histogram with diagonal features.

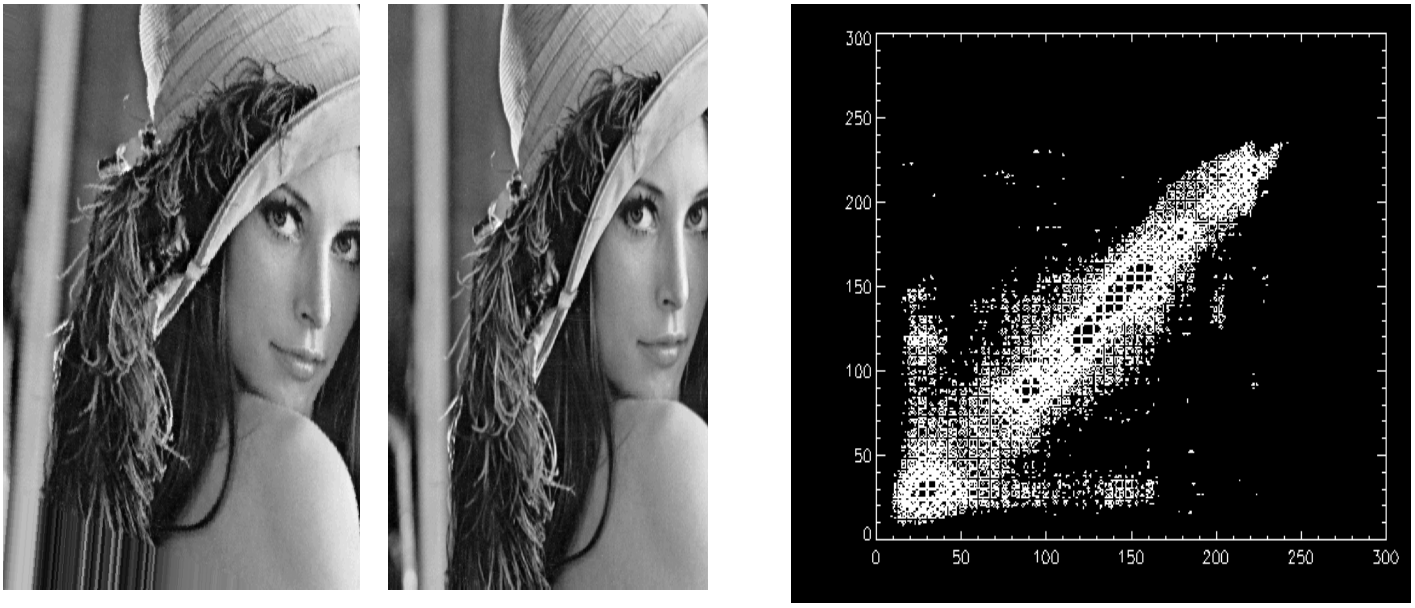


(a) (b) Joint entropy = 5.1

Figure 4.6 Joint grey value of image (b) and itself (a) (ie rotated 0°) their plot of feature space corresponding joint entropy value 5.1

Image rotated 1° Joint entropy is 6.9

Figure 4.7 Joint grey value of image (b) and input image (a) rotated 1° and their plot of feature space with corresponding joint entropy value of 6.9.



(a) (b) Joint entropy of 7.3

Figure 4.9 Joint grey value of image (b) and input image (a) rotated 5° and their plot of feature space corresponding joint entropy value of 7.3

(a) (b) Joint entropy 7.5

Figure 4.10 Joint grey value of image (b) and input image (a) rotated 10° and their plot of feature space corresponding joint entropy value of 7.5

(a) (b) Joint entropy 7.6

Figure 4.12 Joint grey value of image (b) and input image (a) rotated 20° and their plot of feature space corresponding joint entropy value of 7.6

(a) (b) Joint entropy 7.6

Figure 4.13 Joint grey value of image (b) and input image (a) rotated 30° and their plot of feature space corresponding joint entropy value of 7.6

If *F* is the feature space distribution, then *F(x,y)* is the number of pixels in reference image value x and having corresponding value of y in slave image. It is clear from the figure that when image pairs are misregistered, the pixels with value y in the reference image will have corresponding values with different values thus causing dispersion in the histogram as shown in the figures 4.7, 4.8, 4.9, 4.10, 4.11 and 4.12 above. If image pairs are registered then the pixel values x and y will be very similar or the same as shown in figure 4.6.

From the 2D histograms it can be deduced that:

1) Diagonal features diminish with increasing misregistration

2) The spatial dispersion pattern increases in area with increasing misgregistration.

3) Joint entropy increases with misregistration

4) The dispersion patterns show clear bounds.

Mutual information contains the term *–H(X,Y),* it means minimizing joint entropy is increasing mutual information. Since generally joint entropy increases with increasing misregistration, the mutual information decreases with increasing misregistration. In other words registration is assumed to correspond to the maximization of mutual information. If images are aligned the amount of information they contain about each other is maximal.

4.9 Effect of bin size for mutual information computation

The selection of bin size is crucial in the estimation of mutual information using the histogram-based method [39]. If the number of bins is increased, more image intensities are discriminated and the joint histogram becomes sparser because of the finite number of samples. The histogram will consequently seem more complex and small changes in the registration parameters will have larger and coarser effect on the joint histogram and MI criterion will become noisy or even fail. However, in decreasing the number of bins less image details are taken into account to evaluate the registration, since a larger range of intensities are mapped into a single bin.

In practice some researchers chose fixed number of bins which is determined empirically [40]. They choose different number of bins arbitrarily and then select the bin size corresponding to highest mutual information. Others also use a variable bin number based on gaussianity data [41].

In this section, we propose two methods namely Freedman and Diaconis rule and Scotts rule for the selection of bin size. We show the effect of bin size of mutual information. We also show the effect of Gaussian and speckle noise on mutual information and correlation coefficient. We also compare the bin size selection methods.

4.9.1 Bin size selection

For the selection of the bin size using Freedman and Diaconis's [61] rule the class width is determined by:

$$W = 2(IQ)N^{-1/3} \tag{4.54}$$

where *IQ* is sample interquartile range, and N is the number of available samples. The interquartile range is the difference between the third quartile (75^{th} percentile) and the first quartile (25^{th} percentile).

For the Scott's rule method [61], the class width is determined by:

$$W = 2sN^{-1/3} \tag{4.55}$$

where *s* is the sample standard deviation. The main idea is to obtain perfectly smoothed histograms.

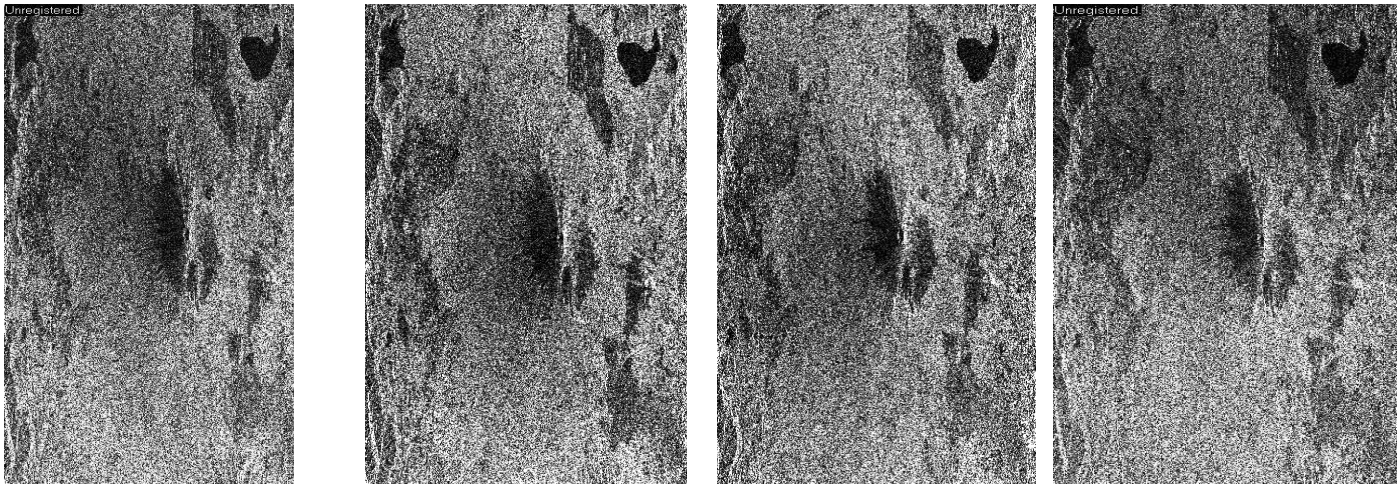
Another method often used in speech processing is the gaussianity rule [41], which proposes that the number of bins

$$Z = \log_2\left(1 + k\sqrt{N/6}\right) + 1 + \log_2 N \tag{4.56}$$

where *k* is the kurtosis. The main idea is to select a bin size that makes the histogram of image as close to the Gaussian distribution as possible.

4.9.2 Experiment showing the effect of bin size

In this section, we present results of different tests, which provide a comparison between MI estimated using Scott's rule, Friedman and Diaconi's rule and chosen fixed number of bins. The experiments are performed on multitemporal images of size 318 x 625 shown in figure 4.4 below. The images are from the JERS1 (Japan Earth Resources Satellite). We register the images using mutual information estimated with a bin size of one (MI_Bin_Size1), bin size of four (MI_Bin_Size4) and a bin size of eight (MI_Bin_Size8). We also used Scotts rule (MI_Scott) and Freedman and Dianconis rule (MI_Freedman) for the estimation of mutual information for the experiment.



(i) (ii) (iii) (iv)

Figure 4.14 Multitemporal JERS-1 SAR training data set taken on (i) February 1993 (ii) April 1993 (iii) July 1993 (iv) August 1999

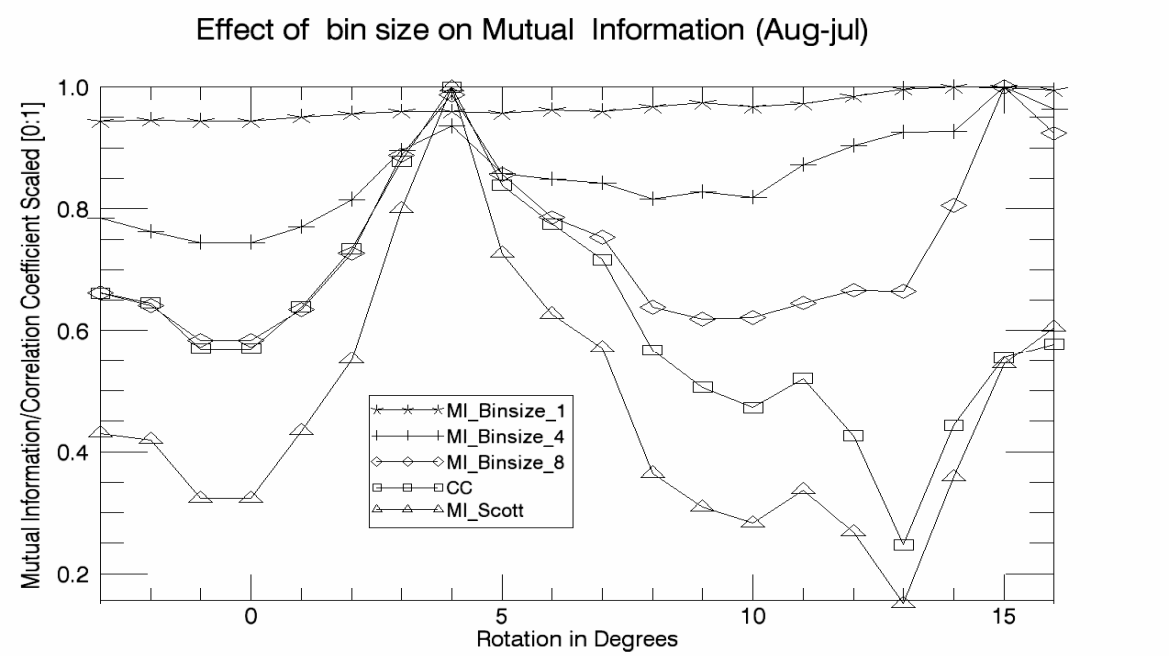


Figure 4.15 Mutual information curves with different bin sizes and CC curve for figure 4.14 (iv) (reference image) and (iii) (slave image) scaled to range [0,1]. MI_Binsize_1 and MI_Binsize_4 failed to determine the correct registration point.The true registration point for rotation is 4°

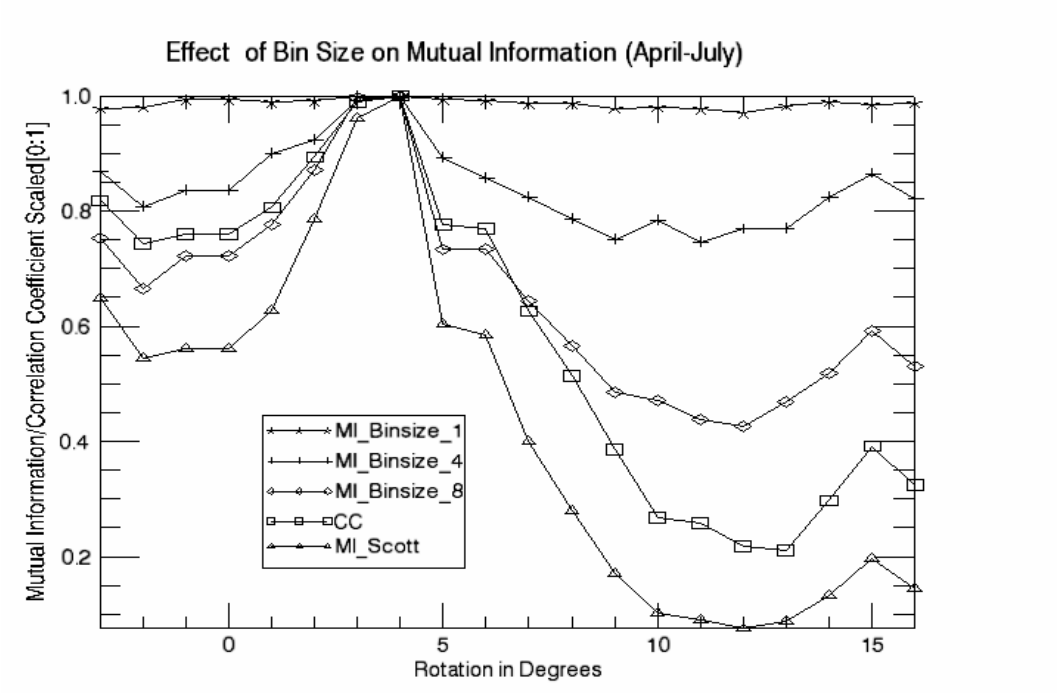


Figure 4.16 Mutual information curves with different bin sizes and CC curve for figure (ii) (reference image) and (iii) (input image) scaled to range [0,1]. All the selected bin sizes used to estimate MI correctly determined the registration point. The correct registration point for rotation is 4°

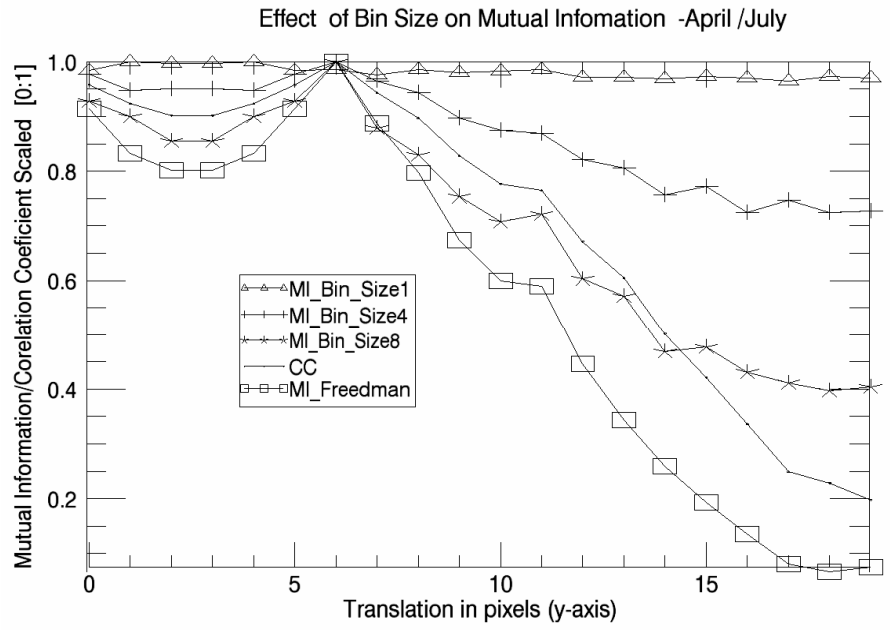


Figure 4.17 Mutual information curves with different bin sizes and CC curve for figure 4.14 (ii) (reference image) and (iii) (input image) scaled to range [0,1]. MI_Binsize_1 failed to detect the correct registration point. The true translation on the y-axis 6 pixels.

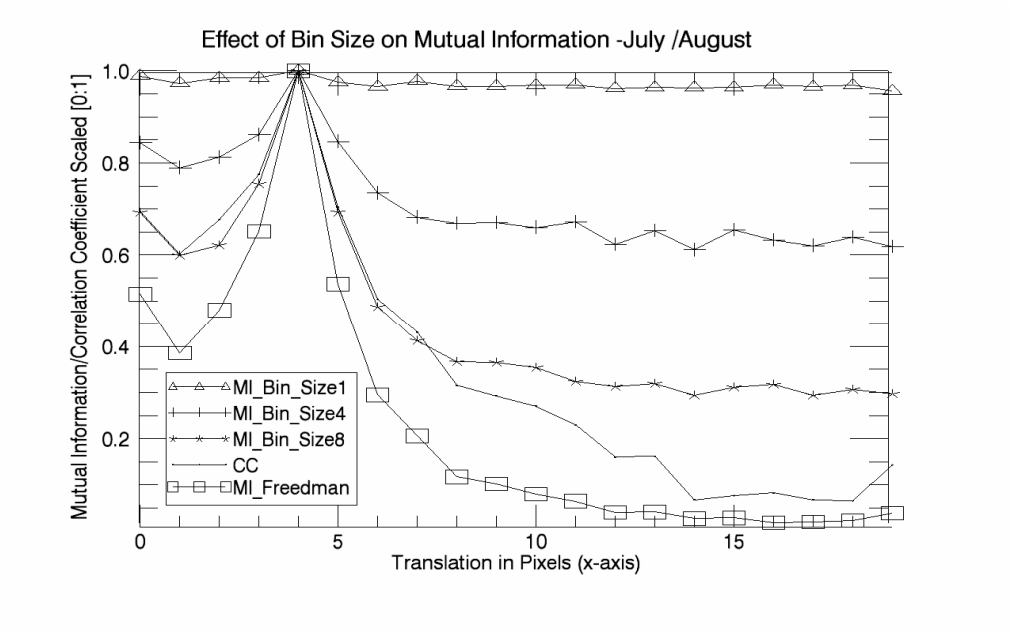


Figure 4.18 Mutual information with different bin sizes and the correlation curve for images in figure 4.14 (iii) (reference image) and (iv) (input image scaled to range [0, 1]. All the selected bin sizes used to estimate MI, correctly determined the registration point. The true translation on x-axis is 4 pixels

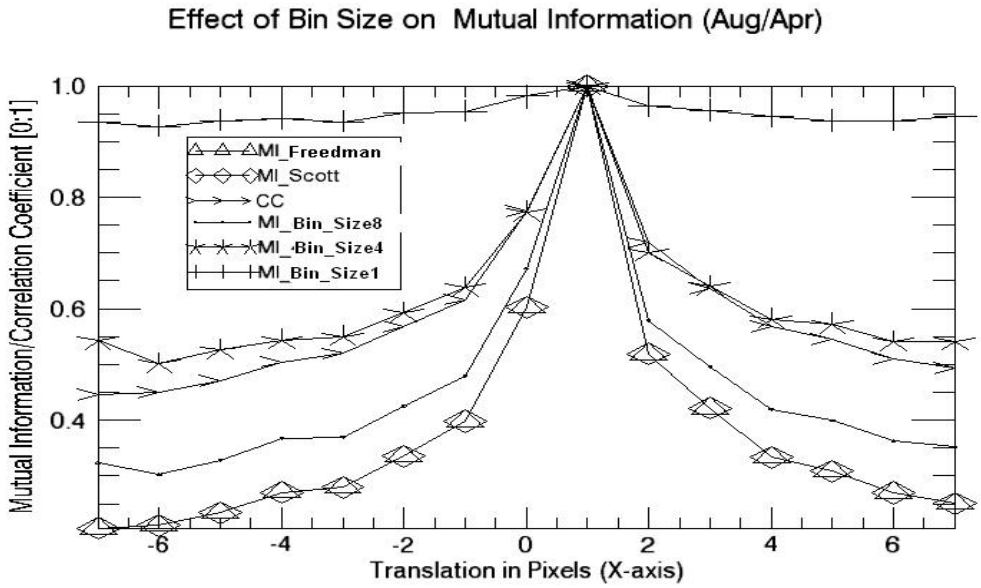


Figure 4.19 Mutual information with different bin sizes and the correlation curve for images in figure 4.14 (iv) (reference image) and (ii) (input image scaled to range [0, 1]. All the selected bin sizes used to estimate MI, correctly determined the registration point. The translation on the x-axis is 2 pixels.

