Technische Universität München
Photogrammetrie und Fernerkundung
Prof. Dr.-Ing. U. Stilla

Automatic Registration of High Resolution SAR and Optical Satellite Imagery in Urban Areas

Sebastian Türmer

Diplomarbeit

Bearbeitung: 01. 11. 2008 - 30. 04. 2009

Studiengang: Geodäsie und Geoinformation (Diplom)

Betreuer: Sahil Suri (DLR)
Uwe Stilla (TUM)

Kooperation: Institut für Methodik der Fernerkundung

2009

Acknowledgements

First I would like to express my sincere thanks to M. Tech. Sahil Suri my supervisor at German Aerospace Center (DLR). He taught me a lot of worth knowing things and exercised nearly infinite patience. As well I like to thank my supervisor at the Technical University Munich Prof. Dr.-Ing. Uwe Stilla who always had some special ideas and stimulated further thought. In addition I like to thank Dr. Danielle Hoja, Dipl.-Ing. Jens Leitloff, Dipl.-Inf. Peter Schwind, B. Tech. Jagmal Singh and Dipl. Phys. Rolf Stätter, all from department of Photogrammetry and Image Analysis at DLR, for their support and advice in diverse situations. My special gratitude goes to Dr.-Ing. Peter Reinartz, head of department Photogrammetry and Image Analysis, who gave me the opportunity to work on this interesting project.
Munich, April 2009

Contents

List of Figures

List of Tables

List of Abbreviations

CCD	Charge Coupled Device
DEM	Digital Elevation Model
DSM	Digital Surface Model
GCP	Ground Control Point
MI	Mutual Information
SAR	Synthetic Aperture Radar
UTM	Universal Transverse Mercator

Abstract

The work presented in this thesis addresses the problem of registration of high resolution SAR and optical satellite imagery in dense urban areas. With the launch of high resolution remote sensing satellites in different modalities like TerraSAR-X, WorldView-1 and IKONOS-2, the contribution of remote sensing for various applications has received a tremendous boost. Specifically, the combined analysis of high resolution SAR and optical imagery is of immense importance in monitoring and assessing catastrophes and natural disaster. Although, latest satellites provide georeferenced and orthorectified data products, still registration errors exist within images acquired from different sources. These need to be taken care off through quick automated techniques before the deployment of these data sources for remote sensing applications. Different registration methods like the ones based on detecting geometrical features like interest points and line detectors to intensity based strategies have been rightfully evaluated. Considering the meticulous task of extracting and matching conjugate features in SAR and optical imagery (especially metric resolution imagery) a general feature based image registration technique for various scenarios might be difficult to develop and implement and therefore region based approach has been selected in this thesis. In general, for high resolution satellite images acquired over urban areas, common city features like wide roads, rivers, big stadiums, play grounds, parks can be expected to appear in considerable sizes and be represented by relatively homogeneous intensity values and therefore the explored strategies in this thesis are based on region detection which in particular show a robust performance for high resolution SAR images. The scheme here is to detect 'on ground' regions, which are not expected to be affected by very different SAR and optical sensor geometries. The following step selects the extracted regions and prepares them for matching with methods based on cross-correlation and mutual information. Selected datasets for the testing and evaluation of the region based registration methods are pre- and post-disaster (2008 earthquake in Sichuan China) IKONOS-2 images along with a TerraSAR-X image acquired within hours after the disastrous earthquake. The evaluated automated methods in this thesis have successfully detected shift parameters within already orthorectified SAR and optical imagery acquired over dense urban areas.

Kurzfassung

Diese Diplomarbeit befasst sich mit dem Problem der Registrierung von hoch aufgelösten SAR und optischen Satellitenbildern in dicht besiedelten städtischen Gebieten. Seit dem Start von hoch auflösenden Satelliten mit verschiedenen Sensoren, wie TerraSAR-X, WorldView-1 und IKONOS-2, hat der Beitrag der Fernerkundung zu unterschiedlichen Anwendungen stark zugenommen. Speziell die kombinierte Analyse von SAR und optischen Bildern, ist bei der Beobachtung und Einschätzung von Unglücken und Naturkatastrophen von immenser Bedeutung. Obwohl neuere Satelliten georeferenzierte und orthorektifizierte Produkte bereitstellen, haben die Daten der verschiedenen Sensoren immer noch einen Registrierungsfehler. Dieser muss vor der Anwendung für Fernerkundungsaufgaben durch schnelle automatische Verfahren beseitigt werden. Evaluiert wurden verschiedene Methoden zur Registrierung, die auf der Detektion von Objekten basieren wie etwa lokale Merkmalspunkte oder Linien, und Verfahren, die eine Abhängigkeit der Intensitätswerte suchen. Da es sehr aufwendig ist, korrespondierende Objekte aus SAR und optischen Bildern zu extrahieren, wurde ein Verfahren basierend auf Regionen für diese Arbeit gewählt. Hoch aufgelöste Satellitenbilder die über städtischem Gebiet aufgenommen wurden, weisen bei breiten Straßen, Flüssen, großen Stadien, Spielplätzen und Grünanlagen eine entsprechende Größe und relativ homogene Intensitätswerte auf. Aus diesem Grund beruht die in dieser Arbeit untersuchte Strategie auf der Extraktion von Regionen und zeigt besonders bei SAR-Bildern eine hohe Robustheit. Ziel des Verfahrens ist möglichst die Detektion von Bodenregionen, da diese nicht durch die stark unterschiedliche Geometrie von SAR und optischen Sensoren beeinflusst sind. Im folgenden Schritt werden extrahierte Regionen ausgewählt und für den Abgleich mit Mutual Information und Kreuzkorrelation vorbereitet. Für das Testen und Bewerten der auf Regionen basierenden Strategie wurde Datenmaterial des Erdbebens in Sichuan (China) aus dem Jahre 2008 benutzt, IKONOS-2 Szenen aufgenommen vor und nach der Katastrophe und eine TerraSAR-X Szene, kurz nach dem Unglück. Die in dieser Arbeit evaluierte automatische Methode hat erfolgreich die Translationsparameter von orthorektifizierten SAR und optischen Bildern in dichten städtischen Gebieten bestimmt.

1 Introduction

1.1 Problem and Motivation

High resolution remote sensing images have found applications in various fields, like traffic studies, treaty and boarder monitoring, agricultural studies, generation of 3D models and topographic maps, early warning systems, urban growth, damage assessment and disaster mitigation. For these specified tasks, images acquired by both passive and active high resolution spaceborne sensors like Quickbird, ALOS, IKONOS-2, COSMO SkyMed and TerraSAR-X are being utilized with great effect. All the mentioned sensors here are nearly in the same league of spatial resolution, about 1 meter, which is expected to improve further in the near future.
Specifically, images acquired both by the passive optical sensors and active SAR sensors (alone and in combination) are major sources for crisis information management. The SAR sensor's active nature gives them the capability to see through clouds and to acquire images at night which might be the only possible option during a catastrophic event. However, images acquired by SAR sensors have very different characteristics from normally used optical sensor images. On top of the very different geometry (sideways looking and measuring distances) from their optical counterparts (measuring angles), images acquired by SAR sensors show a high amount of speckle influence caused by random backscatter of the microwaves emitted by the active sensor. Normally, remote sensing applications might need to accommodate images from different sensors/modalities, depending upon specific application demands or data unavailability. For example, in case of a natural calamity, decision makers might be forced to use old archived optical data with a newly acquired (post disaster) SAR image.
Remote sensing imagery acquired at the time of a crisis might have different accuracy levels for example the IKONOS-2 standard GEO product has an accuracy of about 100 meters (Meinel and Reder, 2001) while the corresponding TerraSAR-X EEC product provides 1 to 4 meter (accuracy tested by corner reflector) (Nonaka et al., 2008). Combining these two data products for various remote sensing applications like mapping, change detection, fusion and GIS overlays might not yield fruitful results and therefore image-to-image registration becomes really important and an absolute necessity. In general, the SAR sensors using longer wavelengths of the electromagnetic spectrum (for X band wavelength is about 3 cm), which are expected to generate much more accurate images as compared to their counterparts. In general, major factors that are responsible for better product accuracy are the determination of the precise orbit and the ortho-rectification process. Optical sensors using the visible part of the electromagnetic spectrum also have

problems with unrecorded satellite jitter during image acquisition and uncertainties in the design of the CCD array (Ayoub et al., 2008).
Multimodal image-to-image registration specifically involving images from SAR and optical sensors is a complicated task. Attempts at registering SAR and optical images using the interest points (Schwind, 2008) or geometrical features (like contours, buildings, lakes) may not be very successful (Lehureau et al., 2008). The limitations of these techniques are quite understandable as the two sensors involved here have very different sensor radiometry (due to different wavelengths utilized) and geometry of the SAR (sideways looking) and optical sensors (downward looking). Slight success for SAR optical registration has been achieved using the intensity based or the areas based techniques. These employ similarity metrics like mutual information and cluster reward algorithm (Inglada, 2002). These algorithms can handle the radiometric differences between optical and radar images, but in their original form are not able to handle the different sensor geometry, which has great influence in high resolution images acquired over urban areas. (Suri and Reinartz, 2009)
Further we would like to mention that manual registration of high resolution SAR and optical imagery using control points poses some difficulties as it is hard to locate the exact corresponding points in the optical and in the radar image. Figure 1.1 shows two images of the same scene, a radar image (TerraSAR-X) and an optical one (IKONOS-2). Due to fine and very different minute details now available with the high resolution sensors, the idea of finding the conjugate features (for registration) becomes much more challenging than ever before. Keeping in mind the upcoming new challenges and driven by joint application demands of SAR and optical imagery, the development of automatic image-to-image registration techniques especially suitable for high resolution imagery acquired over urban areas is encouraged.

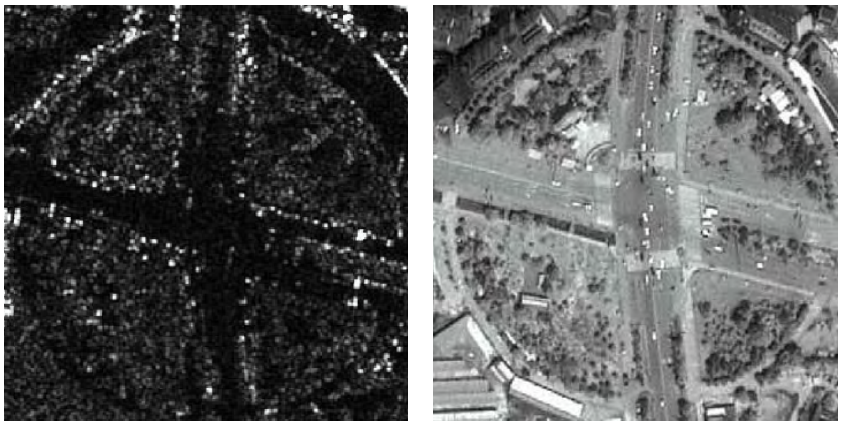


Figure 1.1: TerraSAR-X image (left) and IKONOS-2 image (right)

1.2 Objectives

With the importance of the thesis title discussed in the previous section, the objectives of the presented analysis might be enumerated as:

1. Studying and highlighting the problems encountered in registering high resolution SAR and optical imagery acquired over urban areas.
2. Studying and analyzing the performance of certain standard registration techniques for the explained problem. The knowledge and experience learnt has been further used to develop a concept and related methods which seems to be very useful for registering high resolution imagery acquired over urban areas.
3. Comprehensive testing and critical evaluation of the tested registration strategies on imagery acquired by TerraSAR-X and IKONOS-2 sensors over the city of Sichuan in China.

1.3 Outline

Chapter 1 gives an introduction and explains why co-registration is an important topic.

Chapter 2 has some facts about optical and SAR data along product descriptions available from IKONOS-2 and TerraSAR-X.

Chapter 3 contains definitions and explanations of the tested techniques.

Chapter 4 describes the explored registration strategies.

Chapter 5 includes results from selected datasets and discussion.

Chapter 6 contains a conclusion and scope of future work.

2 Framework Conditions

The main challenge of this thesis is to develop registration strategies for images acquired by very different sensors. There are mainly two categories in space-borne remote sensing sensors, passive sensors in the visible or infrared range of the electromagnetic and active ones in the microwave spectrum. This chapter takes a close look at optical and microwave data and particularly at datasets used for this thesis, namely IKONOS-2 and TerraSAR-X processed image products. Keywords to describe the chief characteristics of the datasets are:

- Space-borne due to the nature of sensors involved
- High resolution due to the volume of data and amount of spatial details available
- Multimodal due to very different sensor modalities
- Urban - images acquired only over urban areas have been considered for analysis
- Disaster mitigation - pre- and post-disaster images have been utilized for treating various registration strategies.