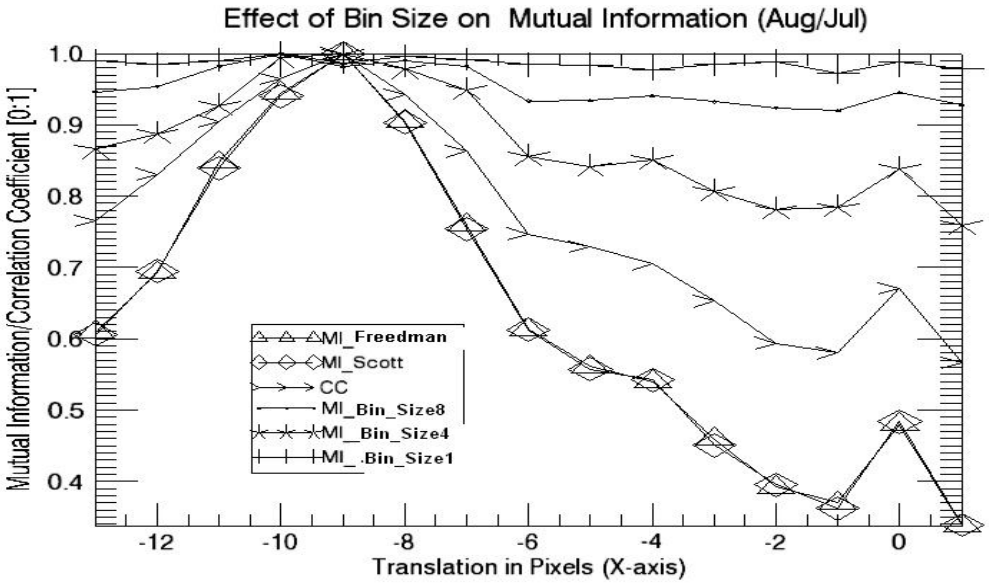


Figure 4.20 Mutual information with different bin sizes and correlation curves for figure 4.14 (iv) (reference image) (iii) (input image) scaled to range [0, 1]. MI_Bin_Size1 failed to determine the correct registration point. The correct registration point on the x-axis is -9 pixels

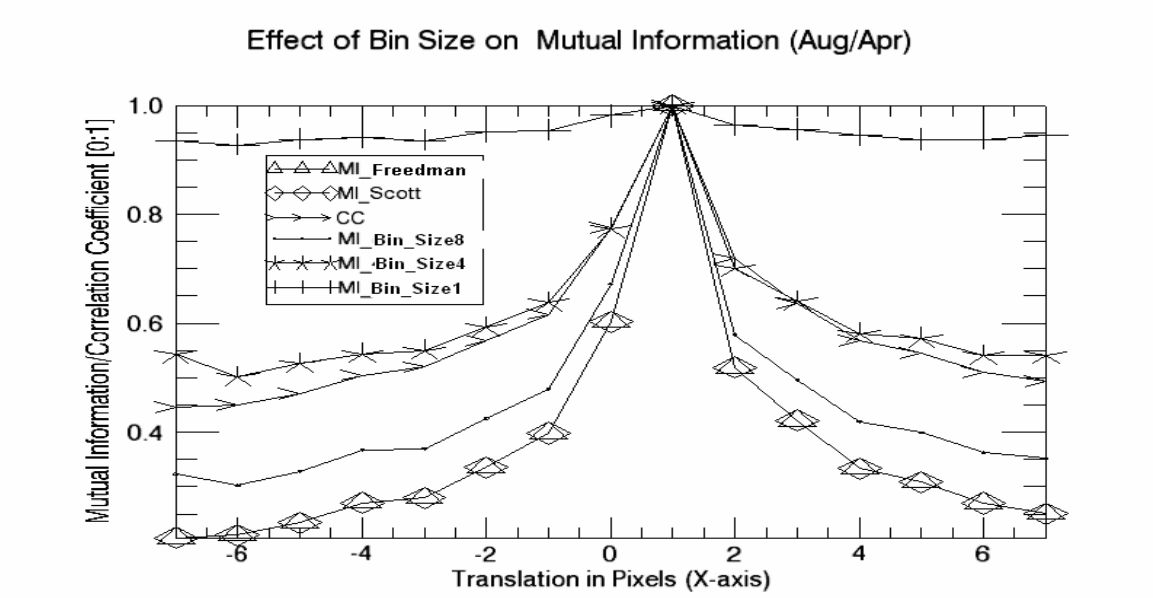


Figure 4.21 Mutual information with different bin sizes and correlation curve for figure 4.14 (i) (reference image) and (ii) (input image) scaled to range [0, 1]. All the selected bin sizes used to estimate MI correctly determined the registration point. The correct registration point on the x-axis is 2 pixels.

4.9.3 Summary of results and findings

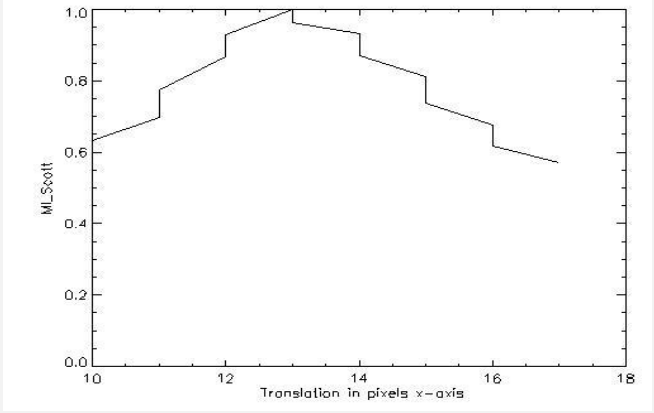
The results show that using Freedman and Diaconis's rule and Scotts rule does not only give more robust result but also generally smoother MI curves. Smoother curves are very desirable when using gradient based algorithms as search space strategies. MI_Bin_Size1 fail to give the correct registration point in figures 4.20, 4.17, and 4.15. In figure 4.17, MI_Bin_Size4 also failed to give the correct registration point. These cases where MI_Bin_Size1 and MI_Bin_Size4 failed to detect the correct registration point can be explained by the fact that some detail information in images may be noise. Generally there was not much difference between the sharpness of MI_Scott and MI_Freedman. MI_Freedman and MI_Scott were consistently sharper than correlation coefficient (CC) and showed the correct registration point in all cases of the experiment.

4.9.4 Subpixel registration

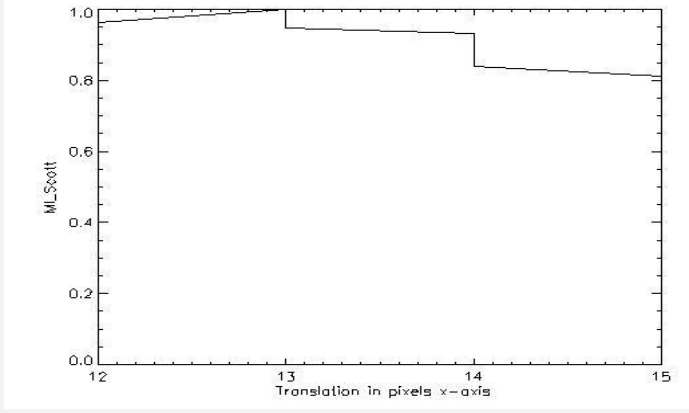
Accurate registration of image pairs is critical in a number of image processing tasks; including image splicing, change detection, image co-addition, and super-resolution processing. In this subsection, we perform some experiments with the images in figure 4.14. We registered image pairs 4.14 (i), 4.14 (ii) (images taken in February 1999 and April 1999 respectively), and 4.14 (iii), 4.14(iv) (image taken in July 1999 and August 1999) to an accuracy of $1/16^{th}$ of a pixel. In the experiment we used MI_Scott to find the registration point. We also used cubic convolution for the interpolation of the images.

The registration points with respect to subpixel accuracy are shown in Table 4.1. The registration curves for MI_Scott are also shown figures 4.22 and 4.23.

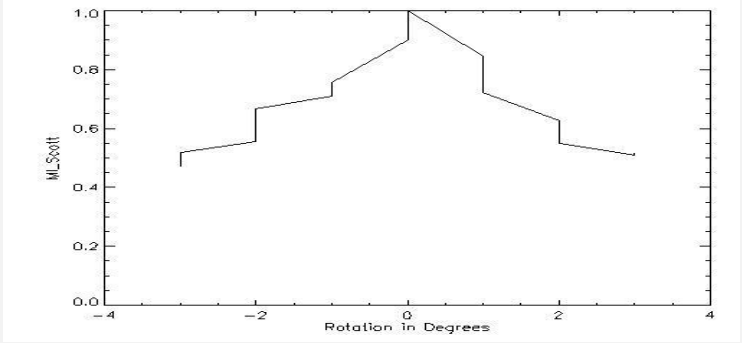
MI_Scott vs. Translation (x-axis) for Feb/April image pair to half pixel accuracy on the x-axis (i)



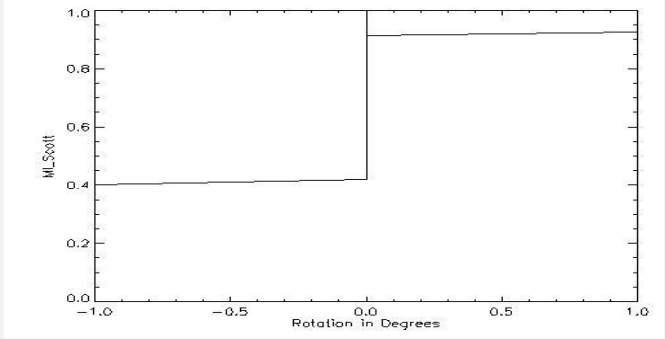
MI_Scott vs. Translation (x-axis) for Feb/April image pair to 1/4th pixel accuracy on the x-axis (iv)



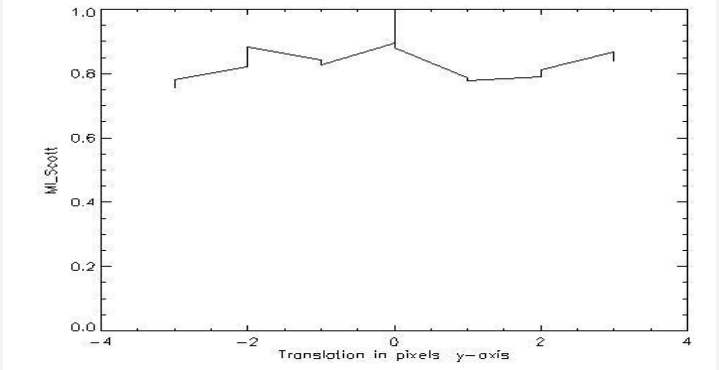
MI_Scott vs. Rotation for Feb/April image pair to 1/4th pixel accuracy (ii)



MI_Scott vs. Rotation for Feb/April image pair to 1/16th pixel accuracy (v)



MI_Scott vs. Translation (y-axis) for Feb/April image pair to Half pixel accuracy on the x-axis (iii)



MI_Scott vs. Translation (y-axis) for Feb/April image pair to 1/4th pixel accuracy on the x-axis (vi)

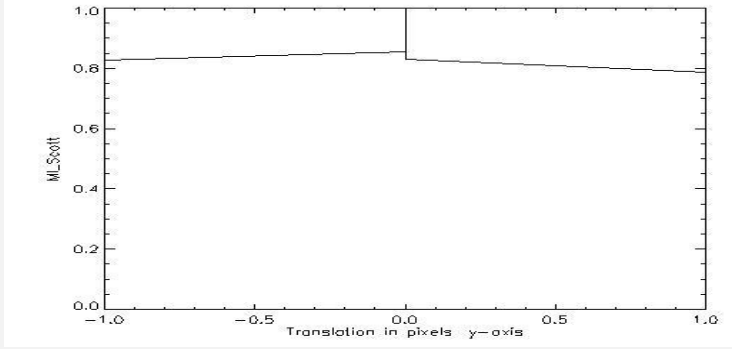
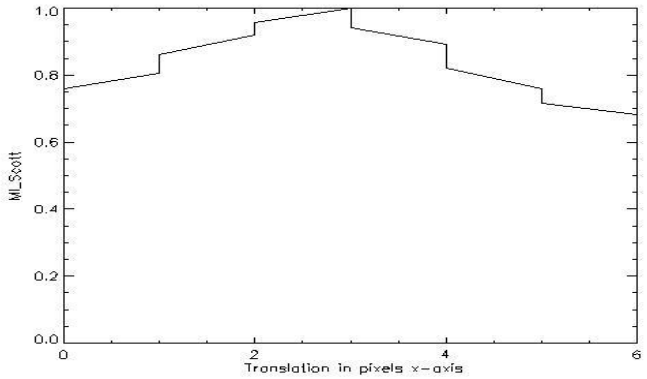
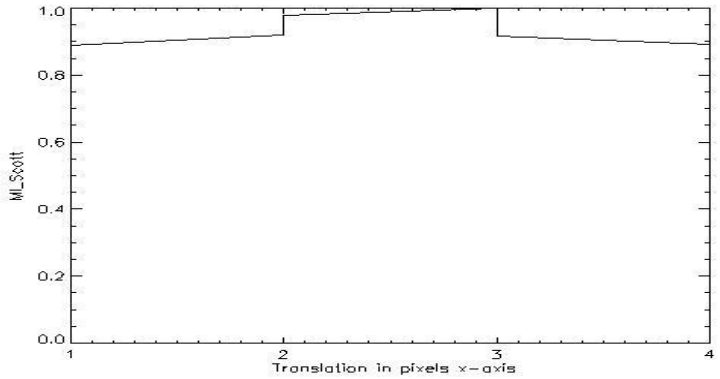


Figure 4.22 Registration curves for February/April image pair to an accuracy of 1/4th pixel (left column) and to an accuracy of 1/16th pixel (right column). The registration curves flatten with increasing accuracy of the registration point. The roughness of MI_Scott curves may be due to the effects of interpolation.

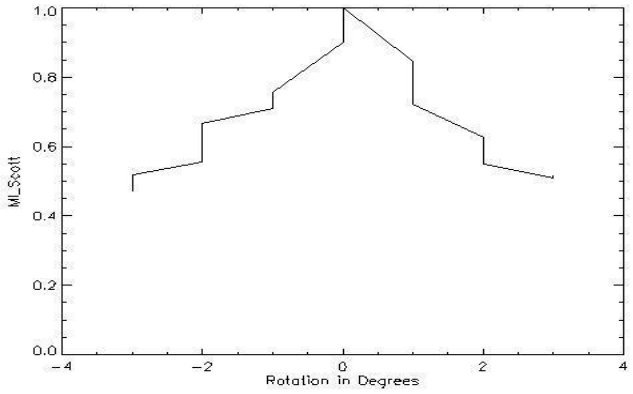
MI_Scott vs. Translation (x-axis) for Jul/Aug image pair to ½ pixel accuracy on the x-axis (i)



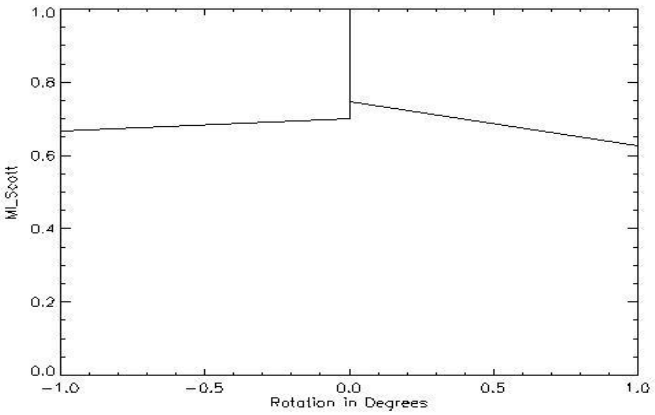
MI_Scott vs. Translation (x-axis) for Feb/April image pair to Half pixel accuracy on the x-axis (i)



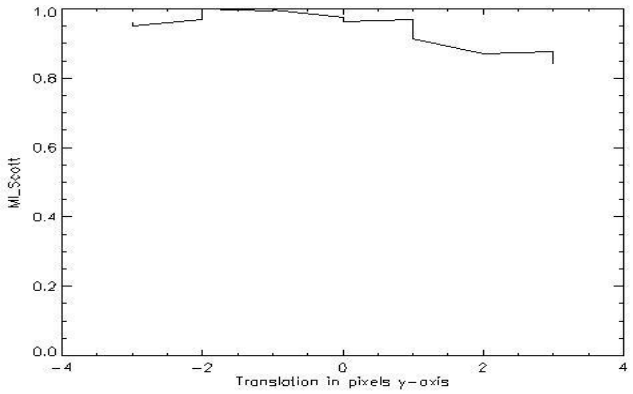
MI_Scott vs. Rotation for Jul/Aug image pair to 1/4th pixel accuracy (ii)



MI_Scott vs. Rotation for Jul/Aug image pair to 1/16th pixel accuracy (v)



MI_Scott vs.Translation (y-axis) for Jul/Aug image pair to 1/2 pixel accuracy on the y-axis (iii)



MI_Scott vs. Translation (y-axis) for Jul/Aug image pair to1/4th pixel accuracy on the y-axis (vi)

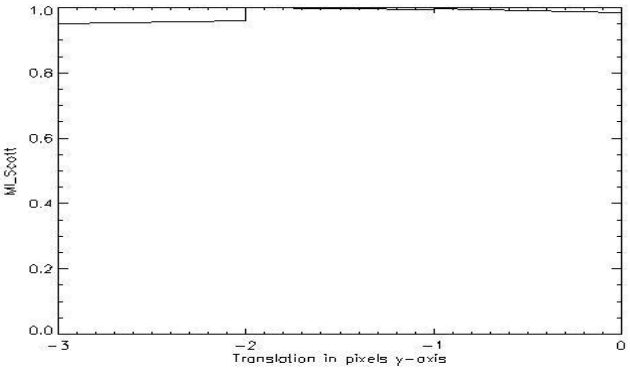


Figure 4.23 Registration curves for February/April image pair to an accuracy of 1/4th pixel (left column) and to an accuracy of 1/16th pixel (right column The registration curves flatten with increasing accuracy of the registration point. The roughness of MI_Scott curves may be due to the effects of interpolation.

Table 4.1 Registration points of image pairs to 1/16th accuracy

Image pairs	Registration points with 1/16th pixel accuracy		
	Tx	Ty	R
Febuary/ April	13.00	0.00	0.00
July/August	3.00	-2.00	0.00

Table 4.1 shows the correct registration points for JERS1 images in figure 4.14. Image taken in April (figure 4.14(ii)) is translated 13.00 pixels to the right on the y-axis as compared to image take in February (figure 4.14(i)). There is no translation on the y-axis or rotation. Compared to image taken in July (figure 4.14 (iii), the image taken in August (figure 4.14 (iv)) is translated 3.00 pixels and -2.00 pixels on the x-axis and y-axis respectively. There is no rotation.

4.9.5 Sensitivity of on mutual information of correlation coefficient to noise

In this section, we compare the sensitivity of MI and CC to Gaussian noise. The reference image used for this experiment is given in figure 3.4. Input images are generated by transforming the image, [Tx,Ty,θ] = [3,3,0], and adding different amounts of Gaussian noise ranging from a SNR of 30 dB to 0 dB, or speckle noise with contrast ratio γ, of 0.71. The Gaussian noise and speckle noise are additive and multiplicative respectively.

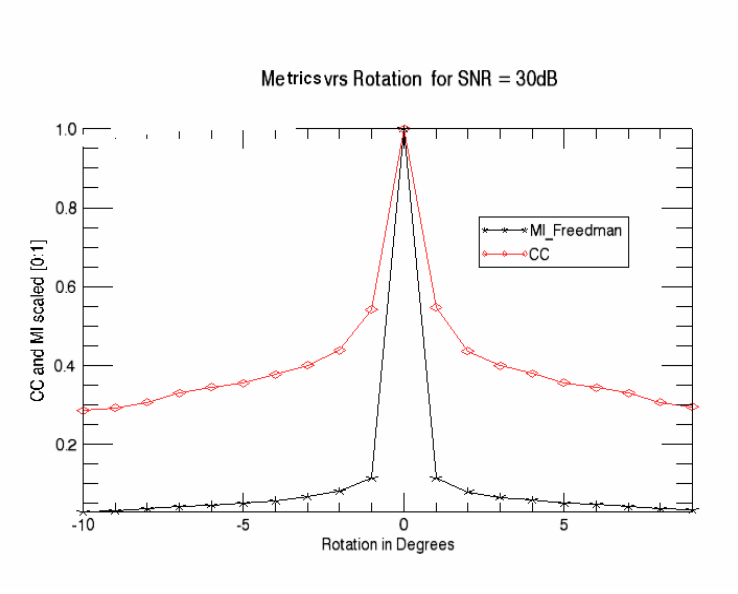
Signal to noise ratio can also be expressed as γ_{SNR}, where γ_{SNR} is the influence of SNR on interferometric coherence. γ_{SNR} is given by

$$\gamma_{SNR} = \frac{1}{\sqrt{(1+N_1/S_1)(1+N_2/S_2)}} \qquad 4.57$$

where S_1 and S_2 are noise free signal intensities of images and N_1 and N_2 are intensities of noise. γ is defined by

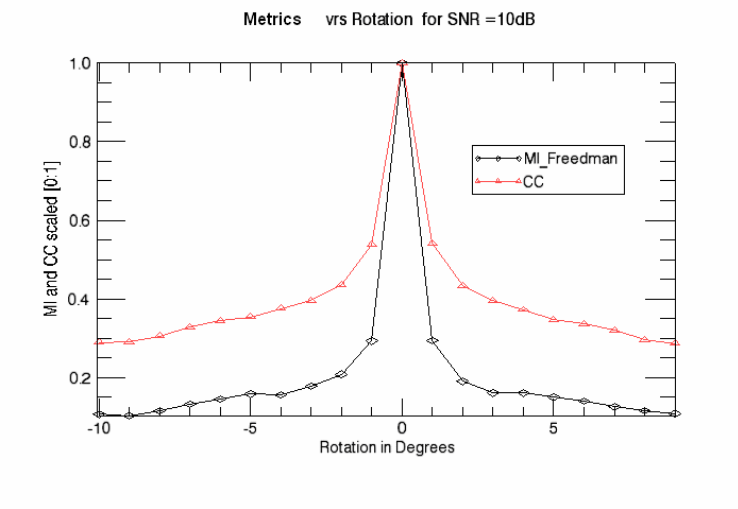
$$\gamma = \frac{S_d}{M} \qquad 4.58$$

where S_d and M are standard deviation and mean of noise respectively.



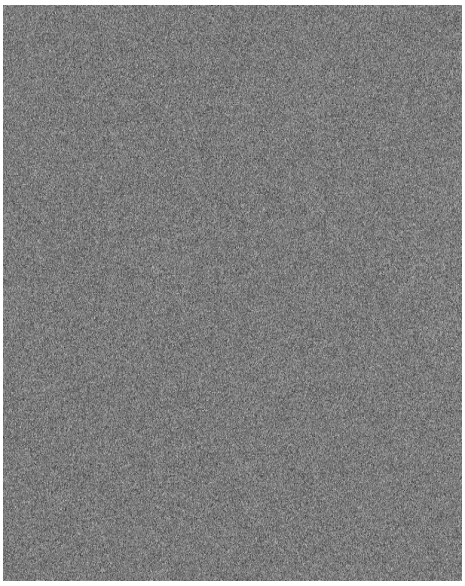
(i) (ii)

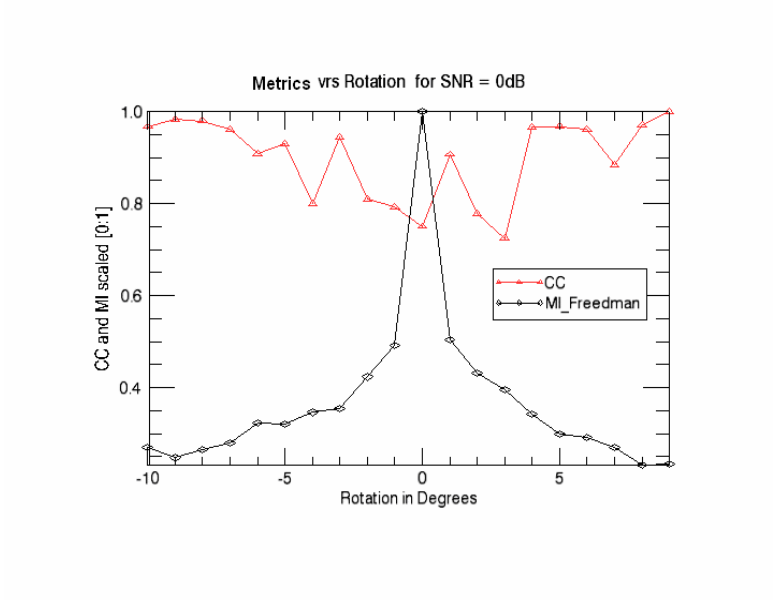
Figure 4.24 Graph showing registration curves of MI and CC (ii) and reference image with SNR of 30 dB (i) (additive noise). Results show MI is sharper than correlation coefficient.



(i) (ii)

Figure 4.25 Graph showing registration curves of MI and CC (ii) and reference image with SNR of 10 dB (additive noise) (i). Sharpness of both curves decreases with increasing noise. Results show that Mutual information remain sharper than CC.

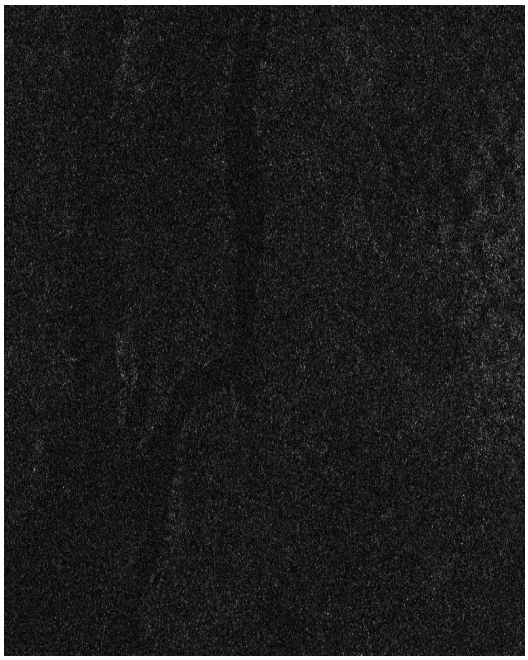


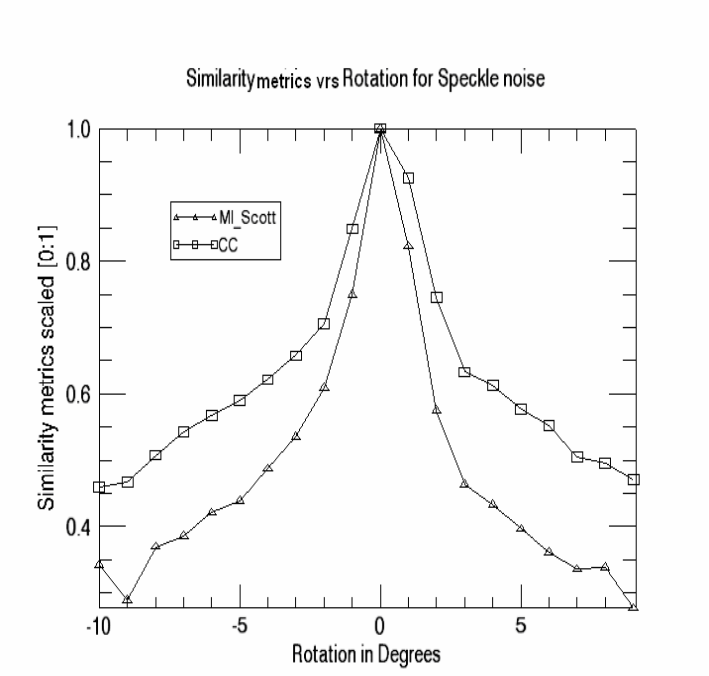


(i) (ii)

Figure 4.26 Graph showing registration curves of MI and CC (i) and reference image with SNR of 0 dB (additive noise) (ii). Correlation coefficient failed to detect the correct registration point. MI is more robust than CC and correctly detect the registration point.

The results indicate that MI produces sharper peaks than CC not only for Gaussian noise but also for speckle noise. MI determined the correct registration values for the selected noise range. For a SNR of 0 dB CC failed to show the correct registration value. In this test, we used Freedman and Diaconis rule or Scott's rule to select the bin size. The same test was performed with a fixed bin size of 2 and MI failed to determine the correct registration value at SNR 4 dB. Given the stochastic nature of speckle noise, we must describe this noise statistically. The statistics used to describe the speckle noise here is drawn from literature from synthetic aperture radar [2].





(i) (ii)

Figure 4.27 Graph showing registration curves of MI and CC (ii) and reference image with speckle noise (multiplicative noise) of contrast ratio 0.7 (ii). MI is sharper than CC

4.9.6 Robustness and accuracy of MI estimation methods

In this section, we performed experiments to compare the robustness and accuracy of MI_Scott, MI_Freedman, MI_4 and MI_Gauss (mutual information estimated by using gaussianity rule to select bin size). The five reference images used in the experiment are SAR1, SAR2, SAR3, SAR4, and Lena. Gaussian noise is added to image to input (slave) images with a SNR of 20 dB. The images are divided into 64 subimages of size 96 x 64.

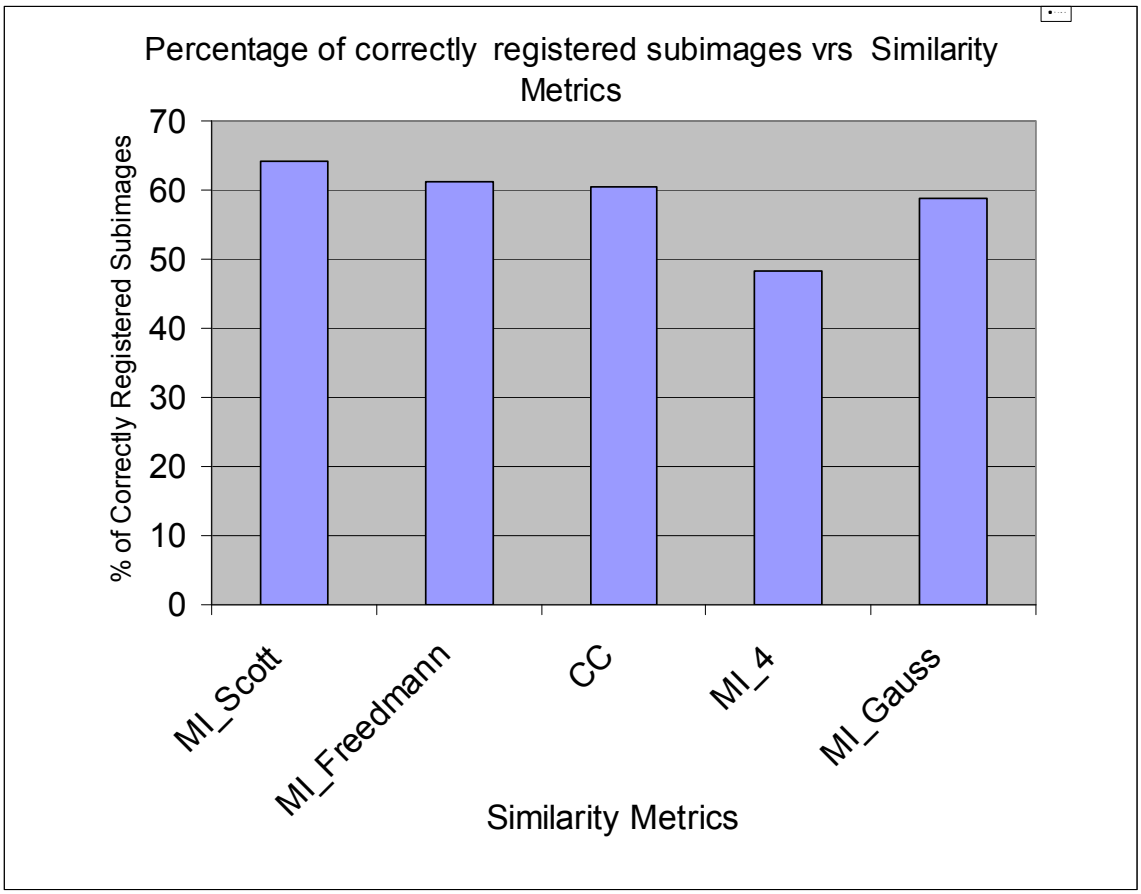


Figure 4.28 Graph showing the percentage of correctly registered images

The percentage of correctly registered of all 64 subimages using MI_Scott, MI_Freedman, CC, MI_4 and MI_Gauss are shown in figure 4.28. The overall percentage of correctly registered subimages for MI_Scott, MI_Freedman, CC, MI_4 and MI_Gauss are 64.25%, 61.25%, 60.5%, 48.3% and 58.8% respectively. The results show that, the two proposed methods, namely MI_Scott and MI_Freedman can be slightly more reliable than MI_Gauss and CC. MI_4 performed relatively poorly.

Table 4.2 shows the mean RMS error of all subimages using the different similarity metrics. The mean RMS error of MI_4 is the highest. The RMS

error for MI_Friedman and MI_Scott are slightly lower than MI_Gauss and CC.

Table 4.2 Similarity metrics and the corresponding Mean RMS error

Similarity Metric	Mean RMS error
MI_Scott	0,389774
MI_Friedman	0,395304
CC	0,402615
MI_4	0,432204
MI_Gauss	0,404654

4.9.7 Computational cost of MI and CC

One of the important parameters used in evaluating similarity metrics is the computational cost. In this subsection we compare the cost of computing MI and CC. The computational cost of spatial correlation coefficient of image pairs is $O(N^2)$, where N is the number of pixels in each image. The cost of computing MI of image pairs depends on the number of bins used to estimate MI. If both image have the same number of pixels, N, then the computational cost of computing the histogram is $O(N)$. The computational cost relative to the number of histogram bins, K, used in the computation is $O(K^2)$.

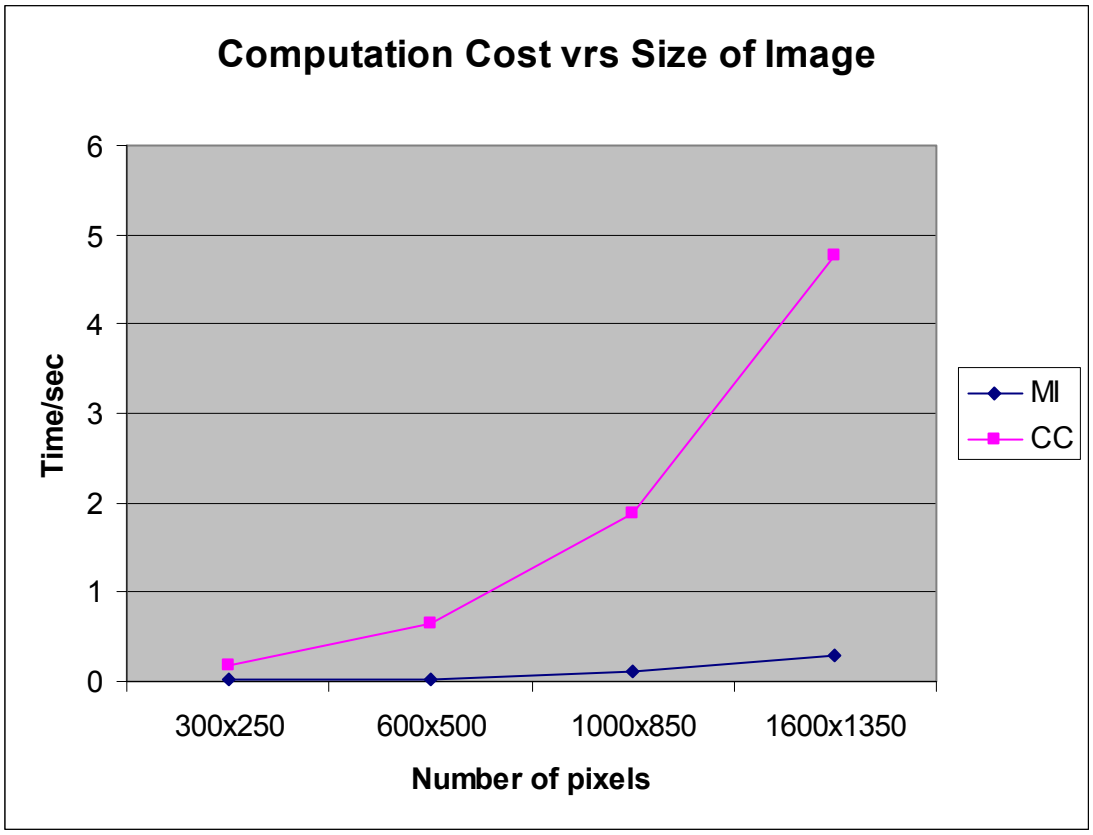


Figure 4.29 Graph showing the time used to compute MI and CC for different image sizes.

In order to compare the computation cost of MI and CC, we measure the time for computing MI and CC for different image sizes using a computer with a CPU of 1.33 GHz. The results in figure 4.29 show that as the size of images increases the computational cost of MI gets much lower than the computational cost of CC. Whereas mutual information increases linearly with increasing image size, correlation coefficient increases quadratically with increasing image size. The results correspond to the big-O notation. The correlation coefficient was estimated in the spatial domain.

4.10 Gradient-based search space strategy

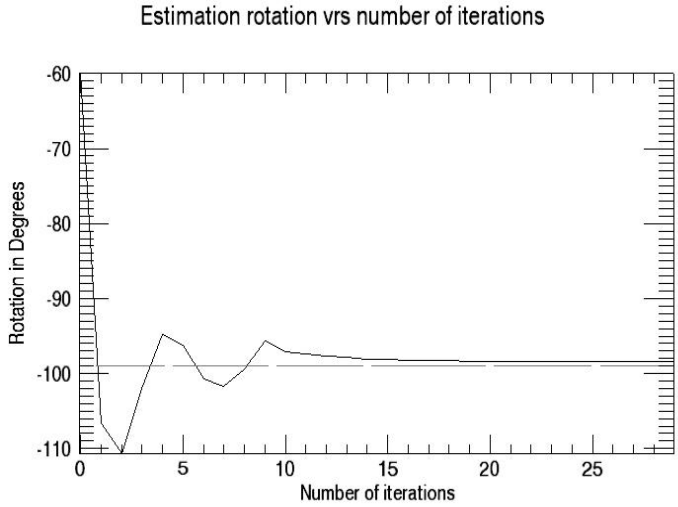
The search for an optimum transformation by exhaustive search is computationally very expensive. In this section, we use a gradient-based

algorithm based on Robbins-Monro stochastic approximation algorithm. It can also be considered as a generalization of steepest decent.

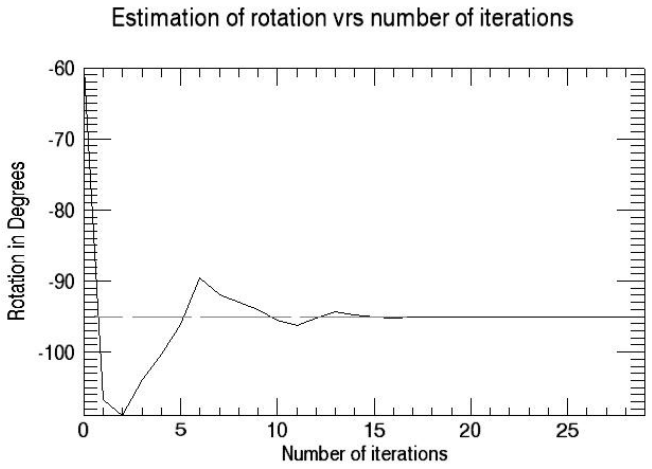
Let MI be the objective function to be optimized. In our experiment, we consider a parameter search space of one dimension, rotation, which is represented by θ. For each iteration, the gradient g_k is based on two objective function measurements. At iteration, k, the update law for rotation (θ) is,

$$\theta_{k+1} = \theta_k + a_k g_k \qquad 4.59$$

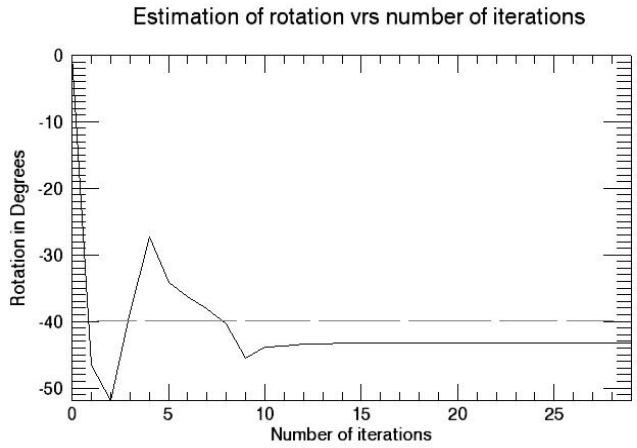
where a_k is the learning rate and satisfies (i) $a_k > 0$,(ii)$\sum a_k = \infty$ and (iii)$\sum (a_k)^2 < \infty$. The update is blocked if the difference the new MI and current MI is less than 0.2 Figures 4.27(a), (b), (c) and (d) show the results of the registration of slave images rotated 99°,95°,40° and 30° respectively. The initial guess for figures 4.27 (a) and (b) is -60° and for figures 4.27 (c) and (a) is 0°. The algorithm either correctly shows or is near the registration point and converges at a fast rate. The algorithm could not, however, overcome some local maxima in some cases.



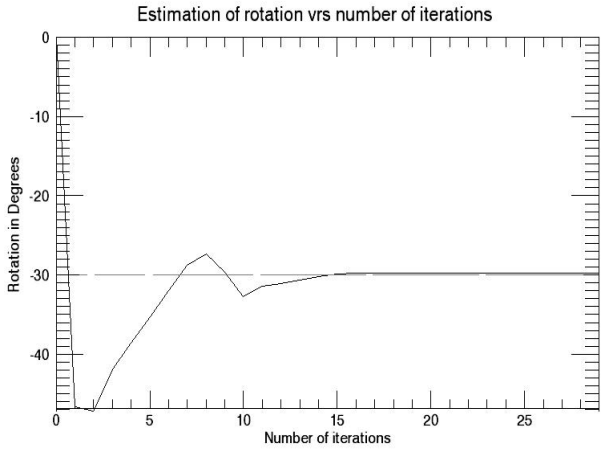
(a)



(b)



(c)



(d)

Figure 4.27 Curves showing convergence and registration points for input images rotated for (a) 99°, (b) 95°,(c) 40° and (d) 30°. The dashed line show the correct registration point

4.11 Summary

We have shown the effect of noise on mutual information and compare the results with correlation coefficient. We have also proposed two methods which formalize the selection of bin size in the estimation of mutual information. The results indicate that using the proposed methods can produce more reliable and accurate results than the fixed bin size method. We also compared mutual information to other F-Information measures with mutual information (figures 4.1, 4.2, 4.3). A gradient-based search space strategy is also presented.

CHAPTER 5

IMAGE REGISTRATION USING A COMBINATION OF MUTUAL INFORMATION AND SPATIAL INFORMATION

Mutual information is a basic concept in information theory. It is applied in the context of image registration to measure the amount information that one image contains about the other. The maximization of the mutual information criterion postulates that mutual information is maximal when images are correctly registered. Mutual information has demonstrated to be a very general and powerful similarity metric that can be applied automatically and very reliably, without prior pre-processing, on a large variety of application.

In recent years, mutual information has emerged as a popular similarity metric in the registration of images. Unfortunately, it takes into account only the relationships between corresponding individual pixels and not those of each pixel's neighbourhood. It ignores the spatial information contained in the images such as edges and corners that might be useful in the matching of images. Thus, it is essential to consider both quantitative and qualitative information in the registration of images which refer to the probability of occurrence and the significance of pixels respectively.

Recently some researchers have incorporated spatial information in the registration of images by multiplying mutual information with a gradient term that measures the coincidence of pixels in both images with strong gradients of similar orientation [42].

Rueckert et al. employ higher order mutual information, using the co-occurrence matrices of neighbouring pixels' intensities. They showed that second-order

mutual-information outperformed first-order mutual information, since it incorporated spatial information into the registration process [43]. Hero et al also use the theory of entropic spanning graphs to directly estimate the entropy of a distribution [44]. This has the advantage of by-passing density estimation. Furthermore, even for large dimensions this algorithm can be easily implemented, whereas histogram methods cannot.

In this chapter, we propose a new similarity metric called spatial mutual information (SMI) which combines mutual information and a weighting function based on image gradient, image variance, and image entropy of local regions. Salient pixels in the regions with high gradient and variance contribute more in the estimation of mutual information of images being registered. We compare the robustness and accuracy to mutual information.

5.1 Quantitative and qualitative information

Shannon's entropy (equation 3.4) determines a quantitative measure of information. It considers all events as abstract random events and neglects the quality of the event. Belis and Gaiasu proposed a qualitative-quantitative measure [45]. The quantitative part is related to the probability of occurrence and the qualitative part is related to the utility of fulfilment of the goal.

Let X= (x_1, x_2, … , x_n) be a discrete random variable. Let P = (p_1,p_2, …, p_n) be the probability of occurrences of the random variable, and U = (u_1,u_2,..., u_n) , $u_i \geq 0$, be the utility of the random variable.

The quantitative and qualitative information is defined as

$$QH(X;U) = \sum_{i=1}^{n} -u_i p_i \log p_i \qquad 5.1$$

If all the utilities are equal the equation becomes Shannon's entropy equation.

5.2 Quantitative and a qualitative relative information

Let Y = (y_1, y_2….y_n) be a second random variable and Q= (q_1, q_2….q_n) be their probabilities, then the quantitative and qualitative measure of relative information or directed-divergence as suggested by Taneja [46] is given as

$$QRH(X,Y;U) = \sum_{i=1}^{n} u_i p_i \log(p_i / q_i) \qquad 5.2$$

When the utilities are the same or equal, the QRH reduces to the Kullback's measure of directed divergence.