For a thorough understanding of the utilized datasets, the following sections elaborate the sensor geometry and the available satellite image products.

2.1 Sensor Geometry

Analysis and development of multimodal image registration is a tricky task, because each sensor can be at a different position in space and have different intrinsic parameters; objects in the scene cannot be assumed to be located at the same position in each image. Due to the innate sensor effects, corresponding objects in each scene may have different sizes, shapes, intensities and positions. In order to combine the information in each image, it is required that the corresponding objects in the scene be aligned or registered. (Foresti et al., 2003)
Before this section explains the basic sensor geometry of optical and SAR sensors a short digression on radiometry is necessary. Due to the very different wavelengths utilized for imaging, one in the microwave (SAR) and the other in the visible part of the electromagnetic spectrum, radiometry of the acquired images is very different.
SAR image radiometry typically depends upon terrain surface roughness, beam incidence angle, wave polarization, moisture content of target area and size of the scatterer. Further,

due to the coherent name of the SAR sensor images radiometry suffers from multiplicative speckle influence (Elachi, 1988). On the other hand optical sensors, functioning very similarly to the human eye, measure sunlight reflected by the earth's surface. That is why optical images radiometry is very comfortable for human eye visualization. In this case the radiometry is affected by a white additive noise usually modeled by a Gaussian model. Detailed reading can be found in Schwind (2008) and Wegner (2007). Due to the very different radiometry some classical registration methods like least-square matching, NCC (chapter 3.3.1) and sum of absolute differences are not applicable for the task in hand.
After briefly discussing the very different radiometric properties of two sensors we continue our discussion to explain the geometric behavior of the two sensors which play a very dominant role in high resolution images acquired over urban areas.

2.1.1 Optical Sensor

It is interesting for our purpose to understand how an urban area is imaged and how real 3D objects appear in a 2D optical image. This understanding is important for the effectiveness of the selected image registration algorithms. Also post registration analysis would not be possible without appropriate understanding of the sensor geometry. The geometric model shows the behavior of optical imagery figure (2.1). Important factors for optical sensors are satellite and sun position as well as the location and size of an imaged object (3D objects such as buildings are the area of interest here).
There are three major factors influencing 3D object appearances in an optical image:

- As observed in figure 2.1, the sensor incidence angle θ (for IKONOS-2 up to 40°), makes rooftops appear at an incorrect position and overlay other objects, like roads. A possible solution would probably be an ortho-rectification with a digital surface model (DSM). For example, with airborne laser scanning it is possible to deliver data for urban DSM to perform ortho-projection (Weidner, 1997).

- The second important factor is the sun elevation angle. Sun elevation produces shadows which influence the intensity values. One feature can possibly have two different intensity value bands. For example a road with an average gray value of 100 in the sunlight might show a value of 50 in the influence of a shadow.

- Another point is the general height of man-made objects in urban areas. With increasing height of the buildings the two mentioned factors gain in weight and lead to a situation where no ground features are imaged properly.

After the sensor geometry influence, the two other major factors that limit absolute image geometry accuracy of finished optical products are unrecorded satellite jitter during image acquisition and uncertainties on the CCD array geometry. The dimensions of these effects are very small and only perfect sensor calibration seem to be a solution for this problem. (Ayoub et al., 2008)

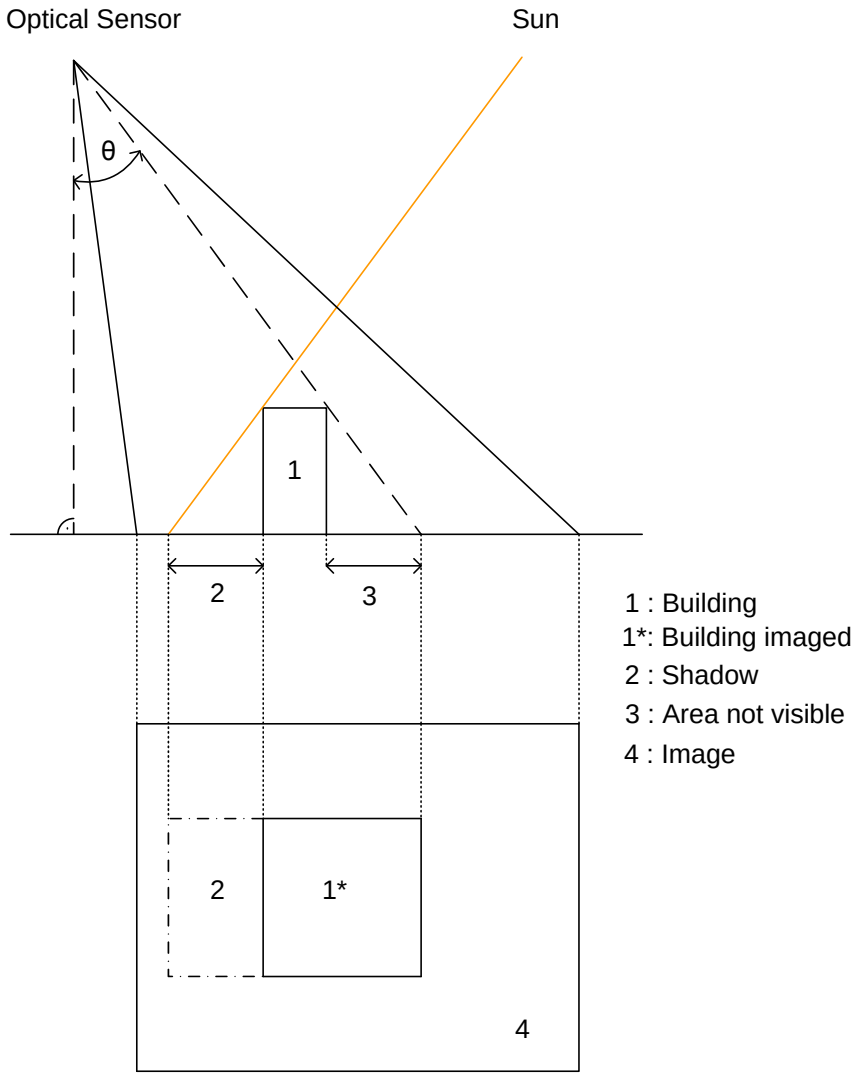


Figure 2.1: Imaging of a building by an optical sensor

2.1.2 Synthetic Aperture Radar

Synthetic Aperture Radar uses microwaves to generate information about the earth's surface. Here we assume that the reader is familiar with basics of SAR imaging. For more detailed information interested readers are referred to Wegner (2007).
Figure 2.2 shows the geometric characteristic of SAR, 2.2(a) for slant image and 2.2(b) for ground range image. For registration application only ground range images are preferred, due to their perpendicular view. Characteristic properties of SAR images are layover, shadow, multi-path signals and speckle (Stilla, 2007). All these phenomena make a SAR image more difficult to interpret; not only for a human interpreter, but for automated methods as well. Generally, the side view of SAR makes objects appear at positions that differ from their positions in the real world. There is a displacement of the elevated 3D objects. In urban areas, façades of buildings or higher objects are affected by the layover

phenomenon (figure 2.2(b) R1). Urban areas provide another common problem, termed as the radar shadow. It is in some ways similar to shadow in the optical sense. To explain this we consider an example of a SAR sensor with an incidence angle of approximately 45°, this means that a building of 5 meter height has 5 meter shadow behind itself in range direction (r_g). This effect is depicted and labeled in figure 2.2(a) and (b).
The very different sideways looking SAR geometry makes it difficult for specific application like urban area road extraction and also image-to-image registration.

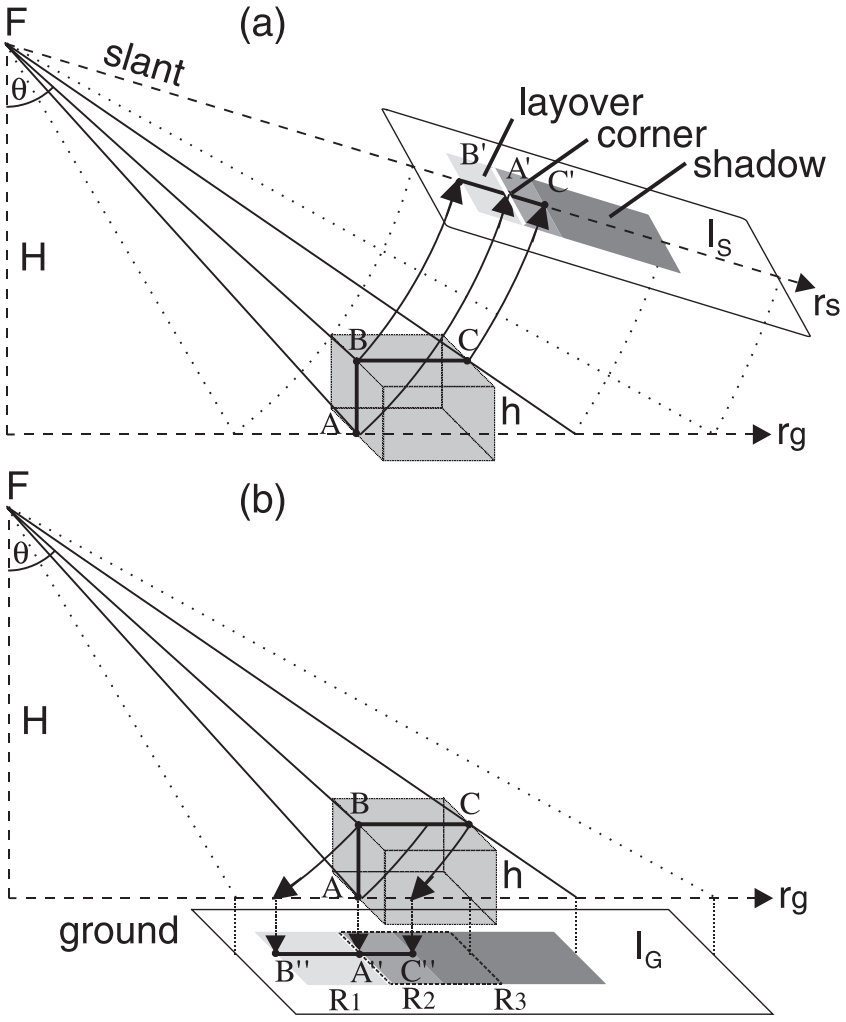


Figure 2.2: Projection of a building into (a) slant image (I_S) and (b) ground image (I_G) (Stilla et al., 2003)

The explanation of the sensors and their geometries thus being accomplished the properties of the utilized IKONOS-2 and TerraSAR-X products are now discussed.

2.2 IKONOS-2

IKONOS-2 is a commercial earth observation satellite, launched in September 1999. The operating company is GeoEye Incorporated, but the European part of the business is handled by European Space Imaging (EUSI). In table 2.1 basic facts about IKONOS-2 are shown and the further section focuses especially on the derived products and auxiliaries. With a ground resolution of about 1 m it is a high resolution space-borne sensor. Special emphasis is on the geocoding accuracy because it is important for different applications. Imagery products supplied by EUSI are listed in table 2.2. Space Imaging categorizes the imagery products to a positional accuracy, which is determined by the reliability of an object in the image being within the specified accuracy of the actual location of the object on the ground. Within each product, the location error is defined by a circular error at 90% confidence (CE90), which means that the locations of the objects are represented on the image with the stated accuracy 90% of the time. This CE 90 accuracy level can also be related to Root Mean Square Error (RMSE).
The GEO products seem to be quite accurate at the first glance, but GEO products are processed without a DEM as shown in table 2.2. Depending upon the quality of the utilized DEM, these products can have a geocode error up to 100 meters (Meinel and Reder, 2001).