5.3. Spatial mutual information

As indicated above, mutual information is calculated on a pixel to pixel basis, meaning that it takes into account only the relationships between corresponding individual pixels and not those of each pixel in the respective neighborhood. The effects of pixels with respect to the goal are not considered. As a result, much of spatial information inherent in images is not utilised. Pixels in regions such as high gradient which enhance reliable and accurate registration [47] are given the same utility as pixel in low gradient. We define the spatial mutual information as

$$SMI(X,Y) = \sum_{x,y} u(x,y) \cdot p(x,y) \cdot \log \frac{p(x,y)}{p(x) \cdot p(y)} \quad , \qquad 5.3$$

where *p(x)* and *p(y)* are the probabilities of pixels *x* and *y* respectively and p(x,y) is their joint probability and *u(x,y)* is the utility. The term *u(x, y) log p(x,y)* is regarded as the useful information gain in attaining the goal. If *u(x,y) = 1* the equation tends to become equation 2.5, which is mutual information. There are

three methods that can be used to estimate SMI. In our work, we use the histogram method.

Let *Hist (X, Y)* be the histogram entries and M is the number of entries. If *X* and Y are the reference and input images, $Hist_x$ (A) and $Hist_y$ (B) are defined as their histograms, and $Hist_{xy}(X,Y)$ joint histogram, then *SMI* is defined by

$$SMI(X,Y) = \frac{1}{M}\sum_x\sum_y u(X,Y)\cdot Hist_{xy}(X,Y)\cdot\log\left(\frac{M\cdot Hist_{xy}(X,Y)}{Hist_x(X)\cdot Hist(Y)}\right) \quad . 5.4$$

5.4 Utility of pixels

The significance of pixels in image registration is not the same. For example, pixels that lie in the region of intensity changes are more useful. However, the characterization of pixels in image analysis remains to be solved. In this section, we propose a method for the computation of the utility of pixels in image registration.

5.4.1 Computing the utility of pixels

The significance of each pixel is determined by calculating the gradient magnitude, entropy and variance of local regions. It has been shown that regions of high gradient edges, high entropy and high variance in an image enhance the determination of registration points in image registration [47]. The algorithm for estimating utilities proceeds as follows:

1. The image gradient is the measure of the rate of change of image intensities between neighbouring image pixels. The gradient of an in image, *I(x,y)* at a location *(x,y)* is defined by

$$\nabla I_i(x,y) = \begin{bmatrix} g_x(x,y) \\ g_y(x,y) \end{bmatrix}_i = \begin{bmatrix} \dfrac{\partial I(x,y)}{\partial x} \\ \dfrac{\partial I}{\partial y}(x,y) \end{bmatrix}_i . \qquad 5.5$$

There are a number of ways to calculate gradient operators for digital images. We use the Sobel operator. In our work, the gradient magnitude of each local region is also calculated according to the equation

$$\left|\nabla I_i\right| = \left[g_x^2 + g_y^2\right]_i^{1/2} . \qquad 5.6$$

The sum of gradient magnitude, G_i is computed for each local region *i.*

2. The entropy of a local region is the information content in the region. If x_k is the k^{th} pixel in a local region and $P(x_k)$ the probability, then the local entropy of local region *i* is defined as

$$H_i = -\sum_{k=1}^{N} P(x_k) \log P(x_k) \qquad 5.7$$

where *N* is the number of pixels in the local region.

3. The variance of a local region i which is proportional to its bandwidth is defined by

$$V_i = \frac{1}{N} \sum_{k=0}^{N-1} \left(x_k - \chi\right)^2 \qquad 5.8$$

where x_k and χ are the pixel values and mean of pixel values of local area respectively.

4. Since high gradient magnitude, high entropy and high variance are preferred, the significance of a pixel $S_i(x_k)$ is determined by the multiplication of the scaled gradient magnitude, scaled entropy and scaled variance of the local region

$$S_i(x_k) = \frac{G_i \cdot V_i \cdot H_i}{G_{\max} \cdot V_{\max} \cdot H_{\max}} \qquad 5.9$$

where H_{max}, G_{max} and V_{max} are the maximum entropy, maximum gradient magnitude and maximum variance respectively of the local regions in the image. The significance of all pixels in a local region is thus the same.

5. The utility of pixels in SMI

Let $S_m(x_k)$ and $S_s(y_k)$ be the significance of master image and slave image respectively. Then the utility of a pair of pixels during registration is defined as

$$u(x_k, y_k) = S_m(x_k) \cdot S_s(y_k) \qquad 5.10$$

where x_k and x_y are corresponding pixels in master image and slave image respectively. There are various ways to combine the significance of the pixel pairs. We use multiplication to estimate the utility of pixel pairs. We will be investigating other methods of combining the significance of the pixels.

5.5 Experiments

We performed a variety of experiments to evaluate the reliability and accuracy of MI and SMI in image registration. In the first experiment, we evaluated the

robustness of *MI* and *SMI* with respect to noise. In the second experiment, image pairs with known input image transformations are registered with SMI and MI, in order to validate the robustness and accuracy of the similarity metrics.

5.5.1 Robustness of MI and SMI to noise.

In order to evaluate the robustness of *MI* and *SMI*, we used the 2D image shown in figure 5.1 below and added various levels of noise. The size of the image is 572 x 768, and the SNR of input images are 10dB, 15dB, 20dB, and 30dB respectively. The input images are also translated 20 pixels in the x-axis and y-axis respectively. To evaluate the performance of *MI* and *SMI*, we plot the changes of similarity metrics with respect to rotation (we did not first determine the translation registration points in order to test the robustness of MI and SMI respectively).

(i) (ii)

Figure 5.1 Reference image (i) and after adding Gausssian noise of SNR 10dB to image (ii)

The results are shown in figures 5.2, 5.3 and 5.4. The results show that SMI generated peaks sharper than MI. The sharpness of MI also decreased faster

than SMI as noise was increased. Figure.5.5 shows that for SNR of 10dB MI failed to point out the correct registration point. Thus, integrating the utilities in mutual information does not only sharpen the similarity metric but also render more robust results. During the estimation SMI, translation and rotation does not affect the utilities of the individual pixels.

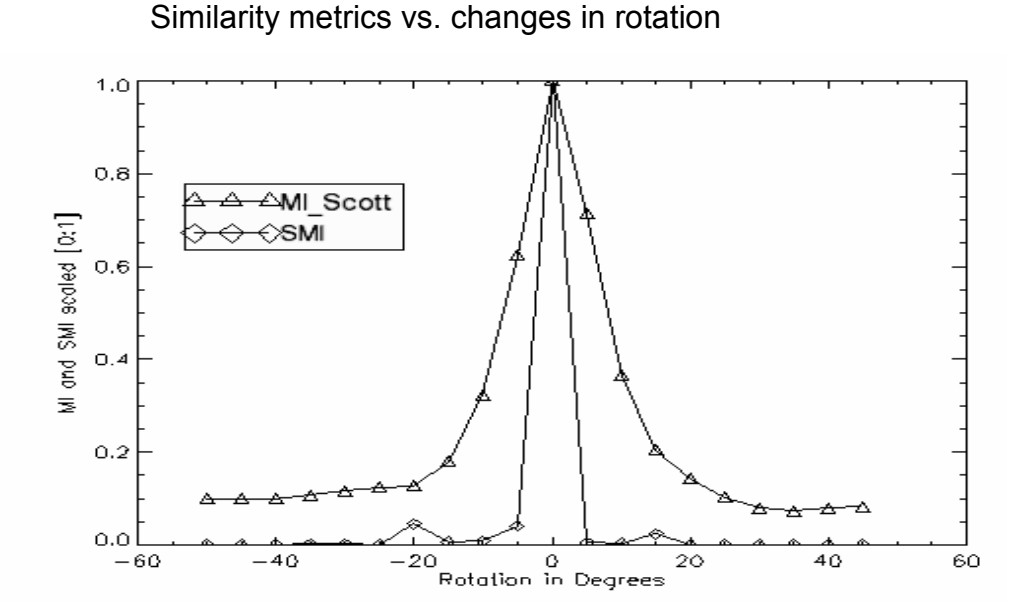


Figure 5.2 Scaled MI and SMI curves for an input image with SNR of 30 dB

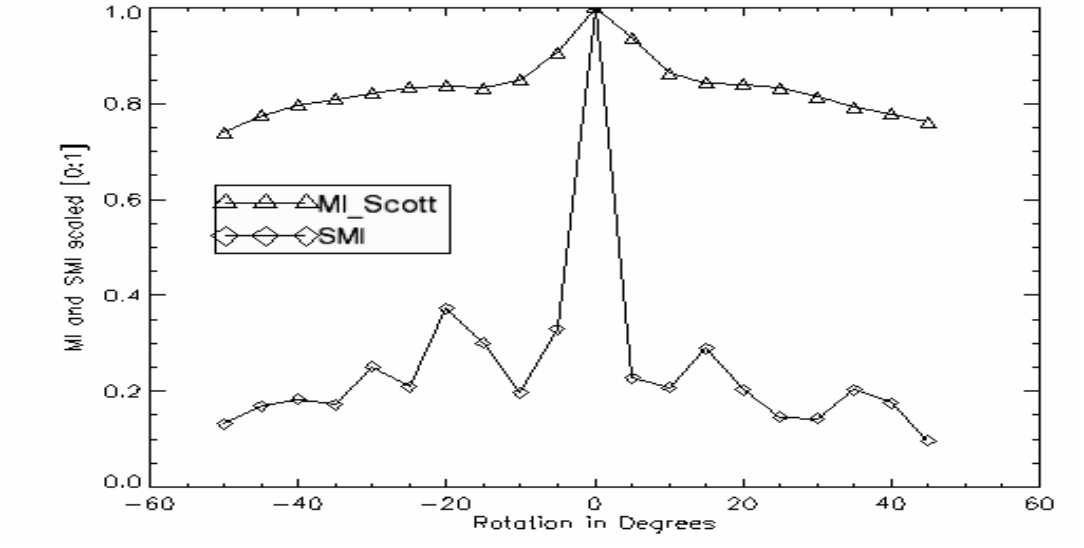


Figure 5.3 Scaled MI and SMI curves for an input image with SNR of 15 dB

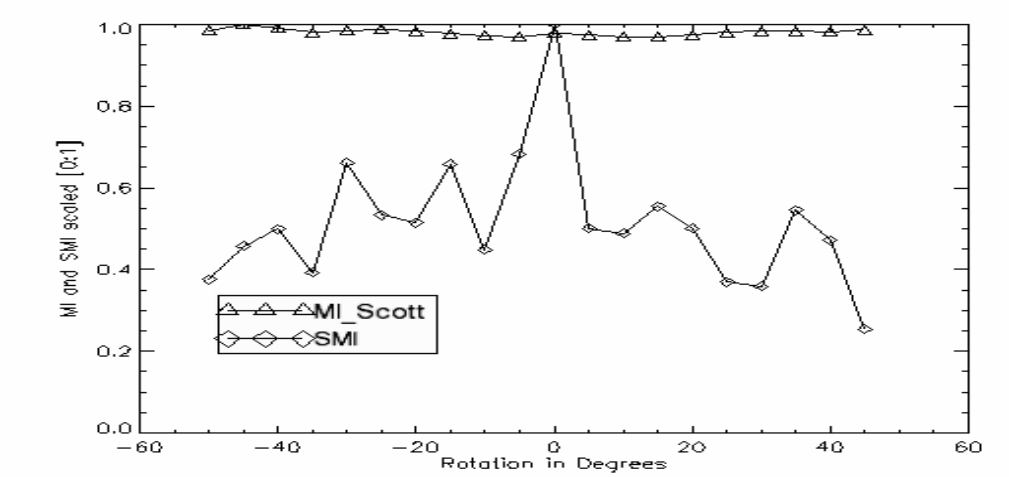


Figure 5.4 Scaled MI and SMI curves for an input image with SNR of 10 dB

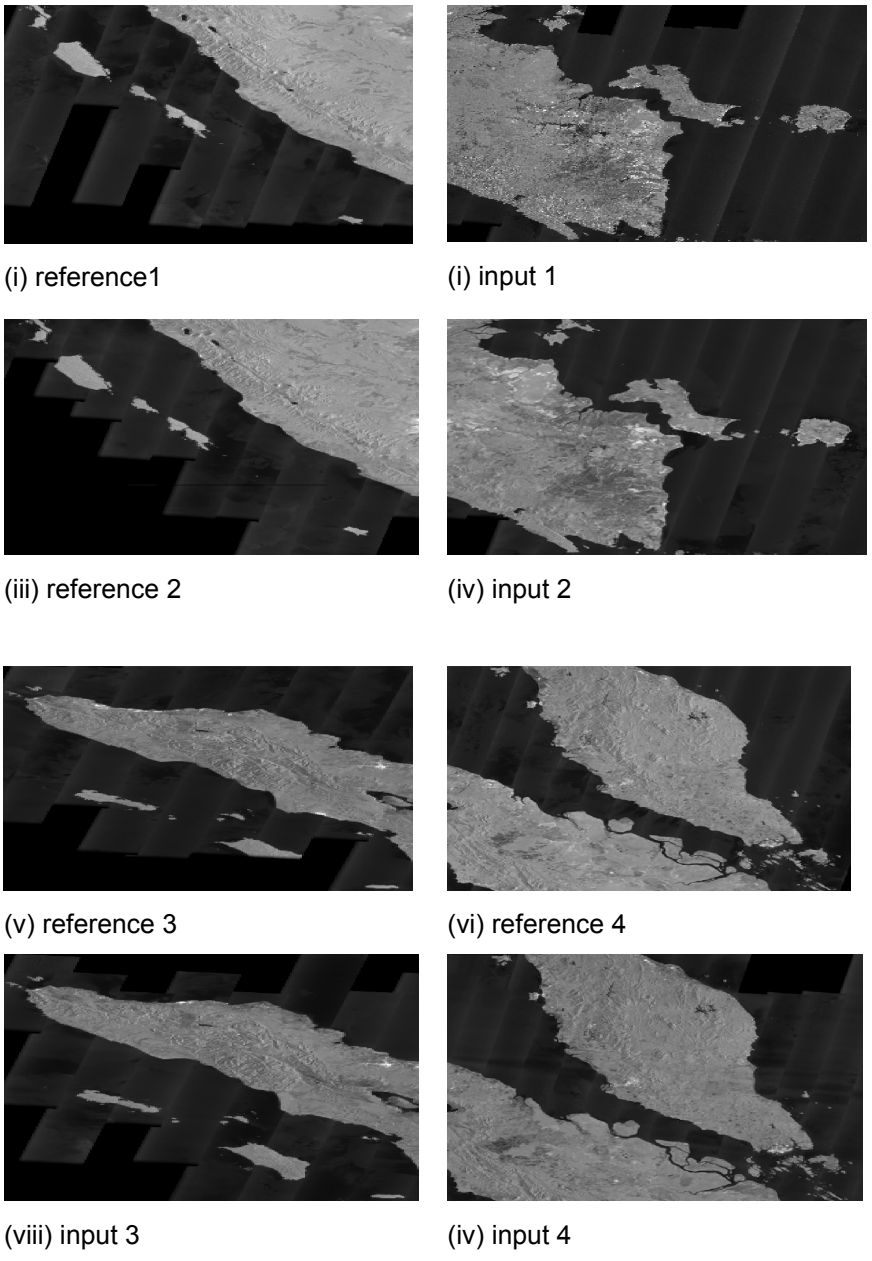
5.5.2 Accuracy and reliability

In the second experiment, a study was conducted to assess the reliability and accuracy of the SMI and MI respectively. We registered 10 SAR images against transformed versions of these images with known transformations. The transformations were composed of random translations and rotations. Each image generated 300 transformed input images. We used MI and SMI to independently register the input image and the reference image. We calculated the percentage of correctly registered images and the root mean square errors (RMSE). Image pairs are correctly registered if the estimated registration points correspond to the ground truth. The percentage of correct registration and the means of RMSE for rotation (R), translation on the x-axis (Tx) and translation on the y-axis (Ty) are shown in figure 5.8 and table 5.1 respectively.

Table 5.1 Comparison of Mean RMSE for SMI and MI

Transformations	Mean RMSE for MI	Mean RMSE SMI
R	0.4234	0.4622
Tx	0.6890	0.6288
Ty	0.7241	0.7322

Figure 5.5 shows multitemporal image pairs from JERS1 SAR taken over South America between May and July 1996. The images are all of size 512 x 1024 pixels. Figure 5.6 shows examples of reference images and their transformed images used for the experiment. The translation and rotation parameters of transformed image varied between 0 to 5 pixels and 0 to 5° respectively



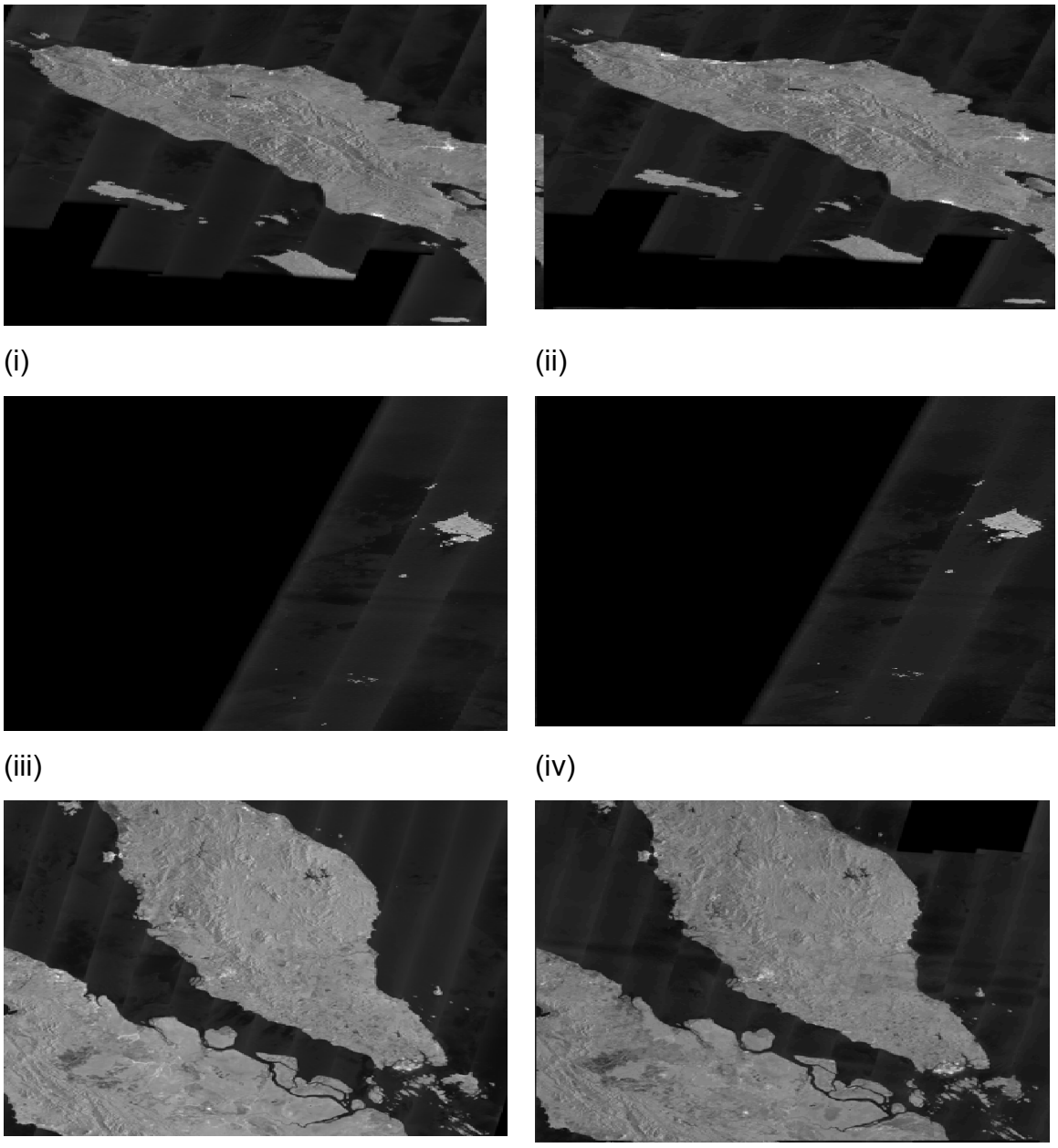
(i) reference1 (i) input 1

(iii) reference 2 (iv) input 2

(v) reference 3 (vi) reference 4 (vii) reference 5

(viii) input 3 (iv) input 4 (iv) input 5

Figure 5.5. Data Set: Series of multitemporal JERS1 SAR taken over South America



(i) (ii)

(iii) (iv)

(v) (vi)

Figure 5.6 Examples of reference images in the left column ((i), (iii), (v)) and their transformed images in the right column ((ii), (iv), (vi)) used for the experiment

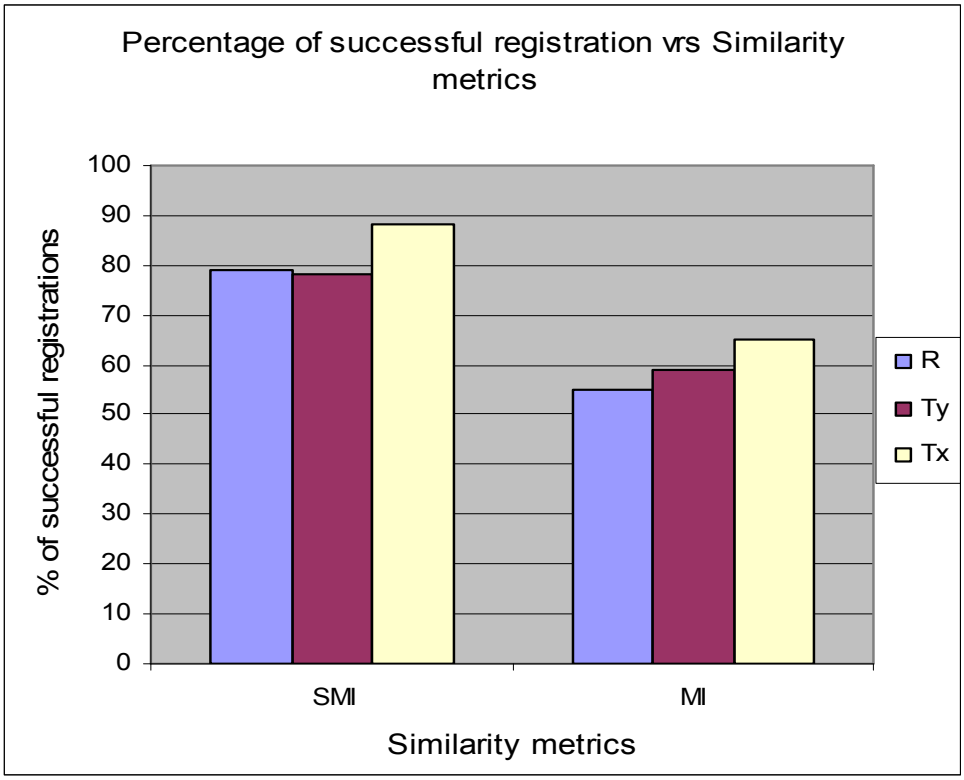


Figure 5.7 Histogram showing successful rates of SMI and MI

From figure 5.7, we can conclude that SMI has a higher success rate than mutual information. The overall percentage of correct registrations for SMI and MI are 82% and 60% respectively. Table 5.1 compares the accuracies of SMI and MI, we can deduce that there is no significant difference between SMI and MI

5.6 Summary

Mutual information as a similarity metric has enjoyed a lot of success in the registration of images. We have presented a new similarity metric, called spatial mutual information (SMI), which not only incorporates the probability of individual pixels, but also the significance or the utility of the pixels. We have shown that SMI can be more reliable in the registration of images. We have demonstrated that SMI can be more robust with respect to noise than MI. In the future we hope to validate SMI on other data sets. We would like also like to apply SMI in other areas such mosiacking and object tracking.

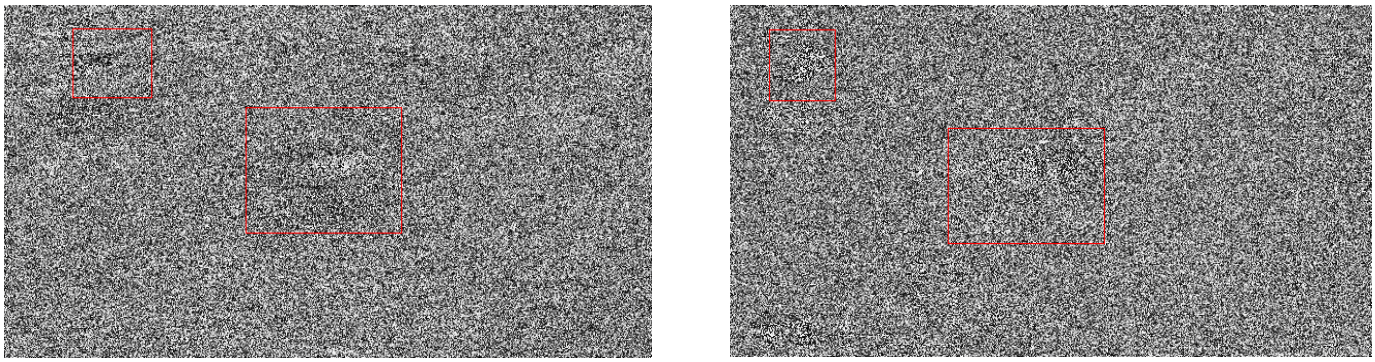
CHAPTER 6

REGISTRATION TEST IMAGES

In this chapter, we show samples of original images, transformed images, difference images, and registered images used in the experiments of the previous chapters.

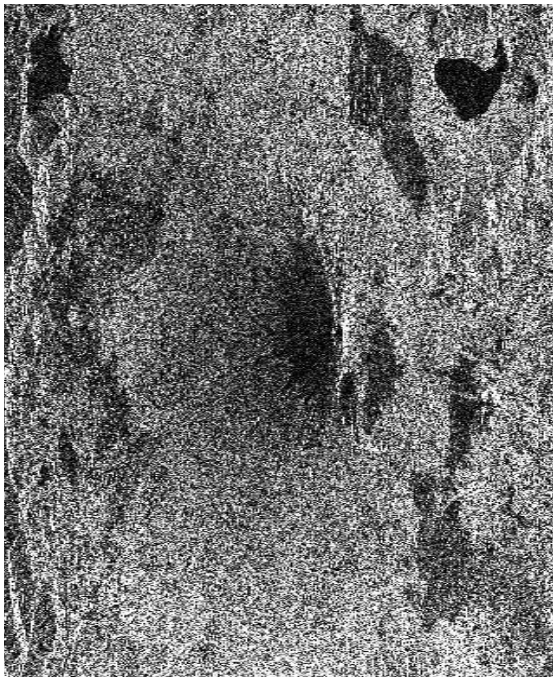
The registered image is the point-by-point correspondence between the reference image and the input image. Image differencing is an image processing technique used to determine changes between images. The difference between two images is calculated by finding the difference between each pixel in each image, and generating an image based on the result. For this technique to work, the two images must first be aligned so that corresponding points coincide. In other words, the image pair must be registered. The difference images shown in this chapter (e.g figures 6.2 (iv)) are scaled.

While reference image and input image may appear the same, they are actually slightly different from each other. The difference graphs in figure 6.1 show how image registration can still detect differences and account for them. The root mean square error before and after image registration are 0.04 and 0.02 respectively.

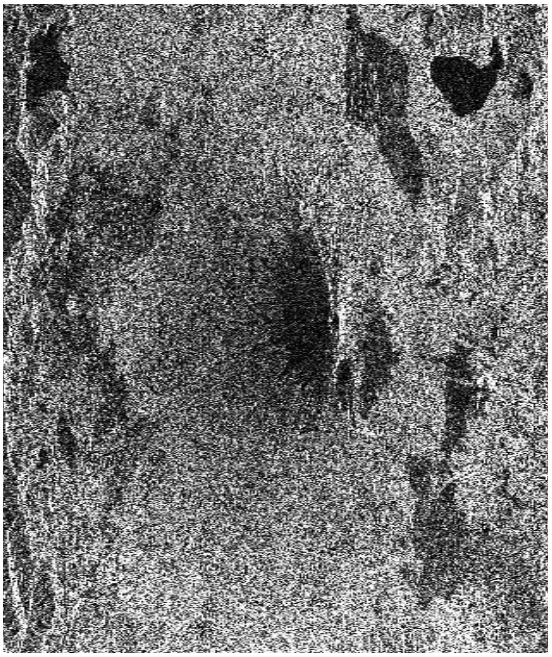


(i) Difference image before image registration (ii) Difference image after image registration

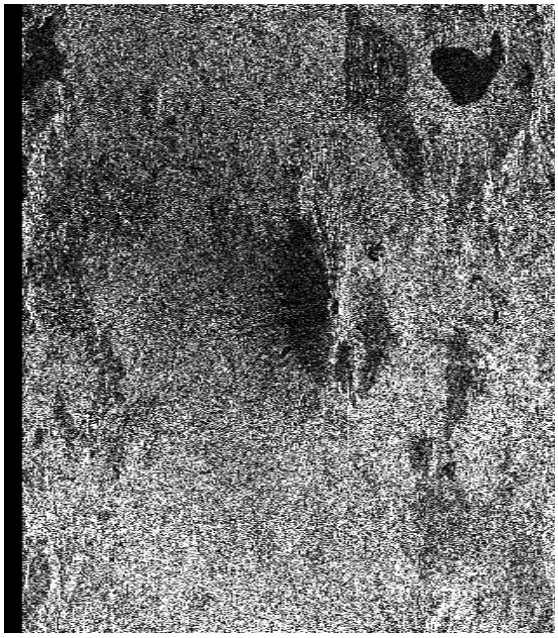
Figure 6.1 Difference graphs of reference and input images in figure 6.2



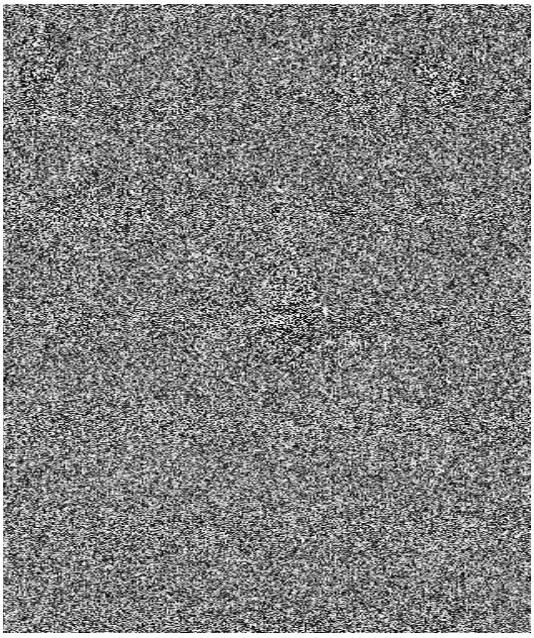
(i) Reference Image



(ii) Input image



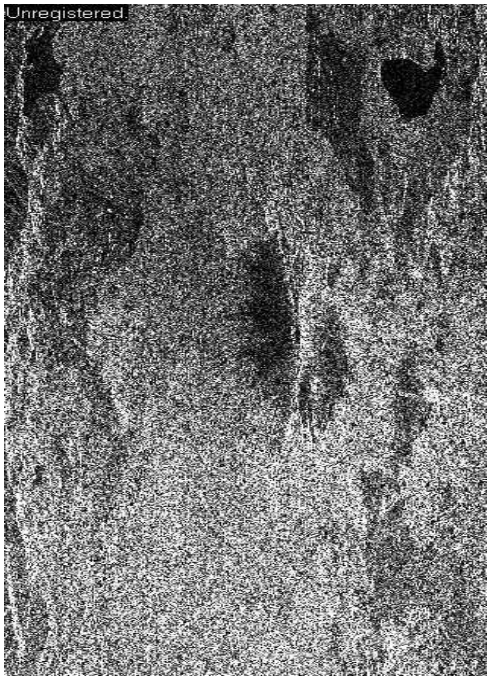
(iii) Registered image



(iv) Difference image

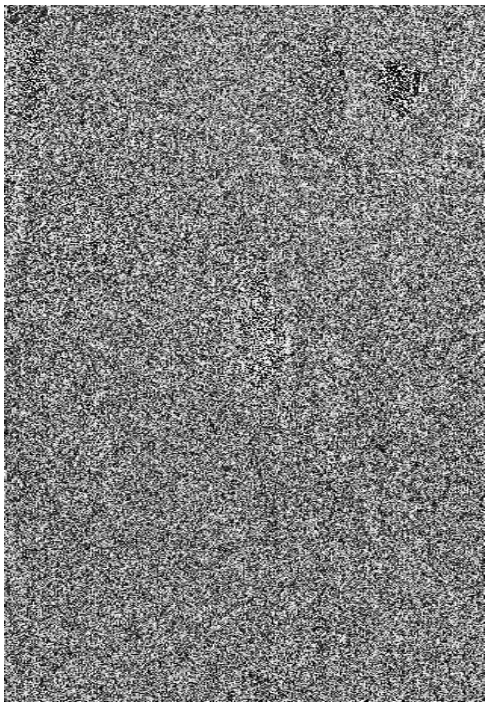
Figure 6.2 Multitemporal image taken in February 1993 ((i)) and April 1993 (ii) and the corresponding registered image (iii) and difference image (iv). Tx = 12, Ty= 0 , R =0

(i) Reference image



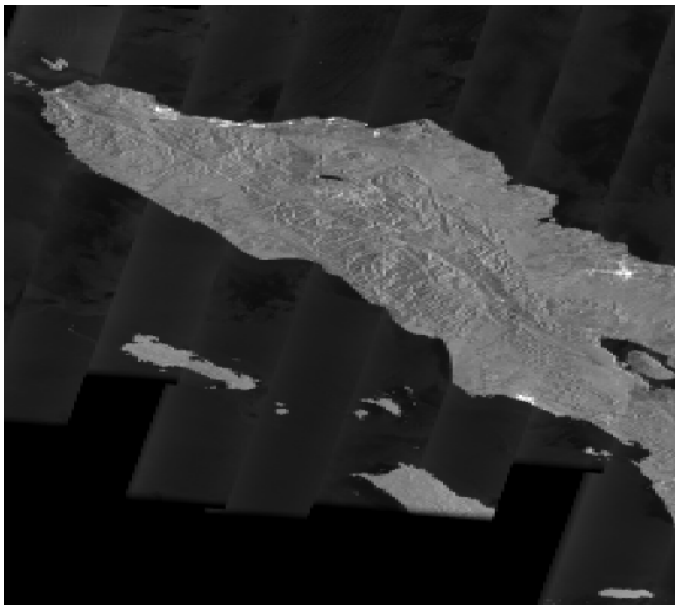
(ii) Input image

(iii) Registered image

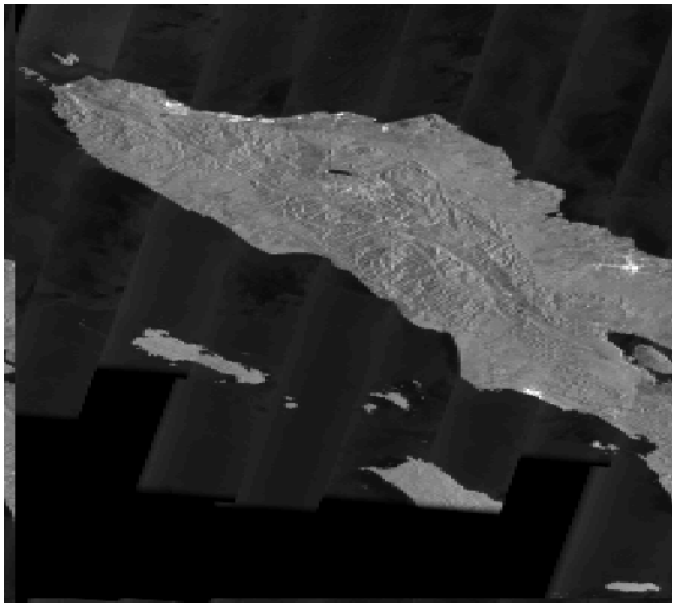


(iv) Difference image

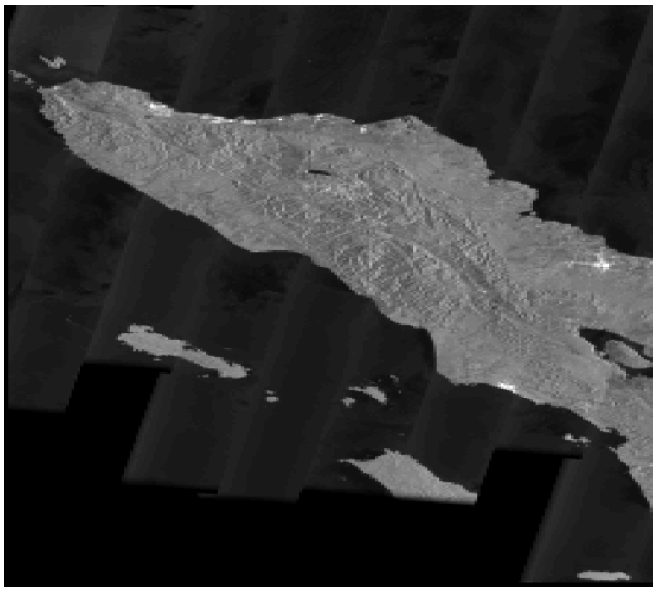
Figure 6.3 Multitemporal images taken in July 1993 ((i)) and August 1993 (ii) and the corresponding registered image (iii) and difference image (iv). Tx =3 ,Ty=-2 ,and R =0



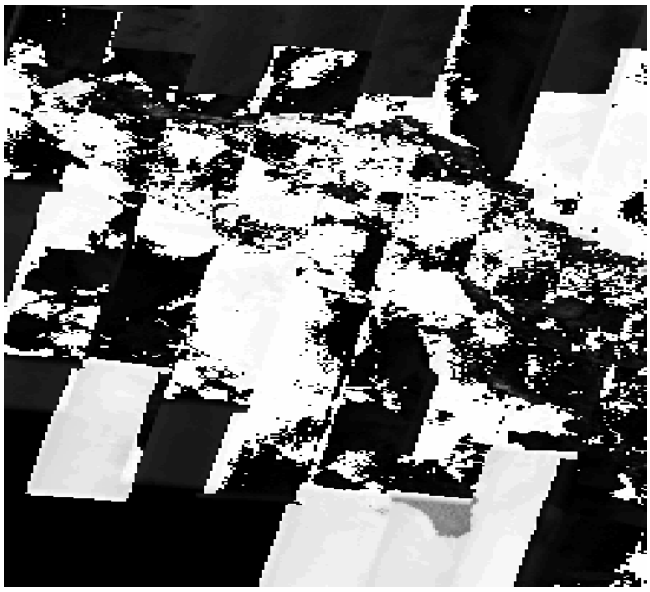
(i) Reference image



(ii) Input image

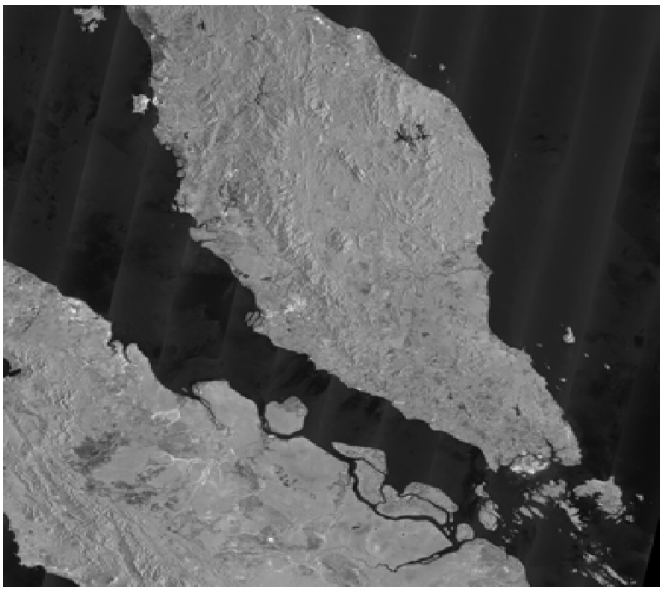


(iii) Registered image

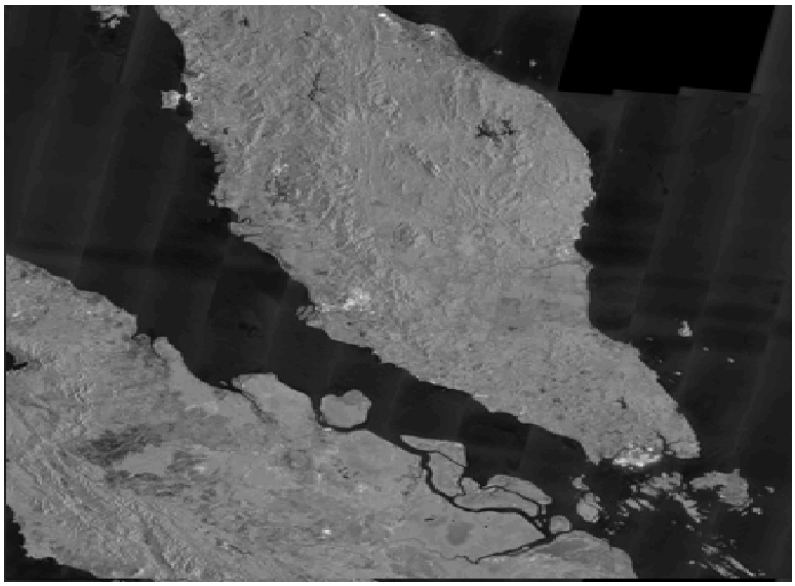


(iv) Difference image

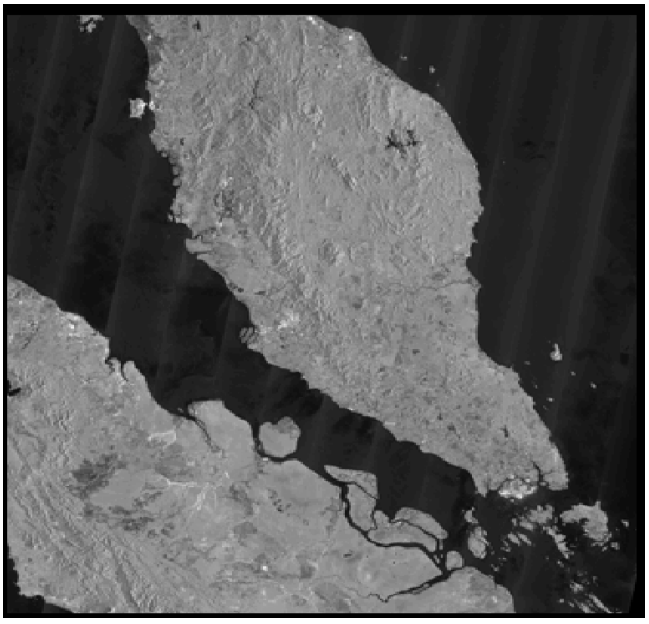
Figure 6.4 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv). Ty =3, Ty=-2, and R = 1°: The pattern structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