Table 2.1: IKONOS-2 basics (Kramer, 2002)

IKONOS-2 ON A GLANCE	
Size	1.83 m x 1.57m
Bands	0.45 - 0.90 μm (panchromatic) 0.45 - 0.52 μm (blue) 0.52 - 0.60 μm (green) 0.63 - 0.69 μm (red) 0.76 - 0.90 μm (NIR)
Resolution	1m (panchromatic) 4m (multispectral)
Power consumption	1500 W
Data storage	64 Gbit
Data transmission	320 Mbit/s X-Band Downlink
Orbit	681 km 98,1° Inclination Sun-synchronous
Repetition rate	14 days (max)
Life time	7 years but still operating

Table 2.2: IKONOS-2 Products (European Space Imaging, 2009)

Products	Product name	Accuracy (CE90/RMSE)	Orthocorrected	Prices per SQ.KM
Geo	Geo	15 m[1]/ N/A	No	20 $
	Geo Archive	15 m[1]/ N/A	No	16 $
Ortho	Standard	50 m / 23.3 m	Yes	26 $
	Reference	25 m / 11.8 m	Yes	30 $
	Pro	10 m / 4.8 m	Yes	32 $
	Precision	4 m / 1.9m	Yes	Call for quote
	Precision Plus	2 m / 0.9 m	Yes	Call for quote

[1] Exclusive of terrain effects

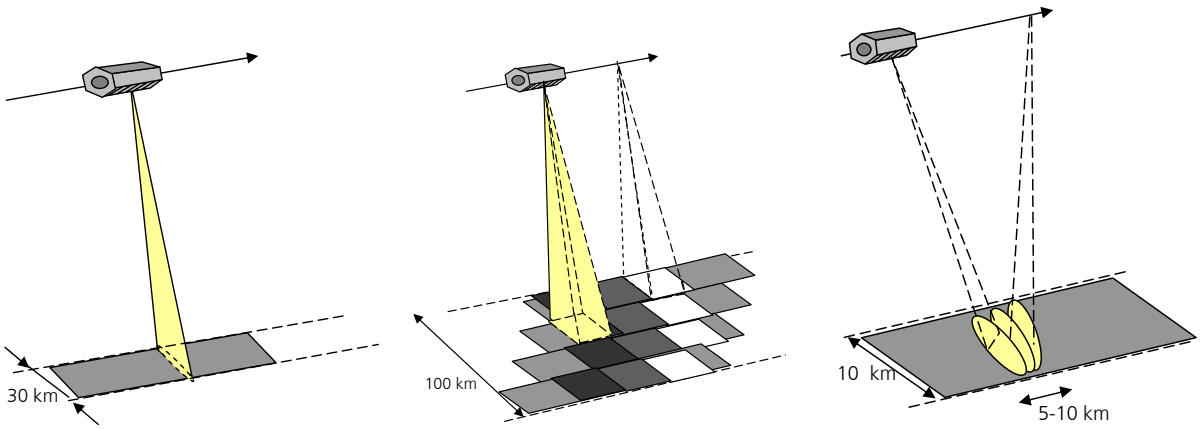
2.3 TerraSAR-X

TerraSAR-X is a German high resolution Earth observation satellite, launched in June 2007. It uses microwaves in the X-Band ($\lambda = 31$ mm) to gather topographic information of the earth surface. Table 2.3 shows the basic facts about the satellite. TerraSAR-X has four nominal imaging modes with different resolutions (figure 2.3) (Eineder and Fritz, 2008):

- StripMap (SM)
  The ground swath is illuminated with a continuous sequence of pulses while the antenna beam is pointed to a fixed angle in elevation and azimuth.
- ScanSAR (SC)
  An electronic antenna elevation steering is used to switch after bursts of pulses between swathes with different incidence angles. In the designed Terra SAR-X ScanSAR mode four stripmap beams are combined to achieve a 100 km wide swath.
- SpotLight (SL)
  This mode uses electrical beam steering in azimuth direction to increase the illumination. The larger aperture results in a higher azimuth resolution at the cost of azimuth scene size.
- High-Resolution SpotLight (HS)
  It is a slower sliding spotlight than SL and the azimuth resolution is enhanced.

Derived from each one of these modes the following basic products are available.

- Single-look complex products in slant range projection (SSC)
- Multi-look detected products in ground range projection (MGD)
- Geocoded ellipsoid corrected products in UTM/UPS projection (GEC)
- Enhanced ellipsoid corrected products in UTM/UPS projection (EEC)



(a) StripMap mode (b) ScanSAR mode (c) SpotLight mode

Figure 2.3: TerraSAR-X modes (Eineder and Fritz, 2008)

The basic products for commercial use are distributed by the Infoterra GmbH which is a wholly-owned subsidiary of EADS Astrium company. All TerraSAR-X products are produced in 16 bit depth. The standard delivery file format is GeoTIFF and the annotation information is provided in XML format. It is noteworthiness that all products used for science are available much cheaper. The program is called COFUR (Cost of Fulfilling User Request) and the distribution of scientifically used products is handled by DLR. For further information see Eineder and Fritz (2008).
For image-to-image registration scenarios, normally the image with higher planar accuracy should be utilized as the reference image. For the present utilized sensors, TerraSAR-X products are expected to be more accurate and henceforth have been utilized as reference image. Parameters that influence geocoding accuracy of TerraSAR-X images mainly are (Schmidt et al., 2008):

- orbit accuracy
- ellipsoid or DEM for ortho-projection

The three different orbit accuracies shown in table 2.5 are used for the production. It becomes clear that the rapid orbit, which is standard for the processing of basic products, has an accuracy which matches the results of Nonaka et al. (2008). One result was a corner reflector measurement based on a EEC product with rapid orbit, that offers a location error of 3.6 meters in an easterly direction and 1.0 meter in a northerly direction. So referring to these results TerraSAR-X images are highly suited to act as reference image in the co-registration process.
In table 2.6 it is shown how sensitive geocode accuracy is to elevation of terrain. This is the reason why EEC products present such a high geometric accuracy. Otherwise GEC products are only rectified with the WGS84 ellipsoid and not with a DEM. This is the reason for the reduced geometric accuracy of GEC products. The GEC products are the recommended product for flat area, marine or coastal applications where topography does

not affect the location accuracy (Roth et al., 2008).

Conclusion

This chapter, beginning from basic radiometric properties of SAR and optical sensors, explained in detail the geometric properties of two sensors. The geometry of the two sensors plays an important role in high resolution registration applications and so was discussed in detail. We have also elaborated on various TerraSAR-X and IKONOS-2 data products and some have been utilized for further analysis.

Table 2.3: TerraSAR-X basics (TerraSAR-X, 2009a)

TerraSAR-X on a glance	
Height	5.0 m
SAR Antenna	4.8 m x 0.7 m x 0.15 m
Resolution	1m @ 5 x 10 km Scene
Power consumption	800 W average
Data storage	256 Gbit
Data transmission	300 Mbit/s X-Band Downlink
Orbit	514 km 98° Inclination Sun-synchronous
Repetition rate	11 days
Life time	5 years

Table 2.4: TerraSAR-X Basic Image Product Categories (TerraSAR-X, 2009b)

Basic Image Product Category	Standard Scene (Width x Length)	Processing level	Resolution
High Resolution SpotLight	10 km x 5 km	SSC, MGD, GEC, EEC	up to 1 m
SpotLight	10 km x 10 km	SSC, MGD, GEC, EEC	up to 2 m
StripMap	30 km x 50 km	SSC, MGD, GEC, EEC	up to 3 m
ScanSAR	100 km x 150 km	MGD, GEC, EEC	up to 18 m

Table 2.5: TerraSAR-X Orbit Typs (Eineder and Fritz, 2008) (Schmidt et al., 2008)

ORBIT TYPE	predicted	rapid	science
ACCURACY	700 m along track	2 m (3D, 1 σ)	20 cm (3D, 1 σ)
AVAILABILITY	Near real time case	Available latest 15 hours after last satellite contact	Operational life showed the availability after 5 to 7 days after acquisition
PURPOSE	Processing of near real time products	Standard processing of basic products	Processing of products from archive if orbit type is available, e. g. for interferometry

Table 2.6: DEM effect on the pixel location accuracy (Eineder and Fritz, 2008)

	20°	26°	32°	38°	44°	50°	Incidence angle
DEM elevation error	2.75	2.05	1.60	1.28	1.03	0.83	displacement factor
2 m	5.5	4.1	3.2	2.6	2.1	1.7	Resulting location error in meters
6 m	16.5	12.3	9.6	7.7	6.2	4.9	
8 m	22	16	13	11	9	7	
16 m	44	33	25	21	17	13	
30 m	82	61	48	38	31	25	
100 m	275	205	160	128	103	83	

3 Theoretical Background

Registration of remote sensing images has been an active field of research for decades (Shan and Stilla, 2008) (Wong and Clausi, 2006). Basically, registration is the process of finding a geometrical transformation between two or more images such that corresponding image structures are aligned properly. These images might have been acquired with the same or different imaging modalities and at the same or different times. There are various applications in diverse fields that need image-to-image registration as a preprocessing step. The most important fields are computer vision and pattern recognition, medical image analysis and remote sensing image processing. For further reading, overview and survey of common image registration methods can be found in Maintz and Viergever (1998), Brown (1992), Zitová and Flusser (2003) and Modersitzki (2004).
The problem of registering input image (I_I) to a reference image (I_R) can be mathematically expressed as (Brown, 1992):

$$I_R(x, y) = g(I_I(T(x, y))) \tag{3.1}$$

where g is a 1D intensity or radiometric interpolation function and T is a transformation function which maps two spatial coordinates x and y, to the new spatial coordinates x' and y'.

$$(x', y') = T(x, y) \tag{3.2}$$

The main objective of the entire process is to estimate the spatial transformation T.
In remote sensing accurate image registration is a prerequisite for bad or even non-geocoded remote sensing images, to proceed with various applications, like fusion, change detection or GIS overlay. Different ways are possible to succeed in registering images automatically. Some common methodologies will be presented in this chapter. These methods are sub-categorized into feature based and intensity based techniques. Certain preprocessing tasks might be inevitable before performing image-to-image registration and these have also been discussed.

3.1 Preprocessing Methods

Preprocessing is performed to improve the image in respect of main operation.

3.1.1 Smoothing

Smoothing is an operation targeted to reduce noise in an image. Certain basic filters that perform smoothing are Mean Filter, Median Filter and Gaussian Filter. Further, there are also adaptive filters that self-adjust their transfer function according to an optimization algorithm. Adaptive filtering uses the standard deviation of the pixels within a local box surrounding each pixel to calculate a new pixel value. Typically, the original pixel is replaced with a new value calculated based on the surrounding valid pixels (those that satisfy the standard deviation criteria). Unlike a typical low-pass smoothing filter, the adaptive filters preserve image sharpness and detail while suppressing noise. More information about statistical filters can be found in Hayes (1996). Popular adaptive filters are the Lee Filter (Lee, 1980) or Frost Filter (Frost et al., 1982). These filters are mainly used for speckle reduction in SAR images.
Here we discuss the Gaussian Filter that has been applied to optical imagery and Enhanced Frost Filter for SAR imagery in the analysis presented in this thesis.

Gaussian Filter

The equation for the Gaussian function is:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp(-(x^2+y^2)/2\sigma^2) \tag{3.3}$$

A Gaussian 5x5 filter with $\sigma = 1.4$ looks like this:

$$G(x,y) = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \tag{3.4}$$

Enhanced Frost Filter

The Frost Filter is an exponentially damped circularly symmetric filter that uses local statistics. The pixel being filtered is replaced with a value calculated based on the distance

from the filter center, the damping factor, and the local variance. The Enhanced Frost Filter is an adoption of the Frost filter and similarly uses local statistics (coefficient of variation) within individual filter windows. Each pixel is categorized into one of three classes mentioned below (Lopes et al., 1990):

- Homogeneous: The pixel value is replaced by the average of the filter window.
- Heterogeneous: An impulse response is used as a convolution kernel to determine the pixel value.
- Point target: The pixel value is not changed.

Parameters that have to be set are:

- Filter size (row x column e. g. 3x3)
- Damping Factor that irreversibly defines the extent of the exponential damping in the weighted average for the heterogeneous class. A larger damping factor produces less averaging.
- Values that define the class cut-offs for the homogeneous (coefficient of variation $\leq C_u$), heterogeneous ($C_u >$ coefficient of variation $< C_{max}$), and point target (coefficient of variation $\geq C_{max}$) classes.
  Estimation of the cut-off values based on the number of looks (L) of the radar image.

$$C_u = \frac{0.523}{\sqrt{L}} \tag{3.5}$$

$$C_{max} = \sqrt{1 + \sqrt{\frac{2}{L}}} \tag{3.6}$$

3.1.2 Clipping

The purpose of clipping is to confine the range of gray values of an image to specific boundaries. It is used to convert an image to a lower bit depth or simply to stretch it to a significant gray value range. The operation may be called histogram stretching, since the gray level histogram reveals the overall appearance in the image. The transformation function $l = T(m)$ (where m is input gray level and l is the output gray level) must satisfy the following conditions (Chanda and Majumder, 2006):

1. $T(m)$ must be monotonically increasing in the interval $[m_{min}, m_{max}]$, i.e.

   $$m_1 < m_2 \Rightarrow l_1 = T(m_1) \leq l_2 = T(m_2)$$

   meaning that the transformed gray level l must preserve the order from black to white.