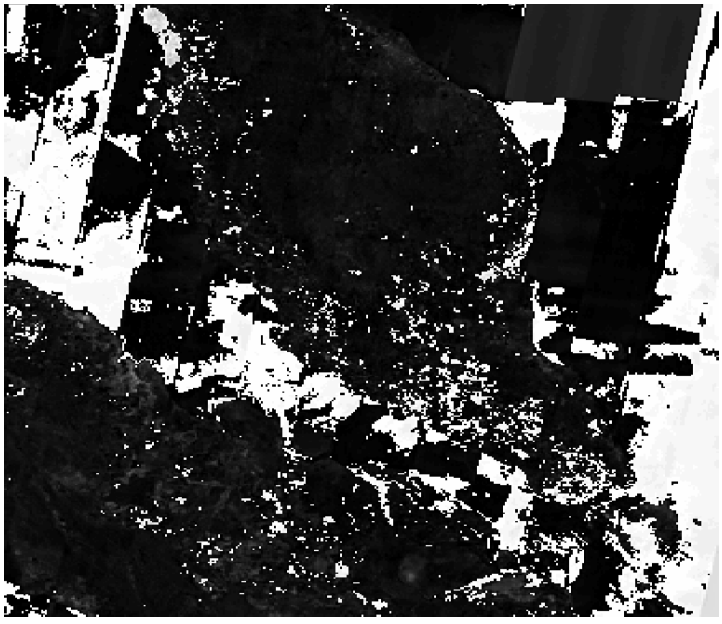
(i) Reference Image



(ii) Input image

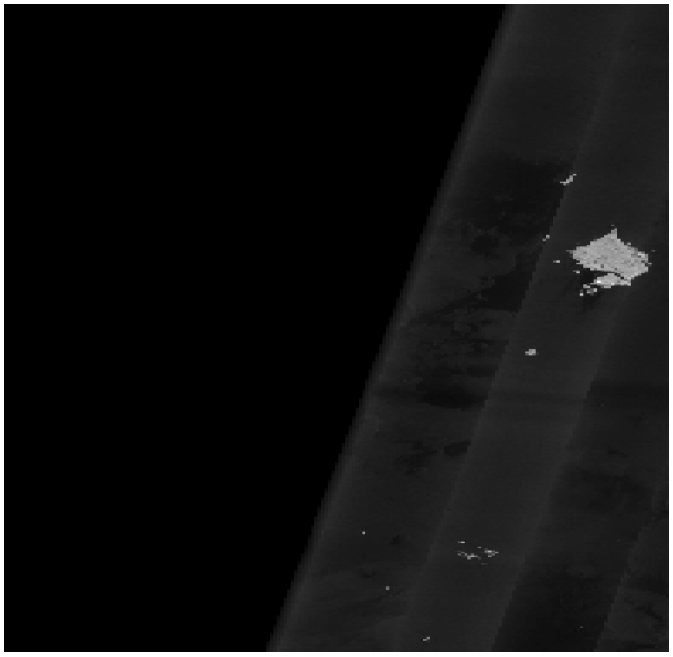


(iii) Registered image

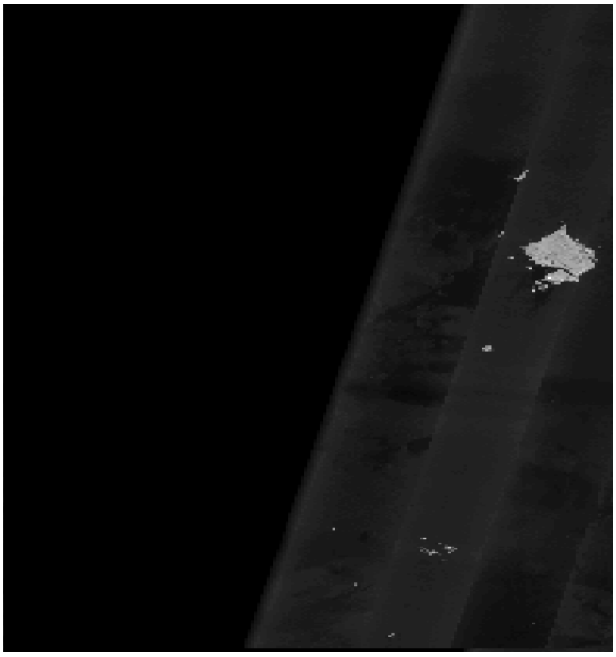


(iv)Difference image

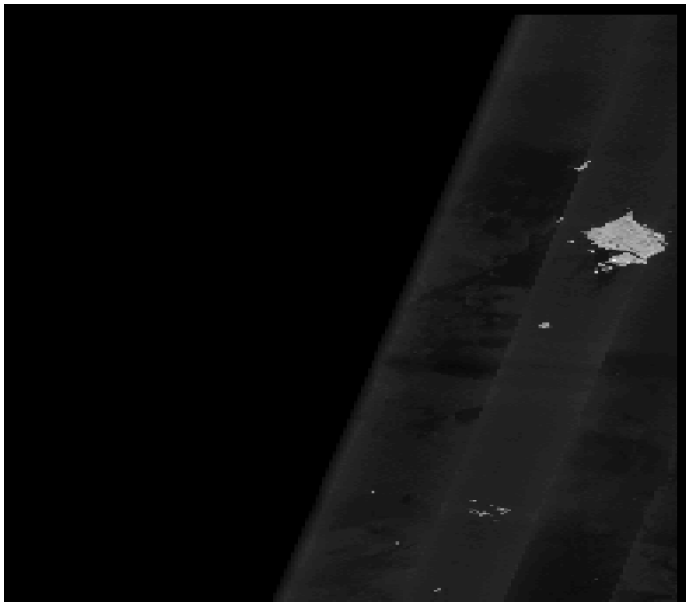
Figure 6.5 Slightly shifted image pair and the corresponding registered image (iii) and difference image (iv) . Tx =2, Ty=-1, and R=0. The structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



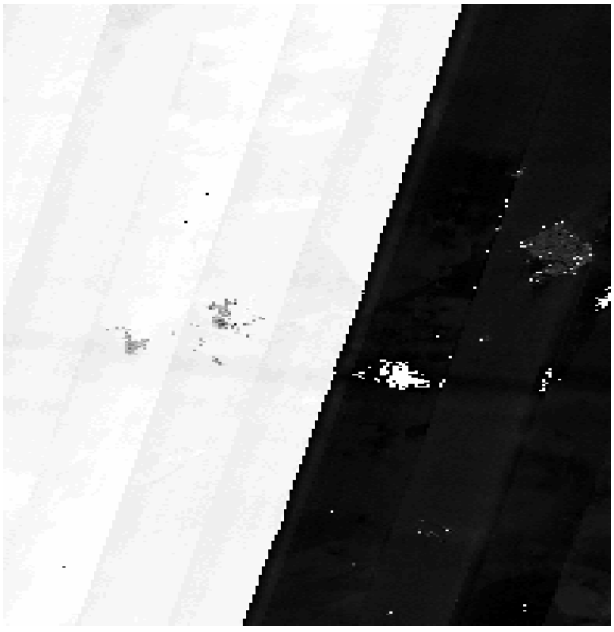
(i) Reference image



(i) input image

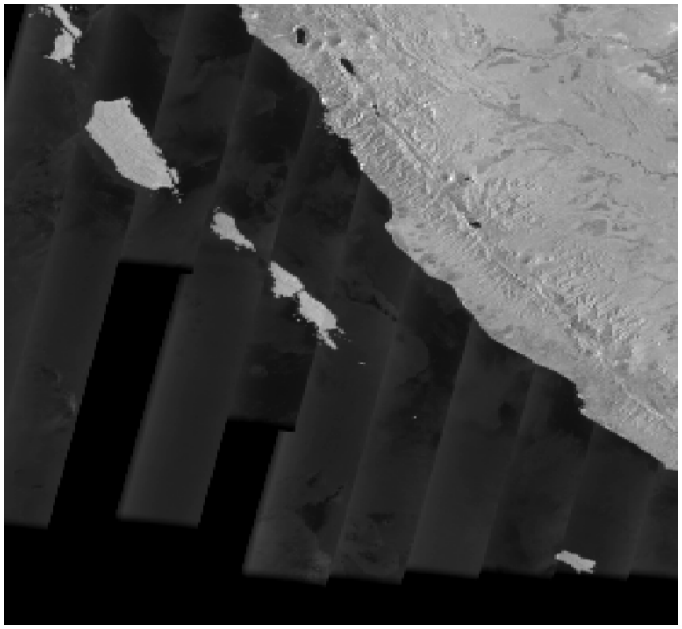


(iii) Registered image

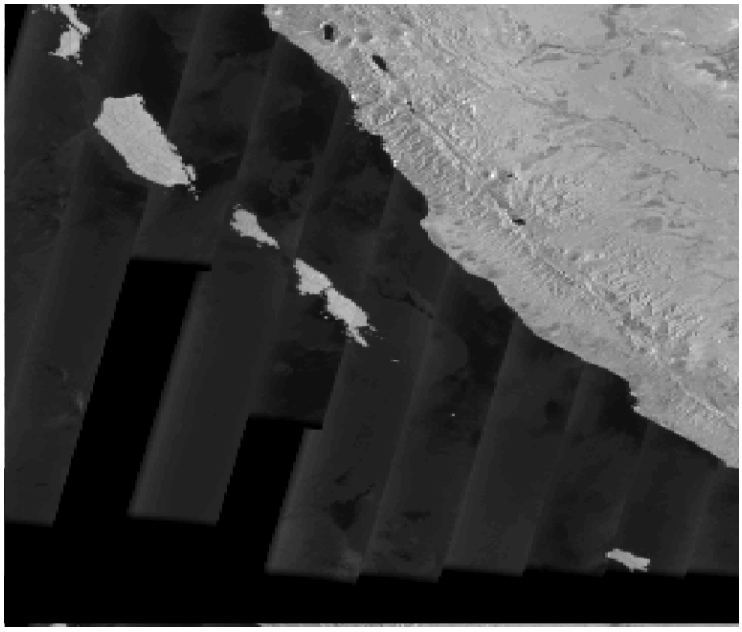


(iv) Difference image

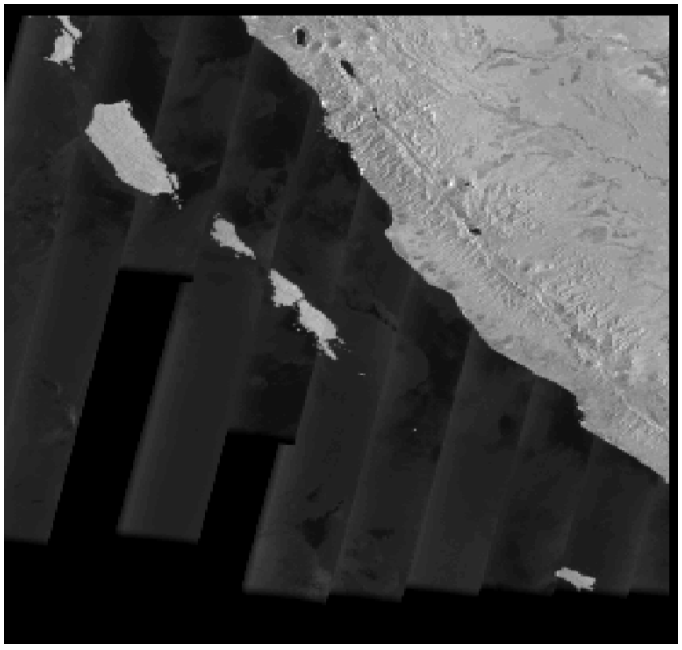
Figure 6.6 Slightly rotated image pair and the corresponding registered image (iii) and difference image (iv). Tx =0, Ty=0 and R=-2°. The pattern structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



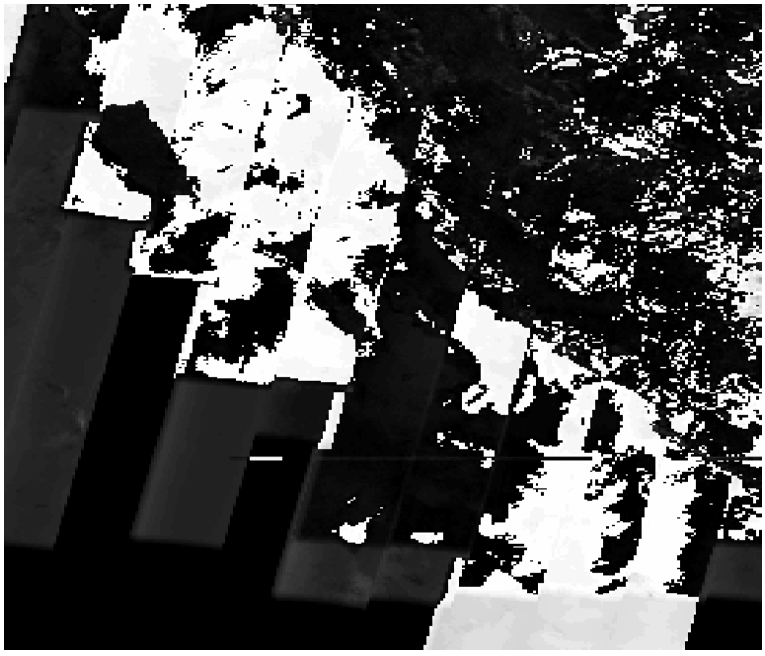
(i) Reference image



(ii) Input image

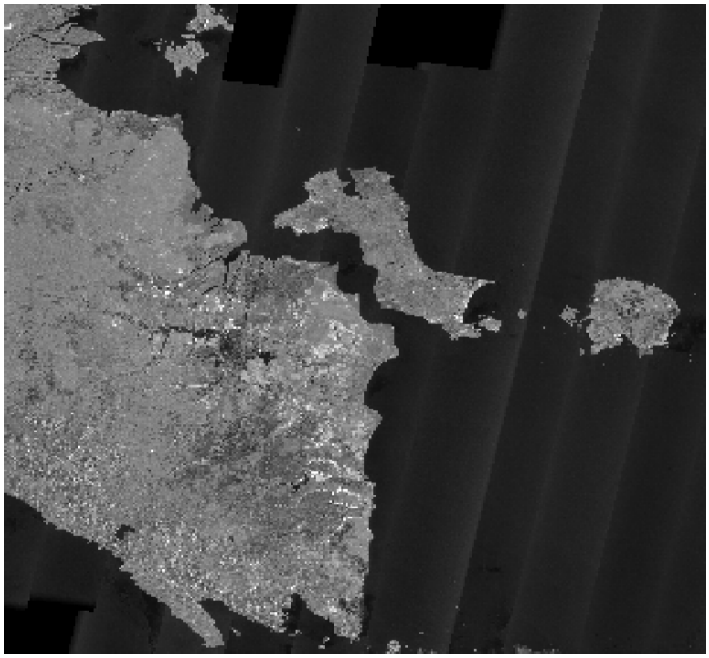


(iii) Registered image

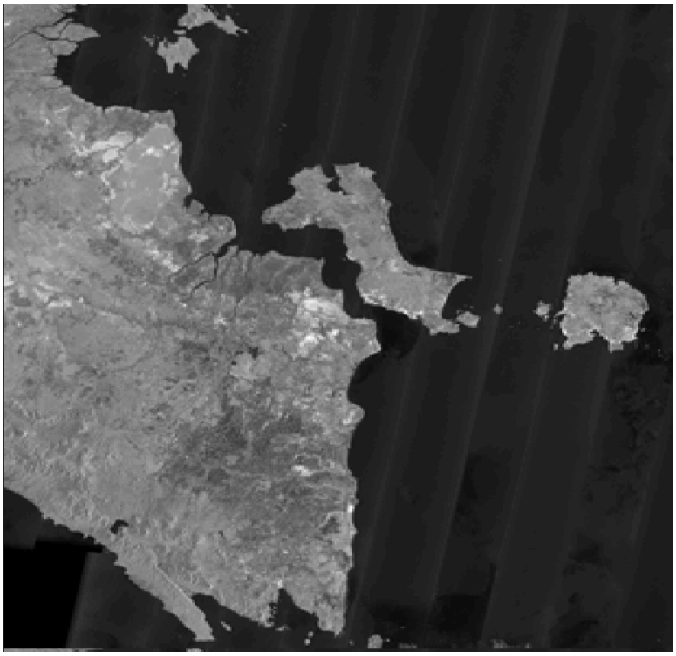


(iv) Difference image

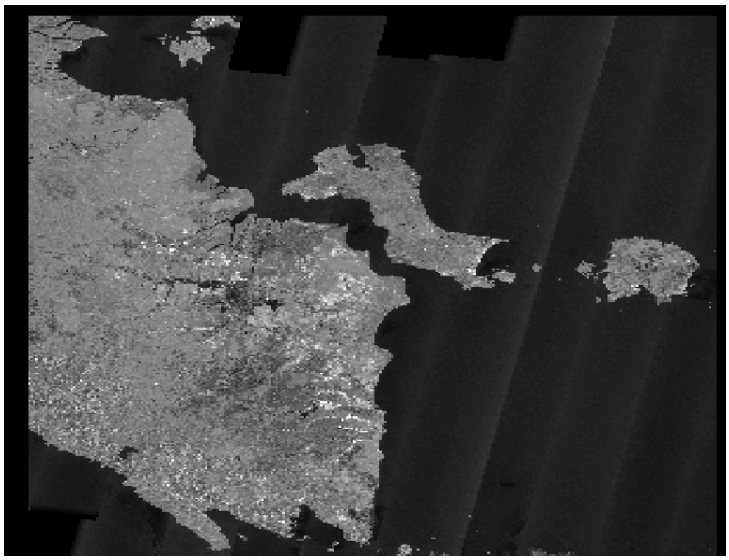
Figure 6.7 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv). Tx =-4, Ty=-1, and R=-2°. The pattern structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



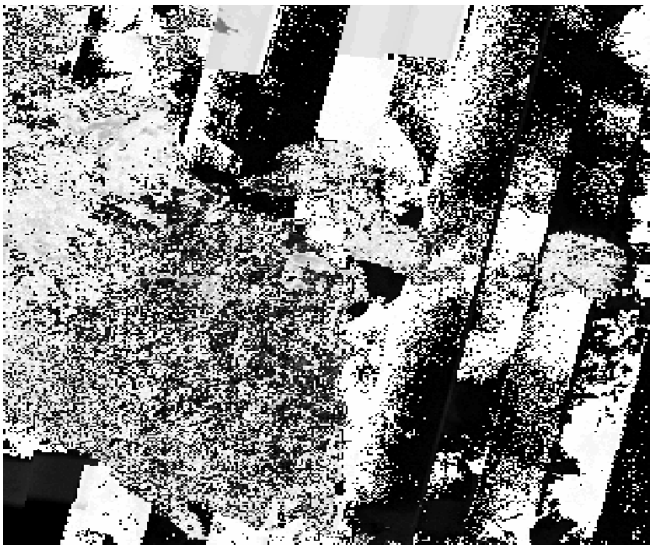
(i) Reference image



(ii) Input image

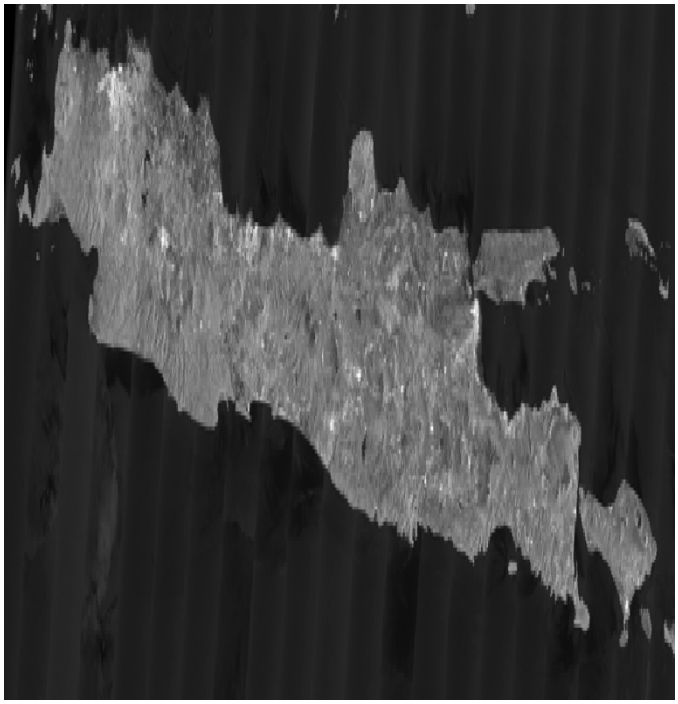


(iii) Registered image

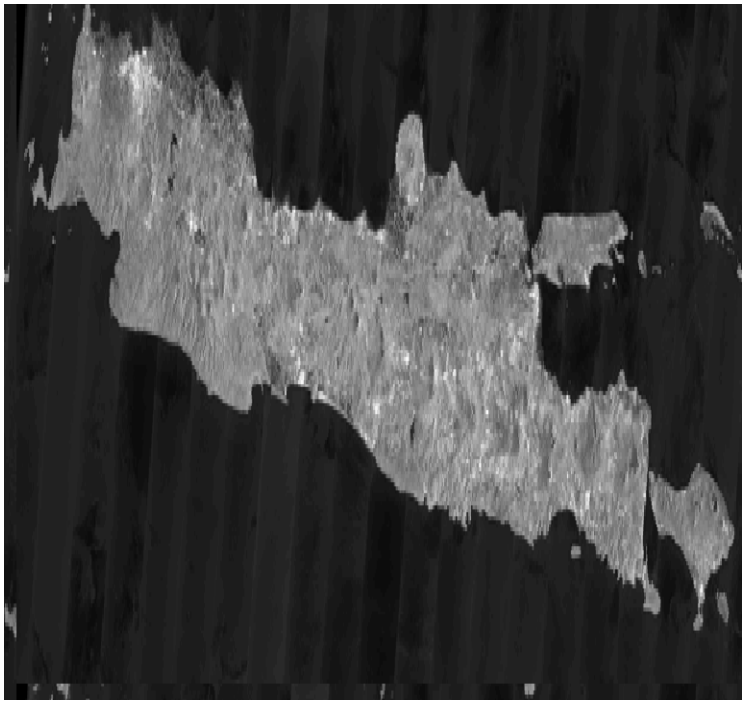


(iv) Difference image

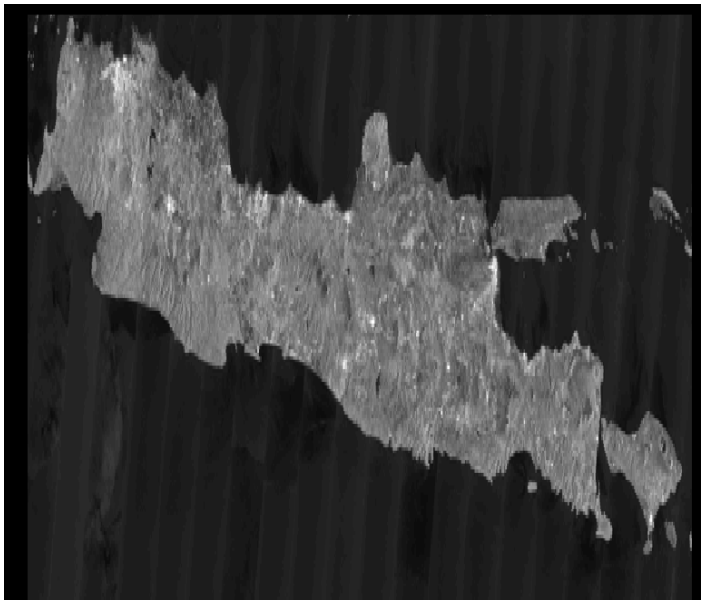
Figure 6.8 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv). Tx= 6 , Ty=0, and R =1°. The pattern structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