2. $l_{min} \leq l \leq l_{max}$, i.e. transformed gray level must lie within the allowed range of gray levels.

where $[l_{min}, l_{max}]$ and $[m_{min}, m_{max}]$ are the available gray level range in the given image, respectively.
In the case of a simple linear stretching the transformation function can be represented as

$$l = T(m) = \frac{l_{max} - l_{min}}{m_{max} - m_{min}}(m - m_{min}) + l_{min} \tag{3.7}$$

This transformation function shifts and stretches the gray level range of the input image to occupy the entire dynamic range $[l_{min}, l_{max}]$. The transformation may also be written as

$$\tilde{g} = \frac{l_{max} - l_{min}}{m_{max} - m_{min}}\{g(r, c) - m_{min}\} + l_{min} \tag{3.8}$$

Clipping is a possible preprocessing method for road extraction from SAR images mentioned in Wessel (2006) and can be useful to reduce the bit depth to gain performance for further processing.

3.2 Feature Based Methods

Feature based methods try to find correspondence between image features, such as points, lines and contours (Goshtasby, 2005). The following subsections explain feature based methods that have been explored in this thesis. Subsections are about detection of interest points, of edges and of regions.

3.2.1 Interest Point Detection

There are a number of intensively proven interest point detectors, like the Förstner Operator (Förstner and Gülch, 1987), Harris Operator (Harris and Stephens, 1988) or SIFT Operator (Lowe, 1999). The last-mentioned of these is currently state of the art in research and as such has also been suggested for use with SAR images (Wessel et al., 2007). Hence only the SIFT Operator is discussed here. David Lowe introduced the Scale Invariant Feature Transform (SIFT) in 1999. A few years later he updated his algorithm. (Lowe, 2004)
The major processing stages of the SIFT processing chain are (Schwind, 2008):

1. Scale space extrema detection: Using the Difference of Gaussian (DoG) function, a scale space is built, out of which extrema are detected. A DoG image is given by

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma) \tag{3.9}$$

$$L(x, y, k\sigma) = G(x, y, k\sigma)I(x, y) \tag{3.10}$$

   Where $I(x, y)$ is the original image and $G(x, y, k\sigma)$ is the Gaussian function.

2. Key point localization: The found extrema are filtered by different criteria to select only key points with high stability.

3. Orientation assignment: By analyzing the local image gradient direction, at least one orientation is assigned to every key point.

4. Key point descriptor: Making use of the location, scale and rotation of the key points, descriptors are created

The SIFT operator offers a very successful method for registering optical to optical images and multi sensor SAR images (Schwind et al., 2009), but it is not capable of handling optical to SAR matching scenarios. Implementation, evaluation and more details can be found in Schwind (2008).

3.2.2 Edge Detection

Edge detectors aim to detect lines, where the image brightness changes or high frequencies appear. The following presents the Canny edge detector for optical images and the Steger operator for SAR images.

Canny Edge Detector

The Canny edge detector is quite a popular and robust edge detector. It was developed 1986 by John Canny (Canny, 1986). The processing steps in detail are:

1. Noise Reduction
   The image has to be smoothed, to make the algorithm more robust.
   A possible filter could be

$$A = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{3.11}$$

The smoothed image is

$$I_2(x, y) = I_1(x, y)A \tag{3.12}$$

2. Finding the intensity gradient of the image
   The image gets convolved with a typical edge detection operator (e. g. Sobel operator (Sobel and Feldman, 1968)). The operator returns a value for the first derivative in the horizontal direction S_y and the vertical direction S_x.

$$S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{3.13}$$

$$S_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{3.14}$$

   Based on this the edge gradient S and direction Θ can be determined:

$$S = \sqrt{S_x^2 + S_y^2} \tag{3.15}$$

$$\Theta = \arctan \sqrt{\frac{S_y}{S_x}} \tag{3.16}$$

   A pixel has 8 neighbors and results in only four edge directions. Therefore the edge direction angle is rounded to 0, 45, 90, 135 degrees.

3. Non-maximum suppression
   In this step the edges are set to a width of one pixel. Every pixel which belongs to an edge doesn't have a neighbor pixel, except in edge direction.

4. Tracing edges through the image and hysteresis thresholding

Finally there has to be a rule to classify intensity gradients as edge or as non-edge. This is realized by hysteresis thresholding, which requires two thresholds T1 and T2 ($T1 < T2$). A value higher than T2 is selected and beginning with this value you follow the edge in both directions. Every value that is higher than T1 represents an edge.

The canny edge detector has been used successfully many times to extract edges from optical images (Lindeberg, 1998).

Steger Operator

The line extraction is done by using partial derivatives of a Gaussian smoothing kernel (equation 3.3) to determine the parameters of a quadratic polynomial in x and y for each point of the image. The parameter σ determines the amount of smoothing to be performed. The parameters of the polynomial are used to calculate the line direction for each pixel. Pixels which exhibit a local maximum in the second directional derivative perpendicular to the line direction are marked as line points. The line points found in this manner are then linked to contours. This is done by immediately accepting line points that have a second derivative larger than a parameter, and points that have a lower value than the lower threshold parameter are rejected. All other line points are accepted if they are connected to previous accepted points by a connected path. (Halcon Reference Manual, 2007)
For a detailed mathematical description please consult Steger (1996).
Steger Operator has been successfully used in Wessel (2006) to extract roads in non-urban SAR images.

3.2.3 Region Detection

A common technique to detect regions is adapted from the region growing family, which graduate images into regions of the same intensity. The input image is rastered into rectangles of different sizes. In order to decide whether two adjacent rectangles belong together only the gray value of their center points is used. If the gray value difference is less than or equal to a given tolerance the rectangles are merged into one region. For example if g1 and g2 are two gray values being considered, they are merged into the same region if:

$$|g1 - g2| < T \tag{3.17}$$

where T is the tolerance (a user defined parameter).
Only regions that contain a minimum of pixels are returned. The method is visualized in figure 3.1 which shows an image with a size of 9x9 pixels. First, a continuous clustering

with size 3x3 pixels starts at the upper left corner. In the next step, the arithmetic middle is calculated and this value of each cluster can be compared to the values in the 4-cell von Neumann neighborhood. If the difference of those neighbor values is less or equal to the tolerance parameter, the clusters are merged to one region.
Region growing offers several advantages over conventional segmentation techniques. The algorithm is very stable with respect to noise. Regions are expected not to have noise and occlusions as long as the parameters are defined correctly. (Castleman, 1996)

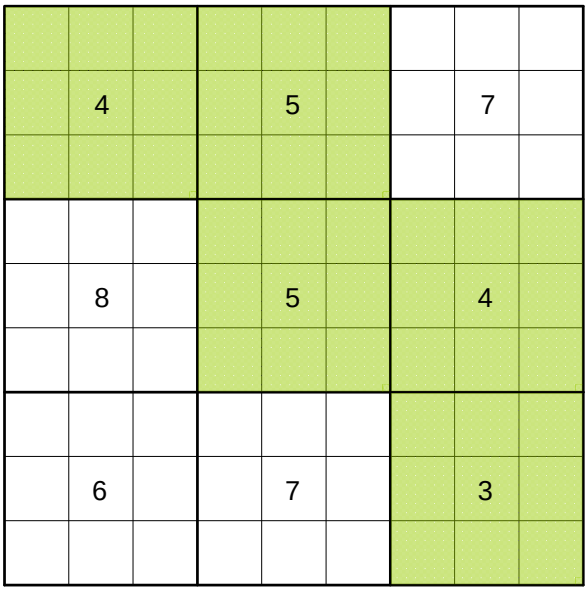


Figure 3.1: Region detection

The result of region growing is a binary image which has been subjected to techniques like skeleton extraction and distance map formation in this thesis.

Skeleton Extraction

This operation takes the medial axes of regions. The idea is that each point on it can be seen as center point of a circle with the largest possible radius (like figure 4.5 in the next chapter) while still being completely contained in the region. (Eckardt, 1988)

Distance Map

For each pixel the distance map operator assigns a number that is the euclidean distance between that pixel and the nearest nonzero pixel. $I : \Omega \subset \mathbb{Z}^2 \rightarrow \{0, 1\}$ is a binary image where the domain Ω is convex and, in particular, $\Omega = \{1, \ldots, n\}$ x $\{1, \ldots, n\}$. By convention, 0 is associated with black, and 1 with white. Hence we have an object A represented by all the white pixels:

$$A = \{p \in \Omega | I(p) = 1\} \tag{3.18}$$

The distance transform is the transformation that generates a map D whose value in each pixel p is the smallest distance from this pixel to A_C. With A_C complement of A that means all black pixels:

$$D(p) := min\{d(p,q)|q \in A_C\} = min\{d(p,q)|I(q) = 0\} \tag{3.19}$$

For more information and a comparative survey of distance map transformations consult Fabri et al. (2008). One common algorithm from Breu et al. (1995) is based on the construction of a Voronoi diagram.

3.3 Intensity Based Methods

Intensity based techniques register images largely on the basis of correlation between the intensity values of the pixels of two images (Inglada, 2002). This makes the registration process to emphasize on feature matching rather than on their detection.

3.3.1 Cross-correlation

The cross-correlation in spatial domain with image $f(x, y)$ and correlation coefficient Φ is defined as:

$$\Phi(x,y) = \sum_{k-a}^{a} \sum_{l-b}^{b} (f(x-k, y-l)g(k,l)) \tag{3.20}$$

Where g(k,l) is the template with indexes for rows from -a to a and indexes for columns from -b to b.
Realization of normalization:

$$\Phi_N(x,y) = \Phi(x,y)\sqrt{Z_f Z_g} \tag{3.21}$$

$$Z_f = \sum_{k-a}^{a} \sum_{l-b}^{b} f^2(x-k, y-l) \tag{3.22}$$

$$Z_g = \sum_{k-a}^{a} \sum_{l-b}^{b} g(k,l)^2 \tag{3.23}$$

Cross-correlation requires a linear relationship between the corresponding intensity values of the images that should be registered. Therefore it can only be utilized to register images with similar modalities.

3.3.2 Mutual Information

Mutual Information (MI) is an information theoretic approach. MI describes a statistical dependence between two random variables (e. g. A and B) expressed in terms of variable entropies:

$$MI(A,B) = H(A) - H(A|B) \tag{3.24}$$
$$MI(A,B) = H(B) - H(B|A) \tag{3.25}$$
$$MI(A,B) = H(A) + H(B) - H(A|B) \tag{3.26}$$

where $H(A)$ and $H(B)$ are the marginal entropies, the so called Shannon entropies (Shannon, 1948) of A and B respectively, $H(A,B)$ is the joint entropy and $H(A|B)$ is the conditional entropy of B given A. Considering an optical image and a radar one to be registered as two random variables, MI is a symmetric relation that always achieves values greater than zero. Registration of two images A and B is based on maximization of $MI(A,B)$.
The marginal entropies and the joint entropy are computed from Chen et al. (2003).

$$H(A) = \sum_{a} -p_A(a) log \cdot p_A(a) \tag{3.27}$$
$$H(B) = \sum_{b} -p_B(b) log \cdot p_A(b) \tag{3.28}$$
$$H(A,B) = \sum_{a,b} -p_{A,B}(a,b) log \cdot p_{A,B}(a,b) \tag{3.29}$$

Here $p_A(a)$ and $p_B(b)$ are the marginal probability mass functions and $p_{A,B}(a,b)$ is the joint probability mass function. These probability mass functions can be obtained from:

$$p_{A,B}(a,b) = \frac{h(a,b)}{\sum_{a,b} h(a,b)} \tag{3.30}$$

$$p_A(a,b) = \sum_b p_{A,B}(a,b) \tag{3.31}$$

$$p_B(a,b) = \sum_a p_{A,B}(a,b) \tag{3.32}$$

Here h is a joint histogram (JH) of the two images involved. It is a 2D matrix with the intensity values of the other image along the other axis. The value $h(a,b)$ is the statistic number of corresponding pairs having intensity value a in the first image and intensity value b in the second image. Instead of equation 3.26, Studholme et al. (1999) proposes the use of this better conditioned variant:

$$MI(A,B) = \frac{H(A) + H(B)}{H(A,B)} \tag{3.33}$$

It is a normalized MI implementation to reduce the sensitivity of MI on changes in the overlap. MI though being computationally expensive has been found useful for registering SAR and optical images (Suri and Reinartz, 2008). Another metric with similar properties is the Cluster Reward Algorithm (Inglada, 2002).

Conclusion

Up to now the selected techniques have been discussed to explain the theoretical part of registration strategies tested in this thesis. Their implementation details are presented in the following chapter.

4 Explored Registration Strategies

In this chapter we detail our explored registration strategies. Figure 4.1 shows the workflow and the following sections explain the implementation and the characteristics of each and every individual workflow component.

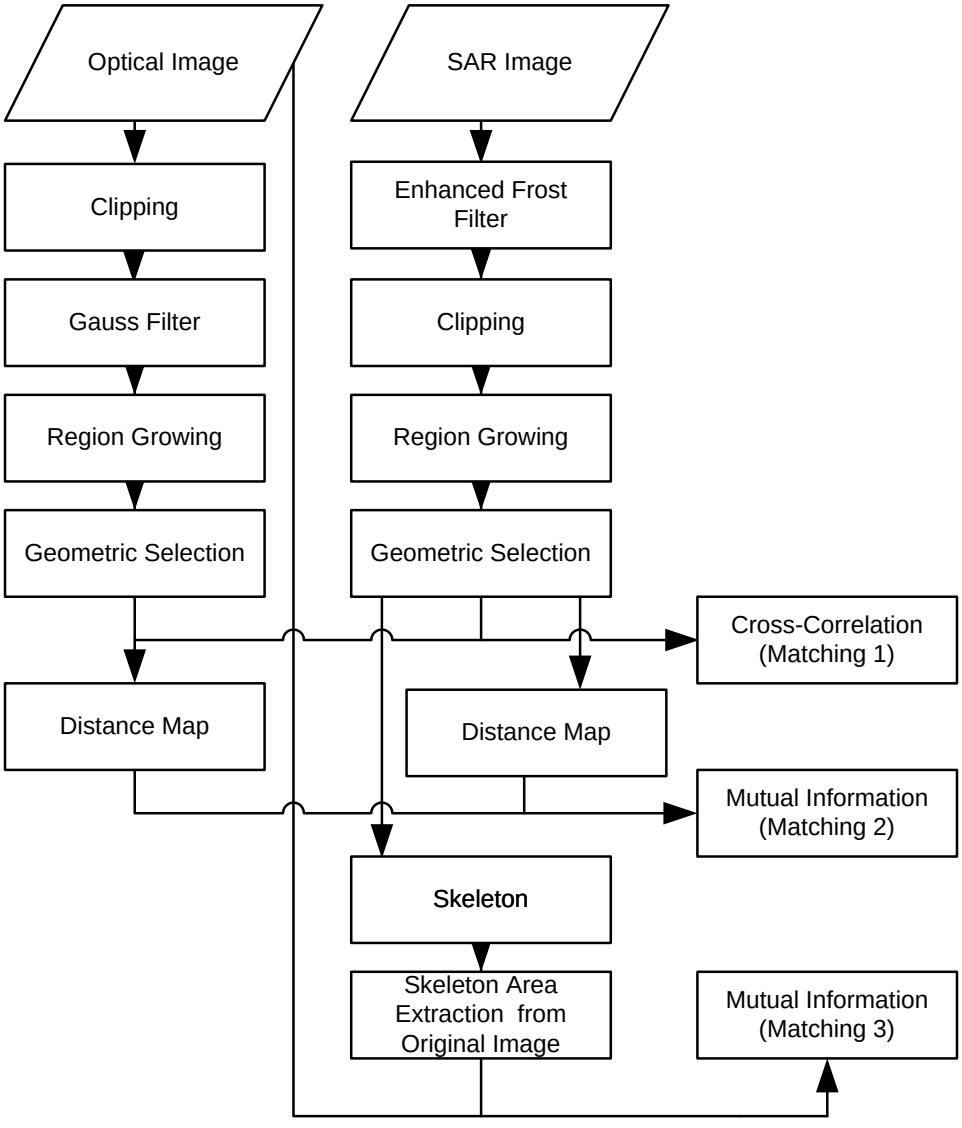


Figure 4.1: Workflow

It can be observed from figure 4.1 that the region growing technique lies at the heart of the three different matching techniques. To implement individual components of the workflow, we have utilized the following software packages:

- ENVI 4.4 by ITT Visual Information Solutions
- HDevelop 8.0 by MVTec Software GmbH
- Matlab R2008b by Mathworks
- Xdibias I by Department Photogrammetry and Image Analysis at the German Aerospace Center (DLR)

4.1 Preprocessing

Preprocessing is done for optical and SAR images, to clear the way for further steps. ENVI implementation of both the smoothing and the clipping operation have been utilized in the presented work.

4.1.1 Smoothing

Optical and SAR images have been smoothed by using a Gaussian filter and an adaptive Enhanced Frost Filter of size 7x7 respectively. The main idea of using the Frost filter is not speckle reduction but to facilitate homogeneous regions detection which is a prerequisite for the three matching strategies.

4.1.2 Clipping

The idea of the clipping technique is to utilize only that portion of the image histogram which contains most of the image content. IKONOS-2 images have 11 bit meaning the gray value of every pixel is between 0 and 2047. However the full range of gray values is never really utilized. For example IKONOS-2 image (figure 4.2(a)) have an average gray value of approximately 450 and most of the image information content is concentrated in a very narrow band. To start off the procedure we first define boundaries surrounding the maximum image information. For figure 4.3 the boundaries are 314 and 579; values outside the thresholds are clipped and not considered for further analysis. Values in between (here 265) are stretched linearly, to fit in the 8 bit space (0 to 255).

As can be seen in figure 4.4, the same technique has been applied to the 16 bit TerraSAR-X image (figure 4.2(b)). The aim in the next processing step (region growing) is to detect large homogeneous areas on the ground. Due to the SAR characteristic these areas are mostly represented by dark pixels. Hence extremely bright pixels are not required. The values of the upper and lower threshold for clipping are zero and 300. This

(a) IKONOS-2 11 bit original data

(b) TerraSAR-X 16 bit after Enhanced Frost Filter

Figure 4.2: Original data for histogram in figure 4.3 and 4.4

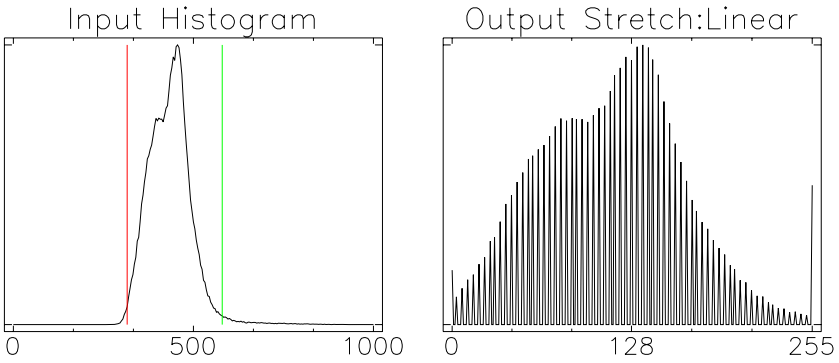


Figure 4.3: Histogram IKONOS-2 11 bit and 8 bit

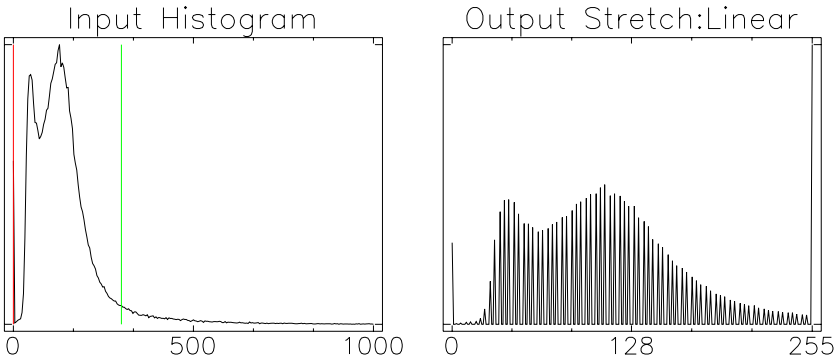


Figure 4.4: Histogram TerraSAR-X 16 bit and 8 bit

results in a lot of 255 values, because all values above upper the threshold become 255. However as mentioned before, these values are not needful.

4.2 Region growing

The region growing algorithm (chapter 3.2.3) has been implemented using HDevelop. The optimal parameters can vary for different images and it can be generally said that setting the optimal parameters would definitely improve the results. But one fundamental goal in this project is to generate a fully automatic method and therefore only the following fixed parameters are used.
Parameters for the SAR and the optical image are:

- Size of rectangle (row x column) = 3x3
- Tolerance = 6 (SAR) / 8 (optical)
- Minimum size of every region = 150 pixels

Usually it has been observed that optical images need a higher tolerance value because roads or rivers appear not as homogeneous as they appear in the corresponding SAR image. Output of this step is a region array containing detected regions. (sample results are shown on page 43 and 51)

4.3 Geometric selection

As registration using regions is only possible with on ground features, geometric selection is important. The expression geometric selection stands for a few consecutive algorithms with the intention to filter out desired regions. Again HDevelop implementation has been utilized for this region filtering step. An overview of the processing steps is given thereafter:

1. Inner circle
   Figure 4.5 shows a segment of pixels, the biggest circle that fits in this segment is the orange one with radius r. If the radius r is smaller than a parameter in pixels, the region is rejected. The advantage is that very small disturbing regions like shadows of buildings can be reduced.

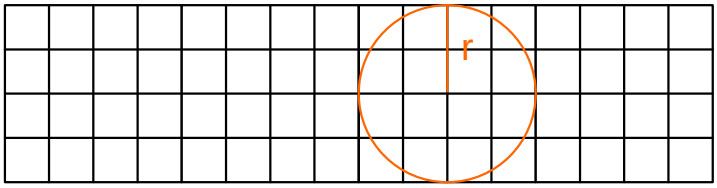


Figure 4.5: Inner circle

2. Outer circle
   The operator determines the smallest surrounding circle of a region as figure 4.6

shows. If the radius R is smaller than a parameter in pixels, the region is rejected. The advantage is that a minimum length of the regions can be determined and regions that are not long enough are rejected. Finally only regions with a desired length remain which can be assumed to be part of a ground feature.

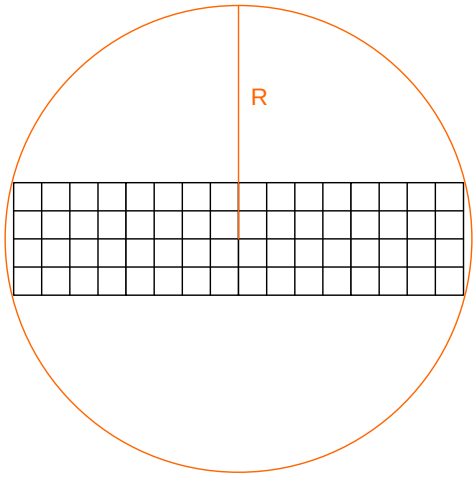


Figure 4.6: Outer circle

3. Anisometry
   Figure 4.7 shows the pixel segment surrounded by an ellipse. The operator determines the quotient of the semimajor axis (a) and the semiminor axis (b). Mathematically expressed: $Z = \frac{a}{b}$. If the value is lower than the parameter(Z), the region is rejected with the advantage that only oblong regions remain.

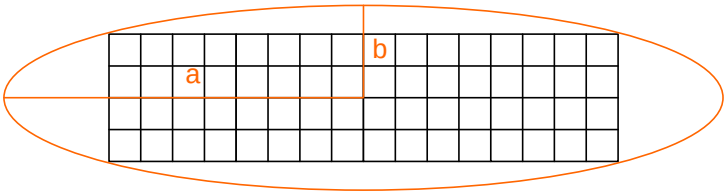


Figure 4.7: Anisometry

The most important step in this part is anisometry. It enables the exclusion of huge bulky blobs which are very prominent in the SAR region image. (sample results can be visualized on page 44) In general optical imagery has been found to be much more sensitive to different parameter settings. SAR images for example are not much affected by different pavements. Regions in SAR images are larger and better connected, on the other hand regions in optical imagery are more fragmented.

4.4 Matching

The region arrays are now transformed to binary images. Pixel values are set to one for region and zero for background. The following presents and evaluates three different methods of performing the matching of the SAR and the optical region. Results are the translation parameters. Rotation has not been considered because datasets used in this thesis are already orthorectified to the same coordinate projection system (UTM).

4.4.1 Binary Cross-Correlation

The following lines explain the functional steps of the HDevelop cross-correlation algorithm. Explicitly used HDevelop functions are 'create_ncc_model', 'set_ncc_model_origin' and find_ncc_model'.

1. 'Create_ncc_model' takes a chip of the SAR image and loads it into memory. Due to the low robustness of cross-correlation, the possibility of down sampling and using pyramid layers is not applicable for this purpose. A rotation of the pattern is not necessary due to the almost perfect north orientation of the input images.

2. The next step is the function 'set_ncc_model_origin'. The final output translation parameters should be column (c) and row (r). With the new UTM coordinates ($UTMIKO_{NEW}$) and the old UTM coordinates ($UTMIKO_{OLD}$):

   $$UTMIKO_{NEW}(x, y) = UTMIKO_{OLD}(x - c, y + r)$$

3. Finally the function 'find_ncc_model' finds the best match of the model and returns translation parameters for row and column. Parameter settings are partially determined by the settings of 'create_ncc_model' therefore no rotation or down sampling is processed. The final output is one value for translation in row direction and one value for translation in column direction.