(i) Reference image



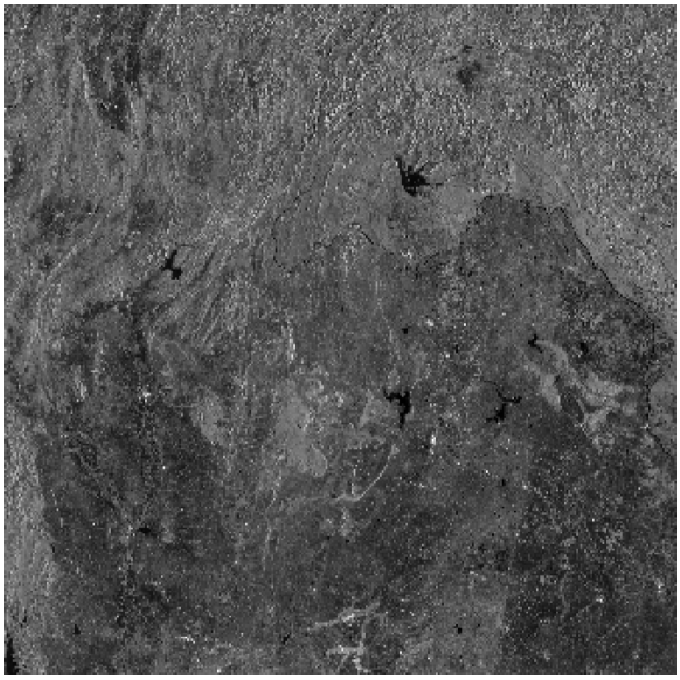
(ii) Input image



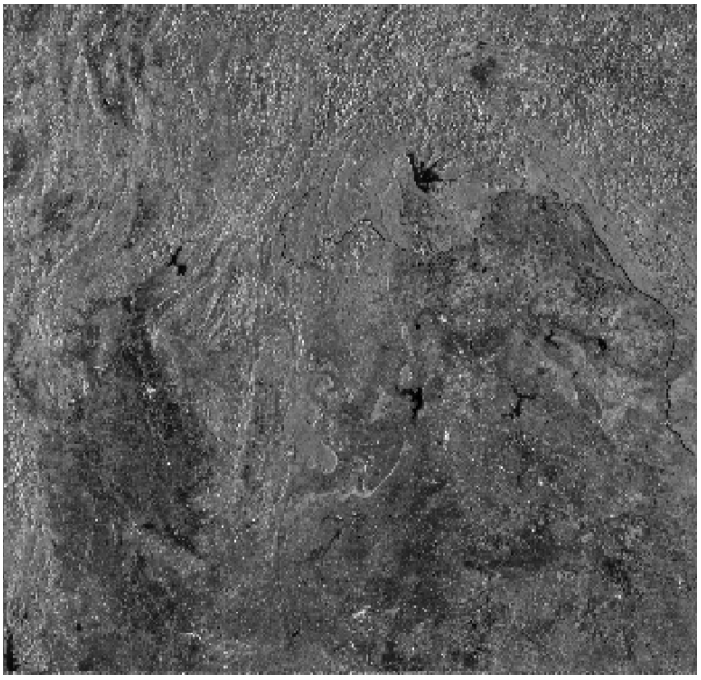
(iii) Registered image

(iv) Difference image

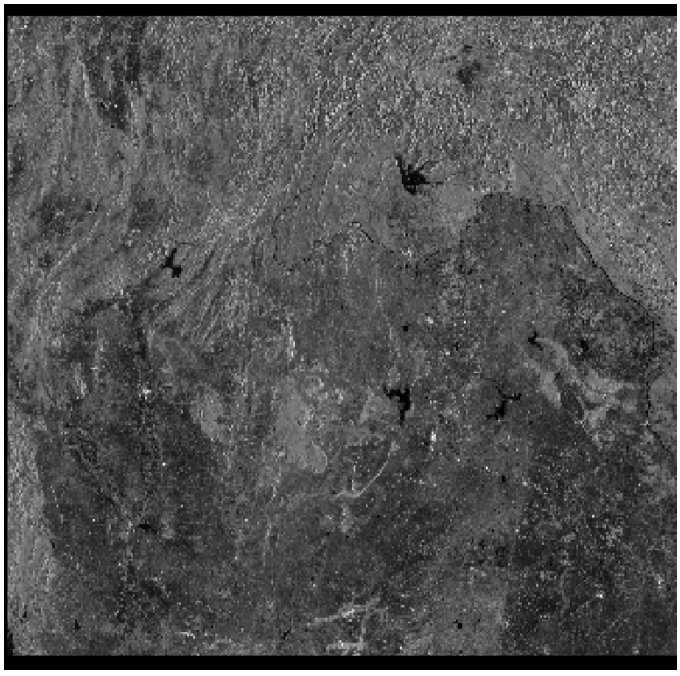
Figure 6.9 Slightly rotated image pair and the corresponding registered image (iii) and difference image (iv) .Tx=3 , Ty =1 and R=2°. The structures in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



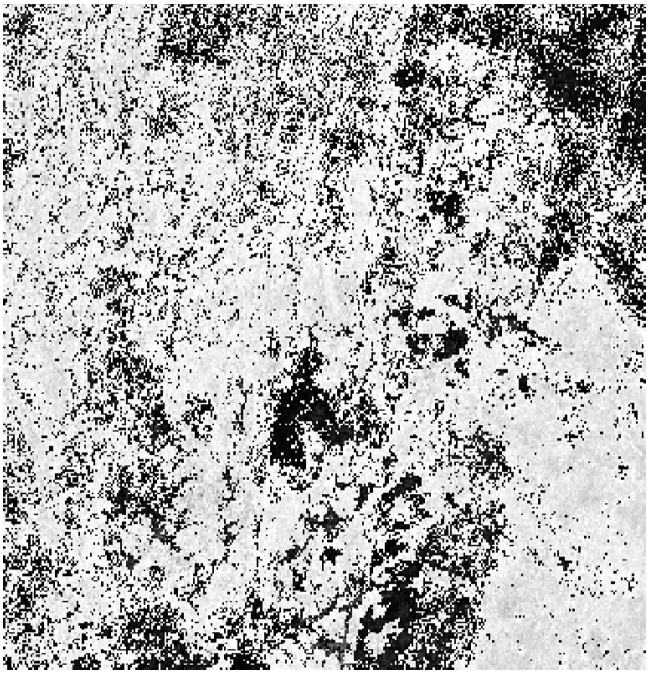
(i) Reference image



(ii) Input image

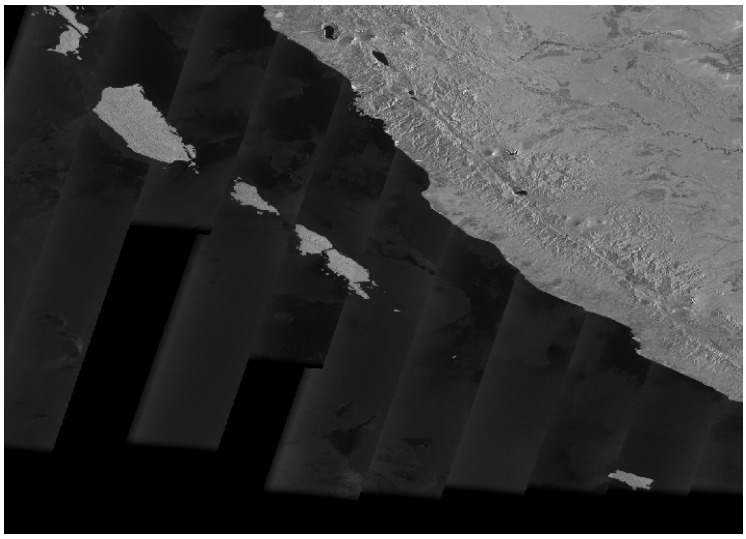


(iii) Registered image

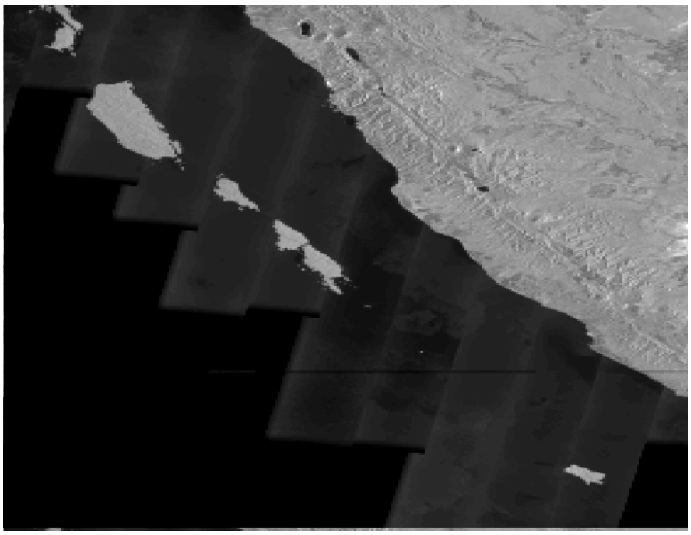


(iv) Difference image

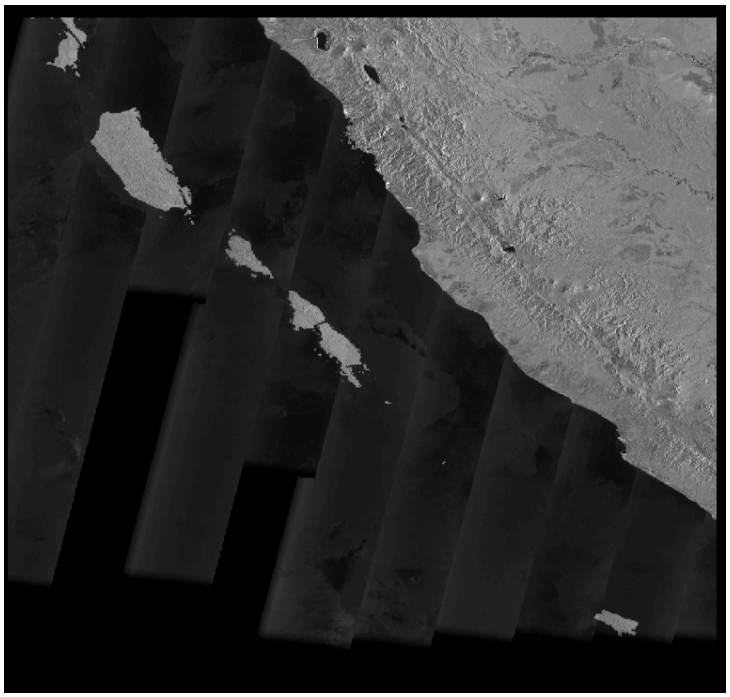
Figure 6.10 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv).Tx=1, Ty =2, and R =-1°. The structures in the difference image are not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255. The correlated nature of noise in difference image may be due to subpixel misalignment.



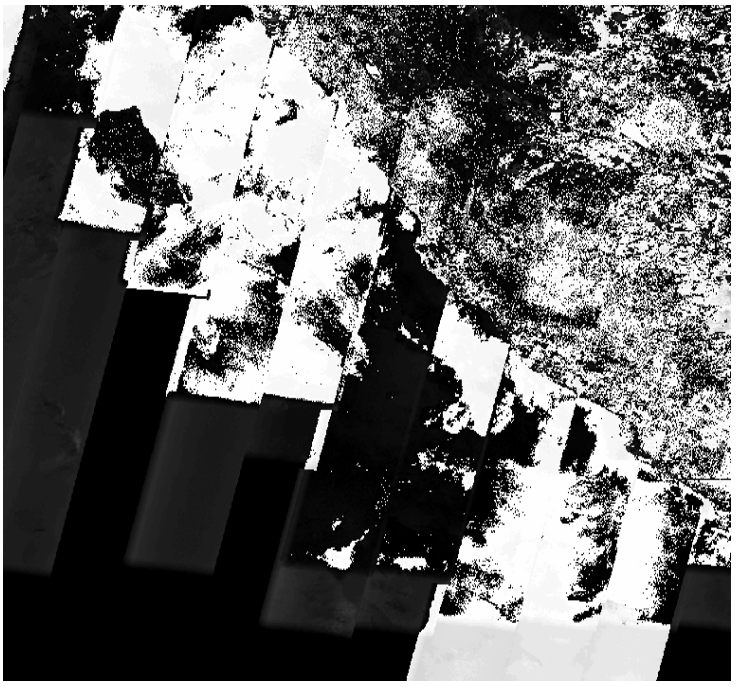
(i) Reference image



(ii) Input image

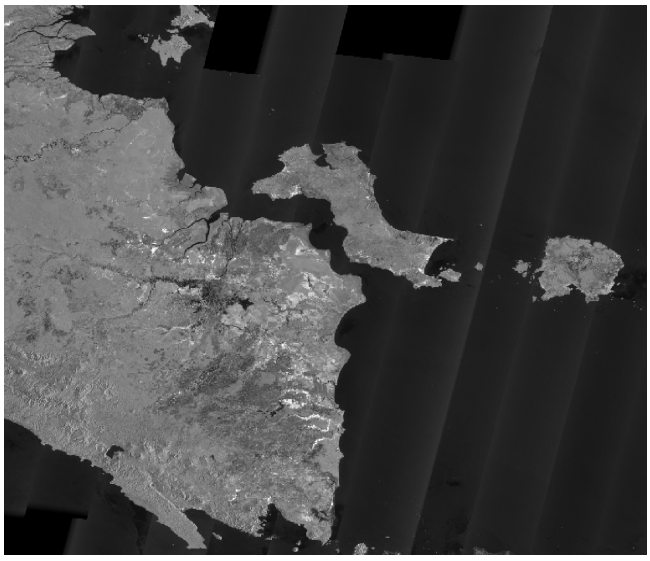


(iii) Registered image

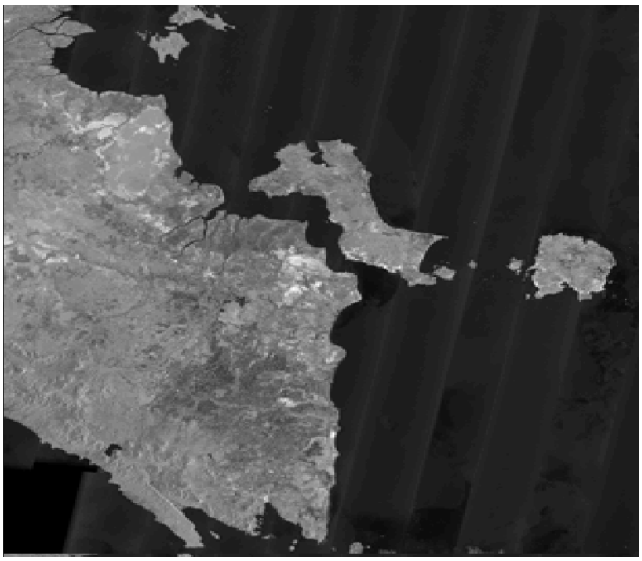


(iv) Difference image

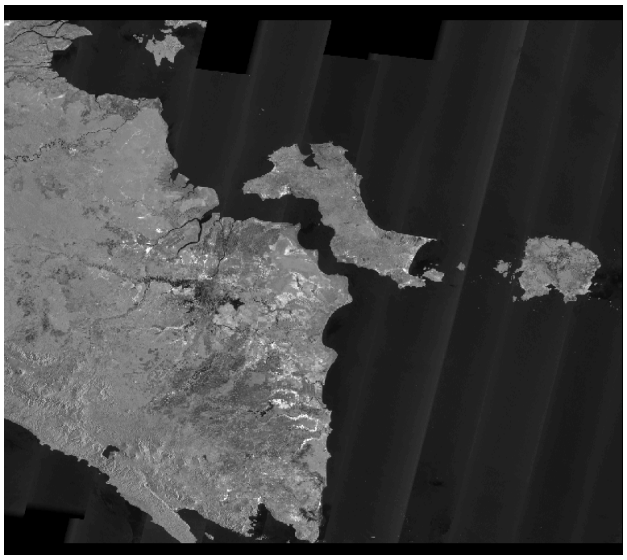
Figure 6.11 Slightly shifted image pair and the corresponding registered image (iii) and difference image (iv). Tx= 2, Ty=2, and R=0. The pattern structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



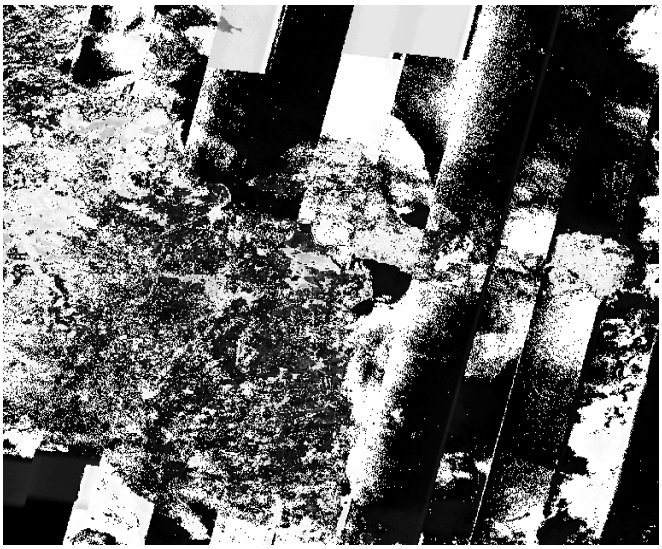
(i) Reference image



(ii) Input image

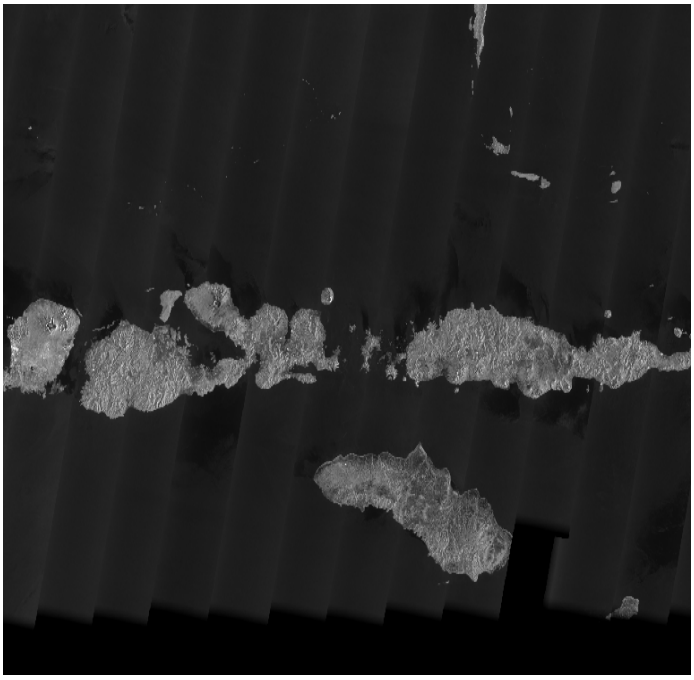


(iii) Registered image

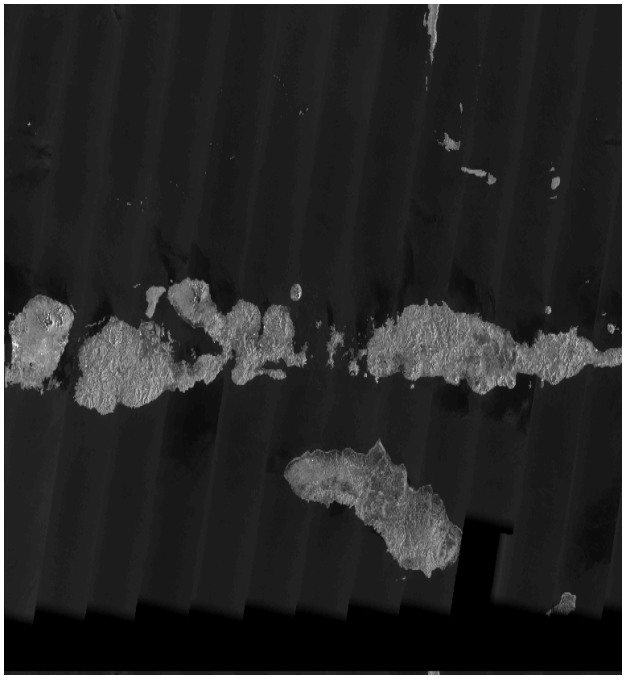


(iv) Difference image

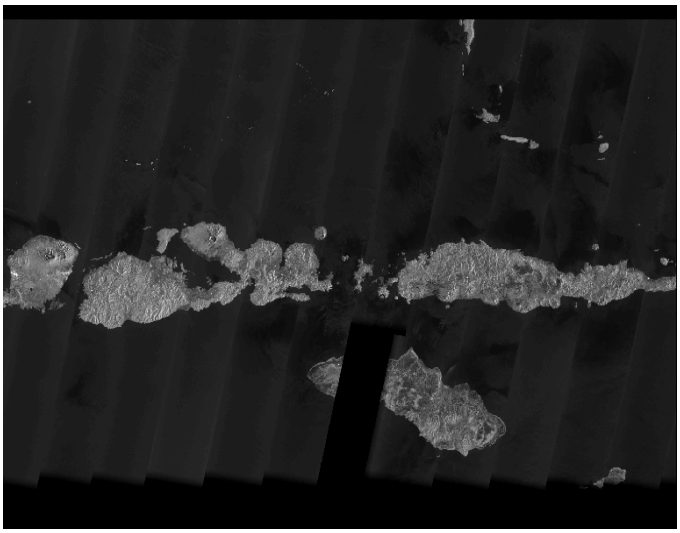
Figure 6.12 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv). Tx=2, Ty=1, and R=1°. The pattern structure in the difference image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