The cross-correlation algorithm (chapter 3.3.1) is provided by HDevelop as well as by Matlab. It is a normalized cross-correlation algorithm which actually is not necessary for binary input data. If a binary cross-correlation algorithm were implemented as well, as the HDevelop and Matlab developers did, performance would be expected to increase.
The turn around time (TAT) on Xeon E5410 2.3 GHz 8 core system is more than ten times faster than on Pentium IV 3 GHz system as table 4.1 shows. The major reason is that HDevelop optimized the algorithm for multi-core processors (parallelization).

Table 4.1: Turn around time for normalized cross-correlation (HDevelop)

Image Size [row x column]	Number of evaluations	Pentium IV 3 GHz (1 core)	Xeon E5410 2.3 GHz (8 cores)
1000 x 1000	20000	27 sec	2.3 sec
2000 x 2000	20000	279 sec	25 sec

4.4.2 Distance Map and Mutual Information

This matching step starts with the binary region images and converts them with the use of Matlab to distance maps (chapter 3.2.3). The implemented algorithm in Matlab is called 'bwdist'. It computes the Euclidean distance transform of the binary images. For each pixel in the binary image, the distance transform assigns a number that is the distance between that pixel and the nearest pixel that is not zero in the binary image. Distance maps convert binary images to 8 bit images thus giving us an option to check the performance of MI for registration based only on the extracted region information. Further the distance map is imported to Xdibias and the module 'hinten' is applied. It searches the registration parameters within two images in a suggested search space. The registration parameters here are assumed to be those for which the MI registration function reaches a maximum value. The graphical user interface (GUI) of 'hinten' is shown in figure 4.8 and the parameters are described. Main parameters to change are the upper and lower search bound of rotation and translation to limit the search space. The calculation time of 'hinten' is shown in table 4.2. This algorithm has a complexity of N^2, that means twice the number of evaluations lead to four times of the run time. In this case we did an exhaustive search; to enhance the run time we can use an optimizer which looks for the maximum (Suri and Reinartz, 2009). The TAT is only listed for a single core processor, because 'hinten' is not yet parallelized.

Table 4.2: Turn around time for mutual information with distance maps (hinten)

Image Size [row x column]	Number of evaluations	Pentium IV 3 GHz (1 core)	Xeon E5410 2.3 GHz (8 cores)
1000 x 1000	1600	105 min	-
2000 x 2000	1600	270 min	-

4.4.3 Skeleton and Mutual Information

This alternative method takes the region of the SAR image and treats it with morphological operators. The first step is to process a closing, filling up gaps and cracks. The closing is a dilation followed by an erosion. It conversely follows that an opening, is an erosion followed by a dilation. The idea is to eliminate the disruptions and generate region medial

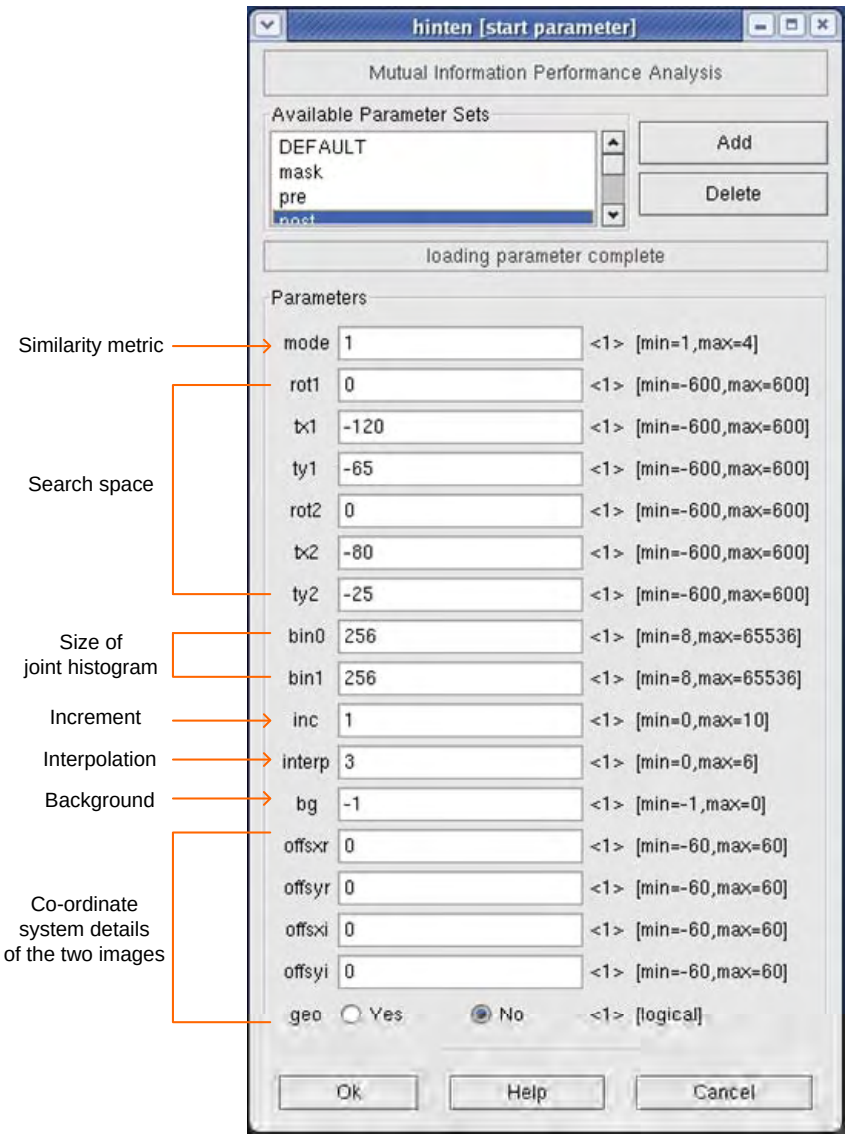


Figure 4.8: GUI of Xdibias module 'hinten'

axis through skeleton application (chapter 3.2.3). These operations are tools implemented in HDevelop, but due to their prominence many other image processing applications could be used in this case.
Now we take the skeleton (e.g. figure 5.12) and broaden it. On every side of the skeleton a fixed number of pixels are added; e.g. 5 pixels on every side. The size of the broadening depends on the desired regions. Further the new region works as a mask and we exactly take those pixels out of the original SAR image which are covered by the region.
Finally the masked SAR image and the original IKONOS-2 image are imported in Xdibias and 'hinten' processes the translation parameters. Adjustment of 'hinten' is the same as one section before, except the background value is zero instead of minus one. This reduces the used pixels and improves performance. The computation time for a single

core processor can be observed in table 4.3. The main benefit of this method is the fusion of the high performance and the use of the robust mutual information similarity metric.

Table 4.3: Turn around time for mutual information with skeleton area (hinten)

Image Size [row x column]	Number of evaluations	Pentium IV 3 GHz (1 core)	Xeon E5410 2.3 GHz (8 cores)
1000 x 1000	1600	5 min	-
2000 x 2000	1600	8 min	-

4.5 Verification

An automatic registration method should be able to detect wrong results by itself. For this purpose the circle shown in figure 4.9 provides a solution. The idea is to register

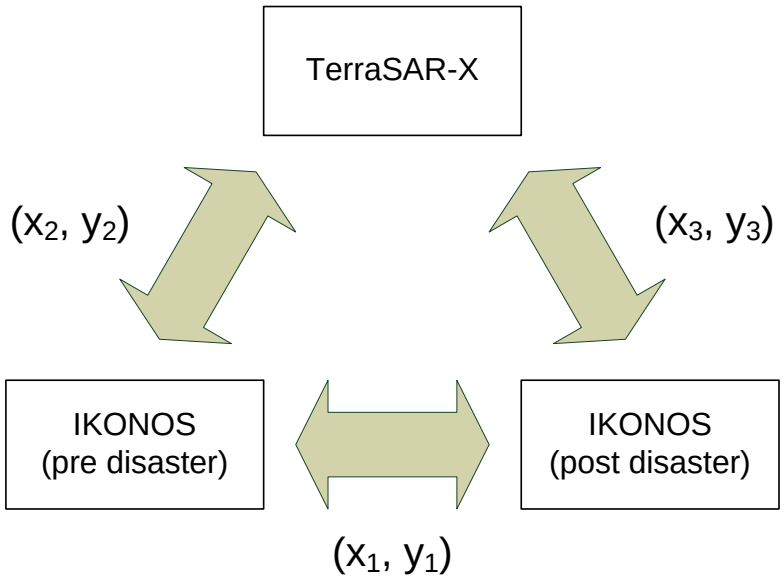


Figure 4.9: Circle of error

pre-disaster and post-disaster IKONOS-2 (x_1, y_1) and subtract the received translation parameters of the pre-disaster IKONOS-2 to TerraSAR-X registration (x_2, y_2) and add the result of post disaster IKONOS-2 to TerraSAR-X registration (x_3, y_3). The Error in X and Y direction can be simply expressed as:

$$Error(X) = x_1 - x_2 + x_3 \quad (4.1)$$

$$Error(Y) = y_1 - y_2 + y_3 \quad (4.2)$$

An error of zero only signifies a consistent performance match technique. It does not have any relation to absolute registration accuracy which cannot be done without having a reliable ground truth. The idea in this thesis would be to evaluate methods on their performance consistency.

Conclusion

Implementation of the proposed strategies has been explained in detail in this chapter. The next chapter presents the experimental results including a critical discussion.

5 Experimental Datasets, Results and Discussion

This chapter shows the results of certain feature matching techniques tested for high resolution SAR and optical images and further compares them to the explored registration strategies based on region growing. Finally all results will be discussed.
The utilized datasets have been acquired over the city of Sichuan in the south west of China. This area came in focus due to the disastrous earthquake on May 12, 2008. Our dataset comprises of three images, one IKONOS-2 image that has been taken before the disaster, one TerraSAR-X scene that has been taken quickly after the disaster and one IKONOS-2 scene that has been acquired more than a month after the earthquake. The meta data are listed in table 5.1. The results presented in this thesis are on small subscenes from the three complete scenes, here announced as dataset I and II. The initial translation error of the pre-disaster IKONOS-2 to the TerraSAR-X scene is about (-130, -27) meters and the error of the post-disaster IKONOS-2 to the TerraSAR-X scene is about (-89, 50) meters in x and y direction. Though these errors are a rough visual measurement and differ for each local area and as already mentioned earlier finding the exact translation parameters might be really difficult.

Table 5.1: Meta data of used datasets

	TerraSAR-X	IKONOS	
		(pre-disaster)	(post-disaster)
Size in pixels (row x column)	8119 x 6985	8160 x 6956	7284 x 7132
Pixel Spacing	1m (resampled)	1m	1m
Bit depth	16	11	11
Mode	HS	Forward Scanning	Forward Scanning
Processing Level	GEC	GEO	GEO
Map Projection	UTM	UTM	UTM
Sensor Incidence Angle	50.80°	26.49°	30.73°
Sun Incidence Angle	-	29.69°	26.75°
Sun Angle Azimuth	-	154.38°	112.93°
Cloud Cover [%]	-	5	2
Polarization	HH	-	-
Date of Acquisition	15/05/08	14/09/07	28/06/08

5.1 Results of Conventional Registration Strategies

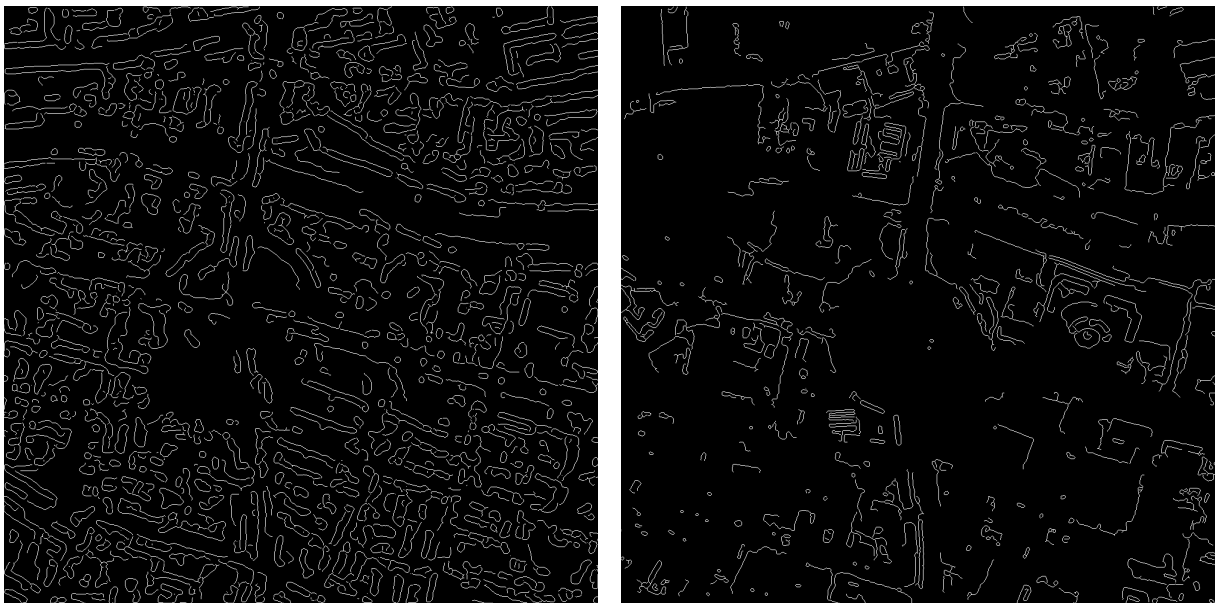
This section introduces the results of conventional registration strategies which are feature based. This methods are clustered into interest point based and edge point based methods. In the case of interest-point operators, chapter 3.2.1 had a description of the latest and most suitable one for our purpose, the SIFT operator. Schwind (2008) highlights the problems of of using the SIFT processing chain for high resolution SAR and optical images. It was shown that SIFT operator can not handle the very different radiometry and the completely different appearance of features.
The following lines of this section continue with results of two edge based methods namely Canny edge detector (chapter 3.2.2) and Steger operator (chapter 3.2.2).

5.1.1 Canny Edge Detector Results

The result of applying the Canny edge detector in a dense urban area can be seen here (same data as used for dataset II on page 50ff). The detected edges from images in figure 5.1 can be visualized in figure 5.2. Different operator parameter settings have been tried and the best parameter combination for SAR and optical image have been shown. Results in Figure 5.2 makes it clear that the extracted Canny edges in SAR image do not belong to the same features in the optical image.

(a) TerraSAR-X (b) IKONOS-2

Figure 5.1: Test dataset for Canny operator



(a) TerraSAR-X contours (b) IKONOS-2 contours

Figure 5.2: Edges detected by the Canny operator

5.1.2 Steger Operator Results

Figure 5.3 depict the performance of a Steger operator with not so dense urban features. In figure 5.3 the lines have been selected with similar methods as used in chapter 4.3 and finally the rest has been linked; as result a road has been extracted automatically. But the power of this operator can only be achieved in rural areas, for dense urban areas it has not been found to be useful.

(a) TerraSAR-X (b) IKONOS-2

Figure 5.3: Selected Steger line

The parameter setting of the Steger operator is more complex than e.g. the parameter setting of a simple Canny operator. It is necessary to have a prior knowledge of the input image and the desired features targeted for detection. Therefore the width and the intensity of desired lines play a role in the process of setting the parameters. Due to this reason only one road has been detected in figure 5.3 and the much broader road alongside could not be detected automatically.

5.2 Results of Explored Registration Strategies

This section presents the results of the explored registration strategies stepwise. The table 5.2 shows the numeric results of the three matching strategies with dataset I and the results of dataset II are tabulated in table 5.3. The last listed method named MI Urban has its roots from the work presented in Suri and Reinartz (2009) and Suri et al. (2009). It registers the original SAR and the optical image by finding a statistical dependency of the gray values. This method is also based on mutual information, but without region extraction. Details of the implemented method can be found in the cited references. The simple idea is to represent another method along with the region based ones to provide the possibility of comparing the results.

Table 5.2: Registration results of dataset I

Dataset I (Size 1000x1001)				
Method	Post IKO to TSX	Pre IKO to TSX	Pre IKO to Post IKO	Error
Binary Cross-correlation	(-84, -45)	(-120, -27)	(-38, 22)[1]	(2, 4)
Distance Map	(-80, -46)	(-123, -27)	(-38, 22)[1]	(5, 3)
Skeleton Area	(-95, -49)	(-126, -25)	(-38, 22)[1]	(7, 2)
MI Urban	(-94, -50)	(-131, -26)	(-38, 22)[1]	(1, 2)

[1] Measured manually by taking average of a number of observations

Table 5.3: Registration results of dataset II

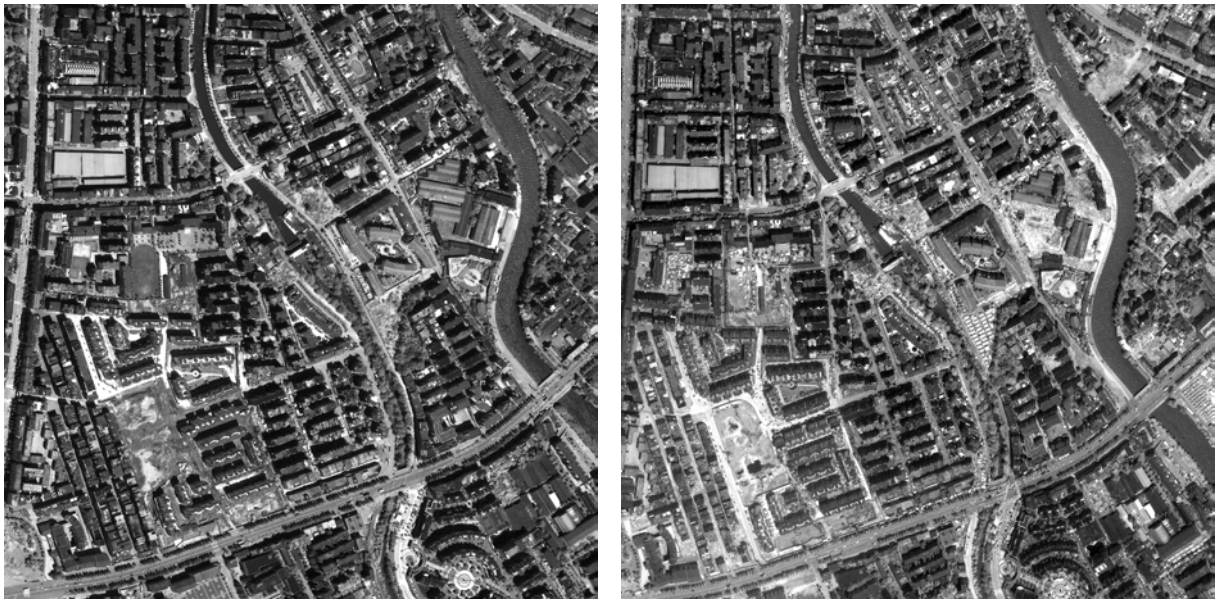
Dataset II (Size 999x999)				
Method	Post IKO to TSX	Pre IKO to TSX	Pre IKO to Post IKO	Error
Binary Cross-correlation	(-93, -48)	(-123, -30)	(-36, 18)[1]	(-6, 0)
Distance Map	(-99, -48)	(-120, -30)	(-36, 18)[1]	(-15, 0)
Skeleton Area	(-90, -45)	(-123, -24)	(-36, 18)[1]	(-3, 3)
MI Urban	(-96, -48)	(-132, -28)	(-36, 18)[1]	(0, 2)

[1] Measured manually by taking average of a number of observations

5.2.1 Dataset I

Figure 5.4 and 5.5 visualize the compressed and enhanced 1000 x1000 pixels images of dataset I.

Figure 5.4: Compressed and enhanced TerraSAR-X image (Dataset I)



(a) pre-disaster (b) post-disaster

Figure 5.5: Compressed and enhanced IKONOS-2 image (Dataset I)

The detected regions of dataset I are shown in figure 5.6 and 5.7. The regions are stored in a region array, so every region can be handled on its own. This is important in the next step, the region selection (explained in chapter 4.3).

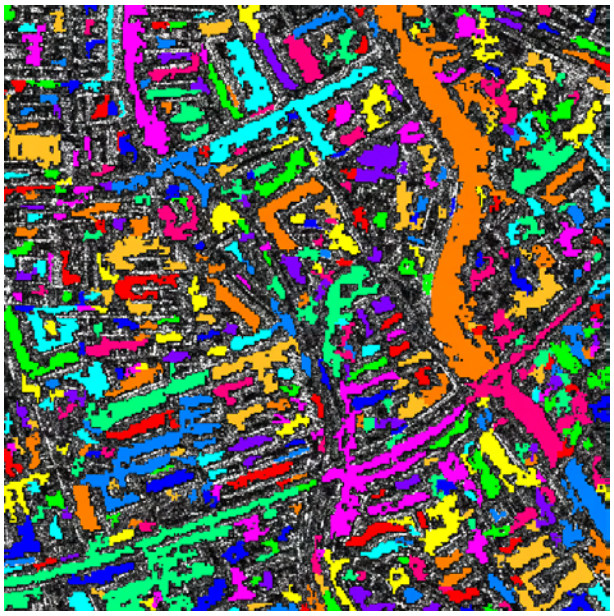


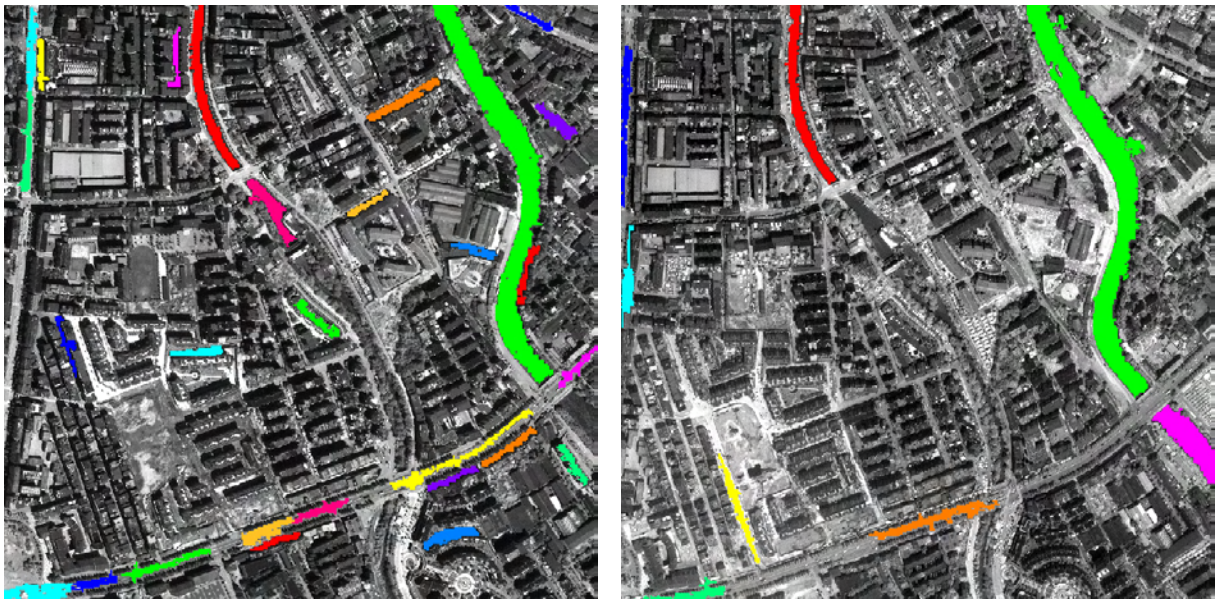
Figure 5.6: Regions TerraSAR-X image (Dataset I)

(a) pre-disaster (b) post-disaster

Figure 5.7: Regions IKONOS-2 image (Dataset I)

Here are the final regions (figure 5.8 and 5.9) after the last selection step. These images are now converted to binary images, if a pixel is associated to a region it has value one and pixels in the background are indicated by a zero. These binary images have been matched using binary cross-correlation which forms the first matching strategy in the presented work.

Figure 5.8: Selected regions TerraSAR-X image (Dataset I)



(a) pre-disaster (b) post-disaster

Figure 5.9: Selected regions IKONOS-2 image (Dataset I)

These figures (5.10 and 5.11) are the calculated distance maps from the previous binary region images (figure 5.8 and 5.9). The substance now is an 8 bit gray value image which has been matched using mutual information (matching strategy two).