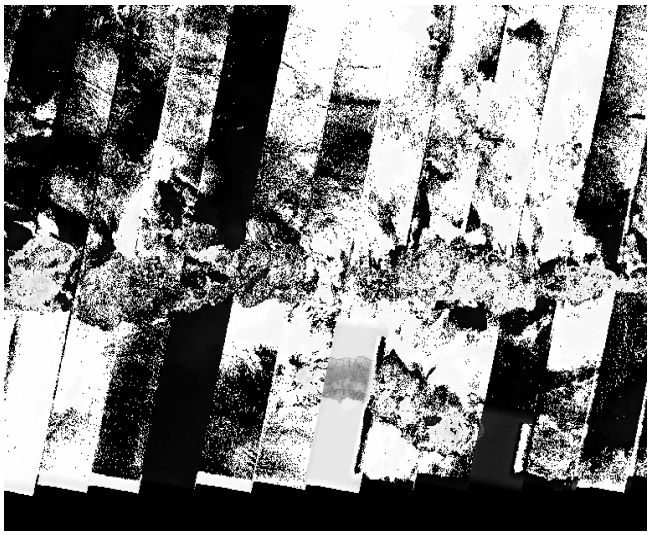
(i) Reference image



(ii) Input image

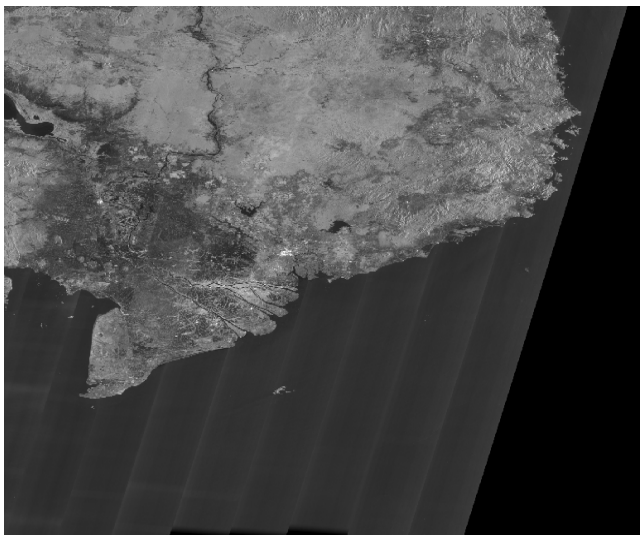


(iii) Registered image

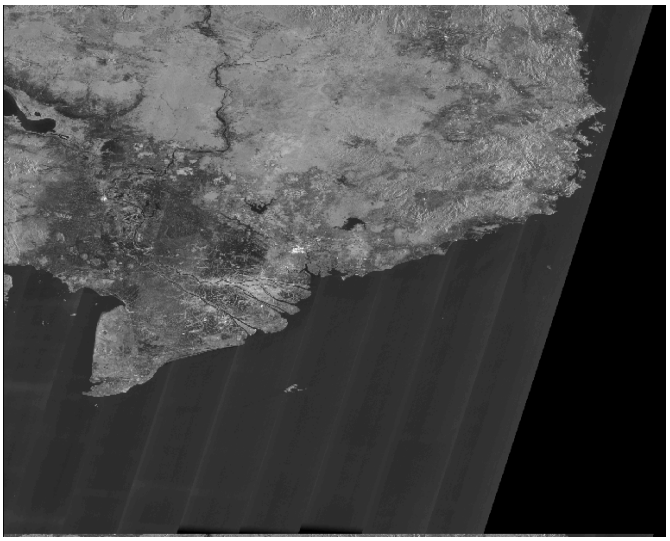


(iv) Difference image

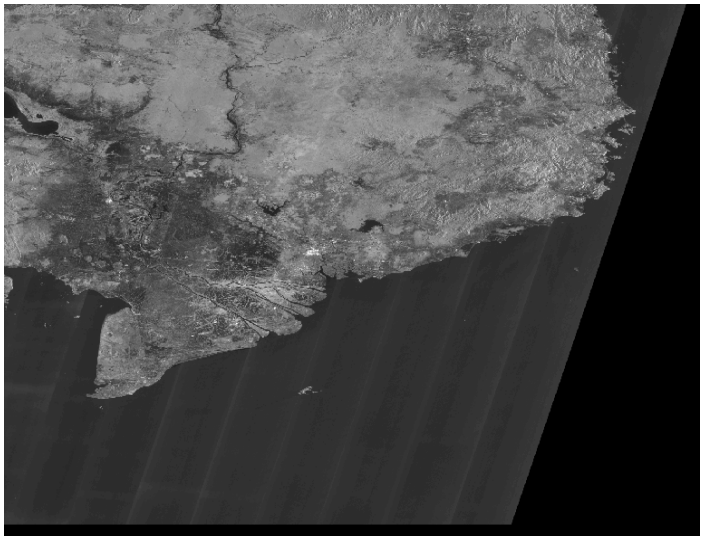
Figure 6.13 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv). Tx=2, Ty=-3, and R=1°. The pattern structure in the differences image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.



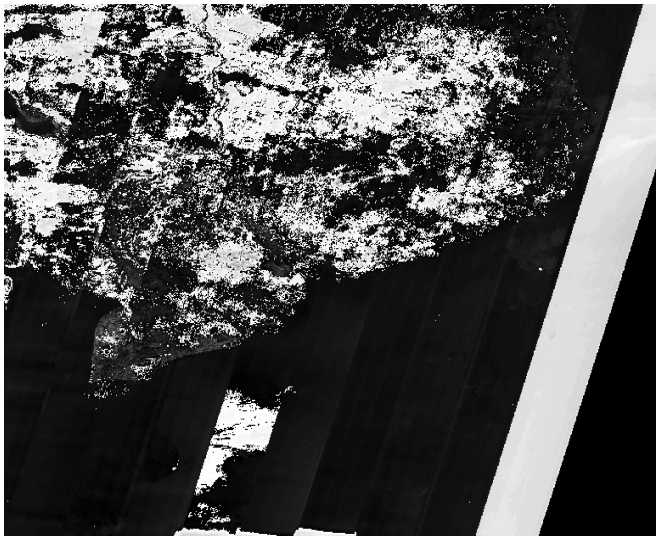
(i) Reference image



(ii) Input image



(iii) Registered image



(iv) Input image

Figure 6.14 Slightly shifted and rotated image pair and the corresponding registered image (iii) and difference image (iv). Tx=0, Ty=-3, and R=0. The pattern structure in the differences image is not only because the little differences between reference and input image but also the difference image is scaled between 0 and 255.

CHAPTER 7

CONCLUSIONS

In this chapter, we summarize the findings and conclusions of this work and suggest future research directions. In this thesis, we investigated not only methods of increasing reliability and robustness of mutual information, but also reducing the computational cost of image registration by using subimages.

7.1 Summary

The summary of each chapter is as follows:

Chapter two gives an overview of image registration methods. The literature review in chapter two shows that not much work has been done on the enhancement of the robustness of mutual information as a similarity metric. Little work is done on incorporating spatial information into mutual information. Even though using subimages is popular (as in remote sensing), the review also shows few efforts to use unique subimages in the registration of images. Therefore, in this work we investigate mutual information as a similarity measure thoroughly. As a search data strategy, we investigate using unique subimages to register images.

In chapter three, we presented automatic subimage selection methods for search data reduction and the enhancement of reliability and accuracy. We propose a new measure, called alignability, which shows the ability of subimages to provide robust and reliable results. We compare this feature to entropy, variance and gradient magnitude respectively. We show that using alginability produces not only reliable but robust results in image registration. The overall percentage of correctly registered subimages for alignalbility-based selection method is 20%,

19% and 32% more than entropy-based method, gradient-based method and variance-based method respectively.

In chapter four, we investigated mutual information as a similarity metric. We showed the effect of noise on mutual information and demonstrated that mutual information can be more robust to noise than correlation coefficient. We also propose two methods for the selection of bin size in the estimation of mutual information. We show that the methods enhance the robustness and accuracy of mutual information. We also compare MI with other F-Information measures (figures 4.1, 4.2, and 4.3). A search space strategy based on Robbins-Monro stochastic approximation algorithm is also presented.

The incorporation of spatial information into mutual information is presented in chapter five. We propose a new similarity metric, called spatial mutual information (SMI), which combines mutual information and a weighting function based on image gradient, image variance, and image entropy of local regions. The results show that SMI can be more robust and accurate than mutual information.
In chapter six, results from registration test images are presented.

7.2 Future research

In order to further improve the proposed methods for fast and robust image registration, the following studies further should be investigated:

- In order to reduce computational cost, we use subimage images. The size of subimages used depends on the type of image pairs being registered. Using a constant subimage size may not be optimal. It is of importance to study if the selection of the size of subimages could be fully automated.

- Using subimages rather than using the whole image for registration does not only reduce the computational effort but can also produce better results. This is true under the assumption that the selected subimage has strong features that can aid the registration. In our work we investigated features such as gradient, entropy, variance and alignability. One can also investigate the effect of features such as eigen values of subimages on the accuracy and robustness of subimage.

- The wavelet transform generates a multi-resolution representation of image data. In using such multi-resolution data, the size of the search data can be reduced by initially searching at lowest resolution images and the proceeding to higher resolution images where search results may be refined. Using wavelets preserves most important features of original data at low and high resolutions. It is of value to study the application of wavelet transforms as a search data strategy and the effects of applying wavelets before and after transformation of reference image.

- An interesting use of generalized divergence concepts is similarity measures. The relevant similarity measure class from f-divergence is the so-called f-information measure. A f-information measure between two datasets can be seen as a resemblance distance, which is minimal when the two datasets are similar. Incorporating Shannon's entropy as a convex function into f-information give mutual information. It is of importance to study the incorporation of other convex functions in f-information and compare with Shannon's entropy which is mutual information

- Normalized mutual information is used in a large number of studies [35]. It is less sensitive to overlap. A main drawback of normalized mutual information is that the dependence of neighboring pixels is ignored. Incorporating the dependence of gray values of neighboring of pixels

could improve registration. It is of importance to investigate the incorporation of spatial information to normalized mutual information.

REFERNCES

[1] Marcus Schwäbisch and Robert Siegmund, “Study on concepts for radar interferometry from satelittes and oceans (and land) applications”. http://www.ifm.zmaw. De/~wwwrs/koriolis_sect4.pdf.

[2] Babara Zitova and Jan Flusser, B. Noble, “Image registration methods: survey ” Image and Vision Computing , vol. 21, pp. 977-1000, June , 2003.

[3] Richard J A, *Remote Sensing Digital Image Analysis,* Springer Verlag, 1986.

[4] Brown.L. “A survey of image registration techniques”, ACM Computing Surveys, 24(4):325—376, 1992.

[5] B.K Ghaferry , A. A Sawchunk , “A survey of new techniques of image registration and mapping “, Proceedings of SPIE: Applications of Digital Image Processing 432 , pp 222-239, 1983.

[6] L.M.G Fonseca , B.S Manjunath,” Registration techniques for multisensor remotely sensed imagery “, Photogrametric Engineering and Remote Sensing. 62 1049-1056, 1996.

[7] E. Gülich , “Results of test on image matching of ISPRS WG”, ISPRS Journal of Photogrammetry of Remote Sensing (46) ,1991.

[8] G. J Penny , J Weese , J. A Little , P. Desmedt , D.L. G Hill and D. Hawkes,” A comparison of similarity measures for 2D and 3D image registration”, IEEE Transaction Medical Imaging vol. 17 No.4 586 595, 1998 .

[9] M. Svedlow C. D McGillem and P. E Anuta, “Experimental examination of similarity measures and preprocessing methods used for image registration”, Proc. Machine Processing Remote Sensing Data, 4A) -17, 1976.

[10] Colligion A ,Maes F , Delaere D Vandermuelen D, Suetens P and Marchal, G Automated , “Multi-modality image registration based on information theory” , Proceedings in the 14th International Conference of Information Processing and Medical Imaging ,pages 263-274, 1995.

[11] Voila P, Wells M, “Alignment by Maximization of Mutual Information”, Proceedings 5th International conference in compuer Vision Pages 16-23, 1995.

[12] F. Maes M Vandermeulen, P. Suestens,” Medical Image Registration using Mutual information “, Proc IEEE Medical Imaging Vol. No. 10 , October , 2003

[13] Davis.W, S.K Kenue, “Automatic selection of control points for registration of Digital Images”, Proc. of International Joint Conference of Pattern Recognition, 4th, Kyoto, Japan, pp.936-938, Nov.7-10, 1978.

[14] E.M Rignot, R.Kwok, J. C. Curlander, and S.S Pang , “ Automated multisensor registration : requirements and techniques” ,Photogrametric Engineering and Remote Sensing ,Vol.58, No.8, August 1991,pp.1029-1038.

[[15] Q. Zheng and R. Chellappa,” A computation vision approach to image Registartion”. IEEE Transaction on Image Processing , 2(3):311-325 ,July 1993

[16] J. leMoigne, “ Parrellel registration of multiple resolution remote sensing data”, Proceedings of IGARSS’95, Firenze, Italy, July 10-14,1995.

[17] S.G Mallat,”A theory for multi-resolution signal approach,” Journal of Photogrammetry and remote Sensing Vol.59, No.5 pp.645-653, May 1993.

[18] E. D Castro, C. Morandi, ”Registration of translated and rotated images using finite Fourier transform “, IEEE Transaction on Pattern Analysis and Machine Intelligence 9,) 700-703 1987.

[19] H. Foroosh, J. B Zerubia, M. Berthod, “Extension of phase correlation to subpixel registration”, IEEE Transactions on Image Processing 11) 188-200, 2002.

[20] Kirkpatrick S, Gelatt C. D ; Vecchi ,M.P, Optimization by Simulated Annealing, Science, Vol.220, no 4598 pp 671-680, May 1983.

[21] Torrence, C. and G.P. Compo, A "Practical Guide to Wavelet Analysis", Bull. Amer. Meteor. Soc., 79, 61-78, 1998.

[22] H. Robbins, S. Monro, “A Stochastic Approximation Method”, Ann. Math Statistics, 382-386, 1954.

[23] J.M Fitizpatrik, “Detection failure assessing success, in :J.V. Hajnal, D.L.G Hill , D.J Hawkes (Eds), Medical Image Registration , CRC Press, Baton Rouge , Florida 2001.

[24] J M Galvez, M. Canton, Normalization and shape recognition of three-dimensional objects by 3-D moments”, Pattern Recognition, vol.26 No.1, pp.667-682, 1993.

[25] R. C, Gonzalez, R. E Woods, *Digital Image Processing*, Addison-Wesley, 19 92.

[26] C. E Shannon; “A Mathematical theory for communication systems”, Bell System Technical Journal, Vol. 27, pp 379-428, 1948

[27] H Wards worth, *Handbook of Statistical Methods for Engineers and Scientist*, 2 Edition McGraw-Hill, 1998

[28] R.P Woods, S.R Cherry, and J. C Mazziotta," Rapid automatic algorithm for aligning and reslicing PET Images" Journal of Computer Assisted Tomography vol.16,no.4,pp 620-633,1992.

[29] C. Studhomle, D.L Hill, and D.J Hawks, "Multiresolutional Voxel Similarity Measures for MR-PET Registration " in Information Processing and Medical Imaging ,Y Bizais , C. Barillot, and R. Di Paola, Eds 1995,pp 287-298, Kluwer Academic Publishers, Dortrecht.

[30] A Colligion , D. Vandermeulen, P. Seutens, and G. Marchal , " 3D multimodality medical image registration using feature spacing clustering" Computer Vision, Virtual Reality and Robotics and Medicine, vol. 905 of lecture of Computer Science Notes , pp195-204, Spriger Verlag, Berlin 1995.

[31] P. Voila and W.M Wells III," Alignment by maximization of mutual information", in International Conference on Computer Vision, pp 16-23 , 1995.

[32] J. Aczel, Z. Darcozy "On the measures of information and their characterizations, Academic Press, New York 1975

[33] Vadja.I," Theory of statistical inference and information", Dordrecht, Netherlands: Kluwer, 1989.

[34]Jianhua .L "Divergence Measures Based on Shannon Entropy" Information Theory IEEE Transaction pp145-152, Vol.37, No.1, 1991.

[35]C. Studholme, D.L.G Hill, and J Hawkes, “ Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures” ,Medical Physics vol.24.no.1, pp. 25-35,1997.

[36] K. Sjögreen, M.Ljungberg, K. Wingardh,K. Erlandsson, S. E Strand ,” Resgistration of emission and transmission whole-body scintillation camera images”, Journal of Nuclear Medicine ,vol.42, no.10,pp.1563-1570, 2001.

[37] R. Bansal, L.Staib, Z. Chen, A.Rangarajan, J.Kinsley, R.Nath, and J. Duncan,” A minimax entropy registration framework for patient set up verification and radiotherapy”, Computer Aided Surgery, vol.4, no.6, pp147-154, 2000.

[38] D. Hill, Studholme, C. and Hawkes, D Voxel, “Similarity measures for automated image registration, Proceedings Third International Conference of Visualisation in Biomedical Computing, 2359:205-216, 1994.

[39] A. Amankwah, O. Loffeld, “Image Registration by Automatic Subimage Selection and Maximization of Mutual Information”, Proceedings IEEE IGARSS05, July 2005.

[40] A. A. Cole-Rhodes ,K.L Johnson, J. LeMoigne, I. Zavorin, ”Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient”, IEEE transaction on Image Processing , Vol..12, pp 1495-1511, December, 2003.

[41] H.H Yang, S.Vann Vurren, S.Shama, and H.hermassky, “Relevance of time frequency features for phonetic and speaker-channel classification” Speech Commun. 31, pp. 35-50, 2000.

[42]. Pluim J, Maintz A, Viergever M, Image registration by maximization of combined mutual information and gradient information. IEEE Trans. Med. Imag., vol. 19, no.8, 809–814, 2000.

[43] Rueckert D, Clarkson M.J., Hill D.L.G., Hawkes D.J, " Non-rigid registration using higher-order mutual information" ,Medical Imaging: Image Processing, K.M Hanson,Ed. Bellinghamn, WA, SPIE Press, 2000.

[44]. Ma B., Hero A.O., Gorman J., Michel O, "Image registration with minimal spanning tree algorithm". IEEE International Conf. on Image Processing, Vancouver, 2000.

[45] M. Belis and S. Guiasu, "A quantitative-qualitative measure of information in cybernetic systems" . IEEE Trans. Inform. Theory, 14, 593-594, 1968.

[46] H. C. Taneja and R.K. Tuteja," Characterization of a quantitative-qualitative measure of relative information". Inform. Sci., 33:217-222, 1984.

[47]. A. Amankwah, O. Loffeld, "On search space and search data strategies in image registration by Automatic Subimage Selection and Maximization of Mutual Information" Proceedings IEEE OCEANS05 , September, 2005.

[48] I.Vadja, *Theory of Statistical Inference and Information* . Dordrecht , The Netherlands :Kluwer Academic,1989.

[49] G. Stockman, S. Kopstein, S. Benett, " Matching images to models for registration and object detection via clustering ", IEEE Transactions on Pattern Analysis and Machine Intelligence 4 ,229–241 ,1982.

[50] A. Goshtasby, G. C. Stockman, "Point pattern matching using convex hull edges", IEEE Transactions on Systems, Man and Cybernetics 15

631–637, 1985.

[51] S. Abdelsayed, D. Ionescu, D. Goodenough," Matching and registration method for remote sensing images", Proceedings of theInternational Geoscience and Remote Sensing Symposium IGARSS'95, Florence, Italy, pp. 1029–1031, 1995.

[52] Q. Zheng, R. Chellapa, "A computational vision approach to image Registration", IEEE Transactions on Image Processing 2 ,311–325, 1993.

[53] B. Zitova´, J. Flusser, F. Sroubek, "Application of image processing for the conservation of the medieval mosaic", Proceedings of theIEEE International Conference on Image Processing ICIP'02, Rochester, MI, pp. 993–996, 2002.

[54] A. D. Ventura, A. Rampini, R. Schettini, "Image registration by recognition of corresponding structures", IEEE Transactions on Geoscience and Remote Sensing 28 (1990) 305–314.Transactions on Medical Imaging 18 419–42, 1992.

[55] R.S. Mitra, N.N. Murthy, "Elastic maximal matching", (Pattern Recognition 24) 747–753, 1991.

[56] A. Goshtasby," Image registration by local approximation methods", Image and Vision Computing 6, 255–261, 1988.

[57]J. Flusser, B. Zitova´, Combined invariants to linear filtering and rotation, International Journal of Pattern Recognition and Artificial Intelligence 13 1123–1136, 1999.

[58] T. Lehmann, W. Oberschelp, E. Pelikan, and R. Repges, Bildverarbeitung f¨ur die Medizin: Grundlagen, Modelle, Methoden, Anwendungen (in German). Berlin, Germany: Springer-Verlag, 1997.

[59] S. W. Rowland, “Computer implementation of image reconstruct [tion formulas,” in Image Reconstruction from Projections: Implementation and Applications, G. T. Herman Ed. Berlin, Germany: Springer- Verlag, 1979.

[60] A. Goshtasby, “Template matching in rotated images”, IEEE Trans. Pattern Analysis and Machine Intelligence, 7(3):338–344 (1985).

[61] Freedman D and Diaconis P On the histogram as a density estimator:L2 theory. Probability Theory and Related Fields. 57(4): 453-476, 1981.

Appendix

A Mutual Inormation IDL Code

```
Compute the average mutual information --- a.m.i. --- of the reference Image
;Compute the average mutual information --- a.m.i. --- of theinput image.
;(i.e. knowing SAR Image1 (Master)  how much can we infer about Slave
Image.)
;INPUTS.
;  S    = Slave Image.
;  T    = Master Image
;KEYWORDS.
;  /plots   = set this to plot the histograms of S and T
;OUTPUT.
;  Returns the Mutual Information between S and T --- a scalar.
;NOTES.Depending on application the bin size would have to be adjusted!!

;Sept. 25, 2004. Amankwah Anthony ( all rights reserved)


Function Mutual_Information,S,T
  mx = max(s)
  mn = min(s)

  histos  = histogram(s, max=mx, min=mn, binsize=4)
  histosT = histogram(T, max=mx, min=mn, binsize=4)
  histo2d = hist_2d(s, T, max1=mx, min1=mn, max2=mx, min2=mn, bin1=4,
bin2=4)
  histos  = float(histos) / float(N_elements(s))
```

```
  histosT = float(histosT) / float(N_elements(s))
  histo2d = float(histo2d) / float(N_elements(s))
  bins = findgen(N_elements(histos)) * 4 + mn
fd=0.0
fsum= 0.00
zz=0.0

 ;if keyword_set(plots) then BEGIN
   ; window, 12
   ; !p.multi = [2, 1, 2]
   ; plot, bins, histos, xstyle=1
    ;plot, bins, histosT, ystyle=1
 ; endif

  info = 0.0
  N= N_elements(s)ssum =0.0

for f=0,N_elements(histos) -1 do BEGIN
for g=0,N_elements(histosT)-1 do Begin

    ps1 = histos[g]

    psT1 = histosT[f]
    p2s1 = histo2d[g,f]
    ;nterm1 = (p2s1 )  * ALOG((p2s1 ) / ((ps1 ) * (psT1 )))
    nterm1 = abs (p2s1-ps1*psT1)

    if (ps1 GT 0) AND (psT1 GT 0) AND (p2s1 GT 0) then begin
    ssum = ssum + nterm1
endif
endfor
```

endfor

Return ,ssum ;// Mutual Information

end