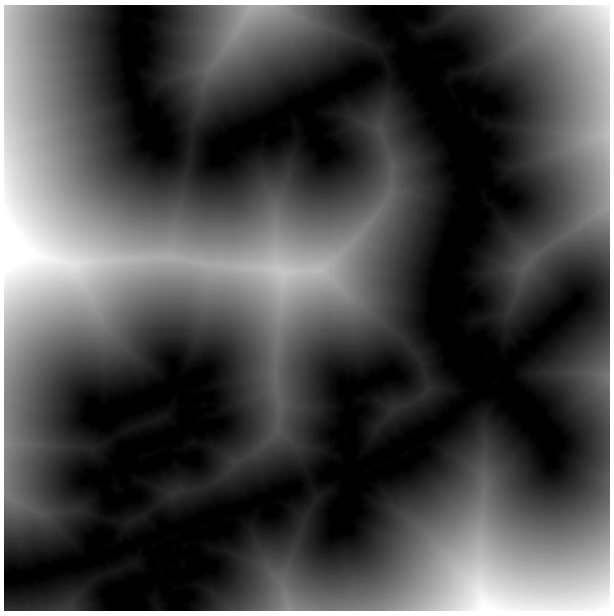
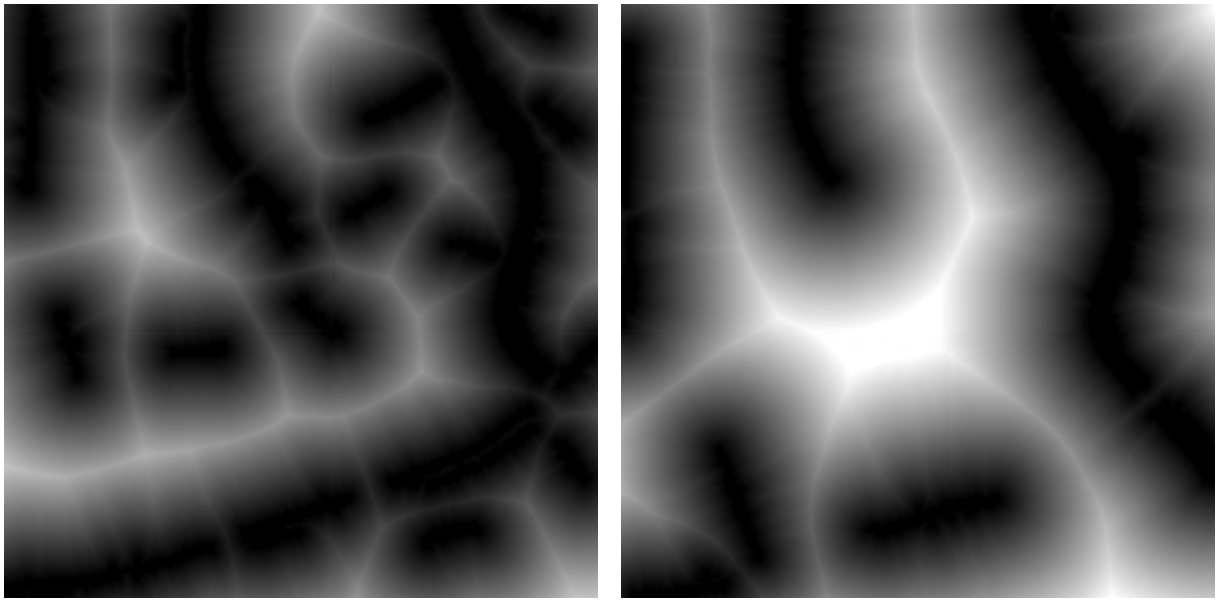


Figure 5.10: Distance map TerraSAR-X image (Dataset I)



(a) pre-disaster (b) post-disaster

Figure 5.11: Distance map IKONOS-2 image (Dataset I)

The sample here (figure 5.12) shows the extracted skeleton from final regions (figure 5.8). An important parameter here is selecting the number of pixels for broadening the existing skeleton (depends on the width of homogeneous regions in the original image). And now the original pixels of figure 5.4 are extracted, exactly those pixels are taken which are covered by the broadened skeleton in the case of an overlay. Finally the extracted SAR image information (skeleton) is matched with the original optical images using mutual information implemented in 'hinten' (figure 4.8).

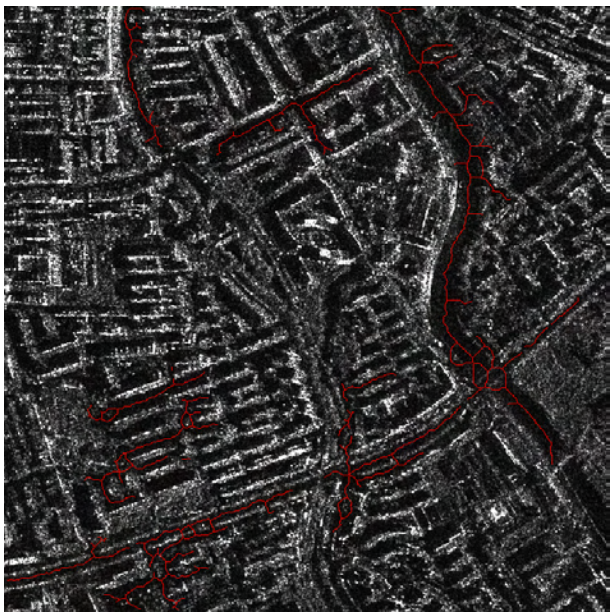


Figure 5.12: Skeleton TerraSAR-X image (Dataset I)

After the presentation of the data before matching, figures 5.13, 5.14 and 5.15 show the graphs of the similarity matrices these have produced when matching post disaster IKONOS-2 to TerraSAR-X image from dataset I. The point of interest is the similarity metric peak which should be clear without any ambiguity. In figure 5.13 the peak is more clear than in the two other ones, that is based on the search space which reaches for cross-correlation over the whole image (-1000 to 1000). This extensive search space is possible due to the very short TAT of this method (see chapter 4.4.1).

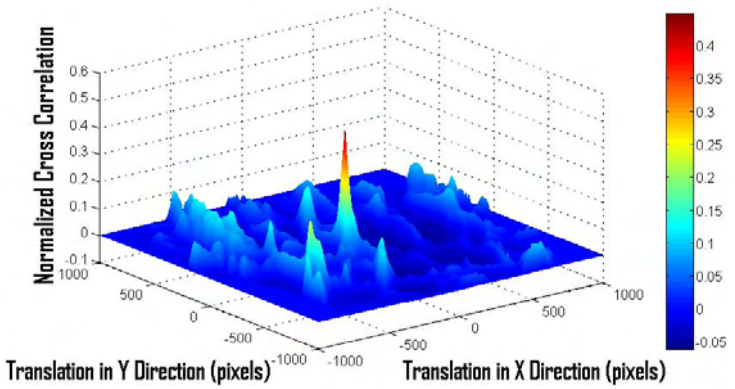


Figure 5.13: Normalized Cross-correlation (Dataset I)

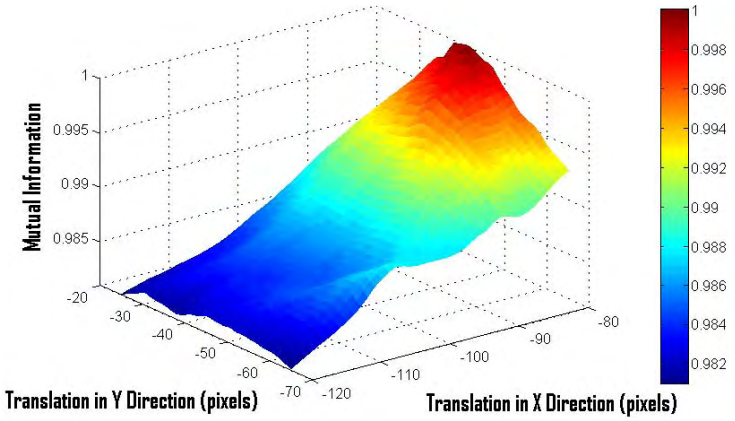


Figure 5.14: Mutual Information with distance map (Dataset I)

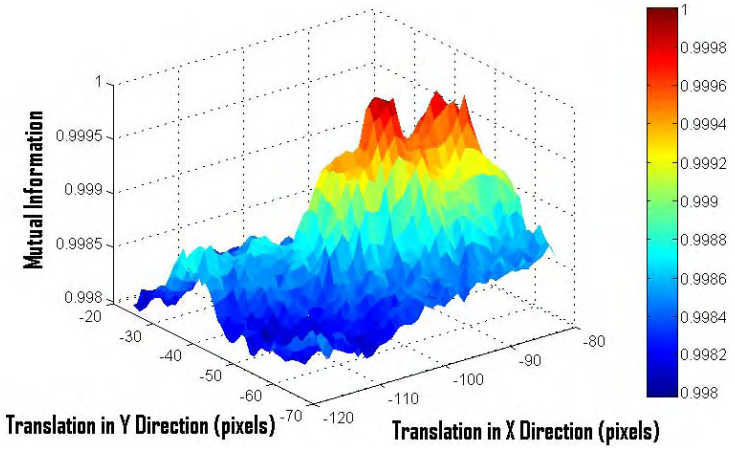


Figure 5.15: Mutual Information with skeleton (Dataset I)

Finally the registration is accomplished and figure 5.16 shows the data before registration. It is the SAR and the post disaster optical image overlayed with old geocode, whereas figure 5.17 shows the overlay with improved geocode after registration. The translation parameters are taken from cross-correlation matching (see table 5.2). Note that the river which runs from south to north in the right part of the image matches its counterpart. Also the street from west to east in the lower part of the image supports the success of a rough global registration at least.

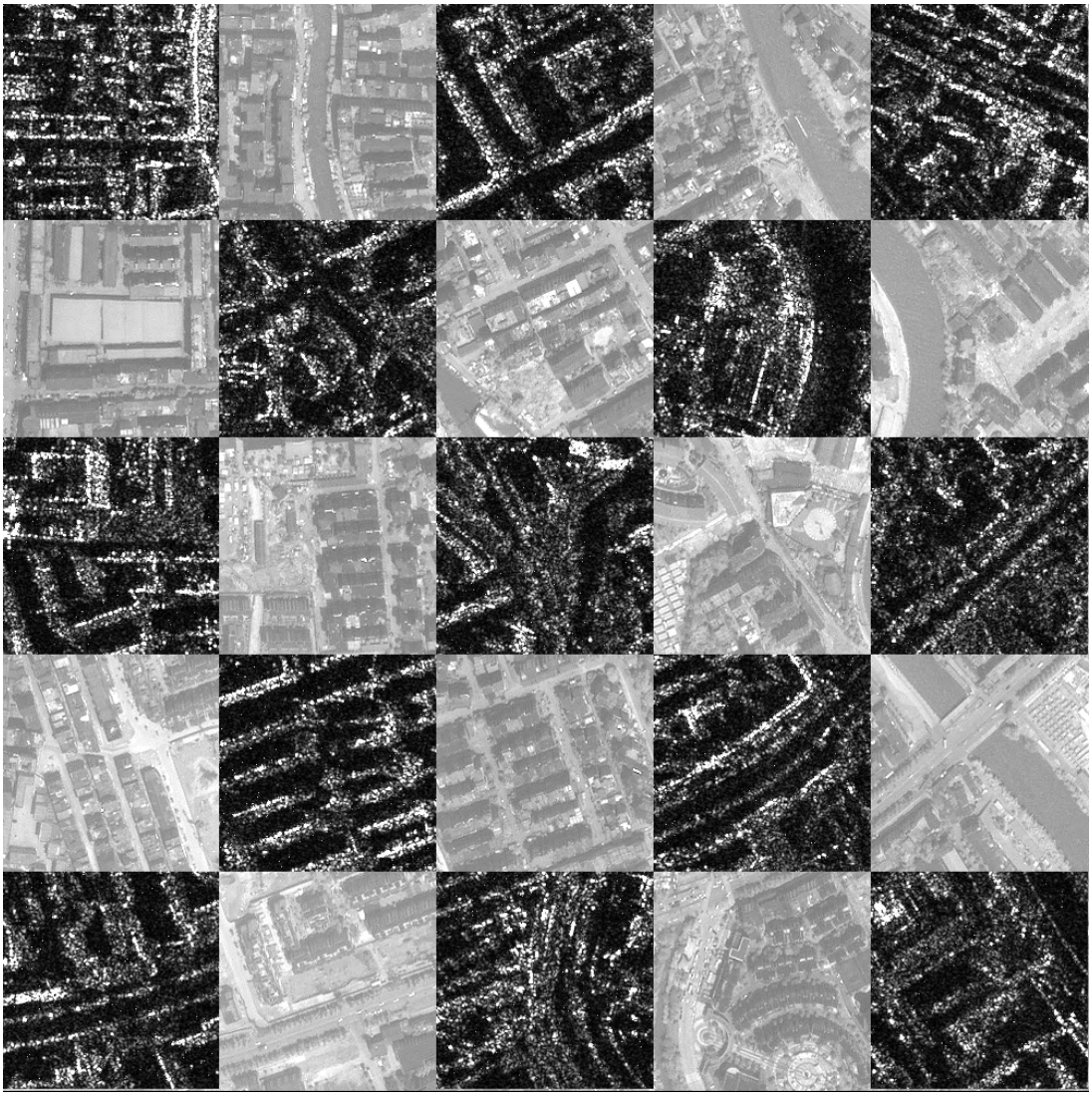


Figure 5.16: Check squared post disaster IKONOS-2 and TerraSAR-X before registration (Dataset I)

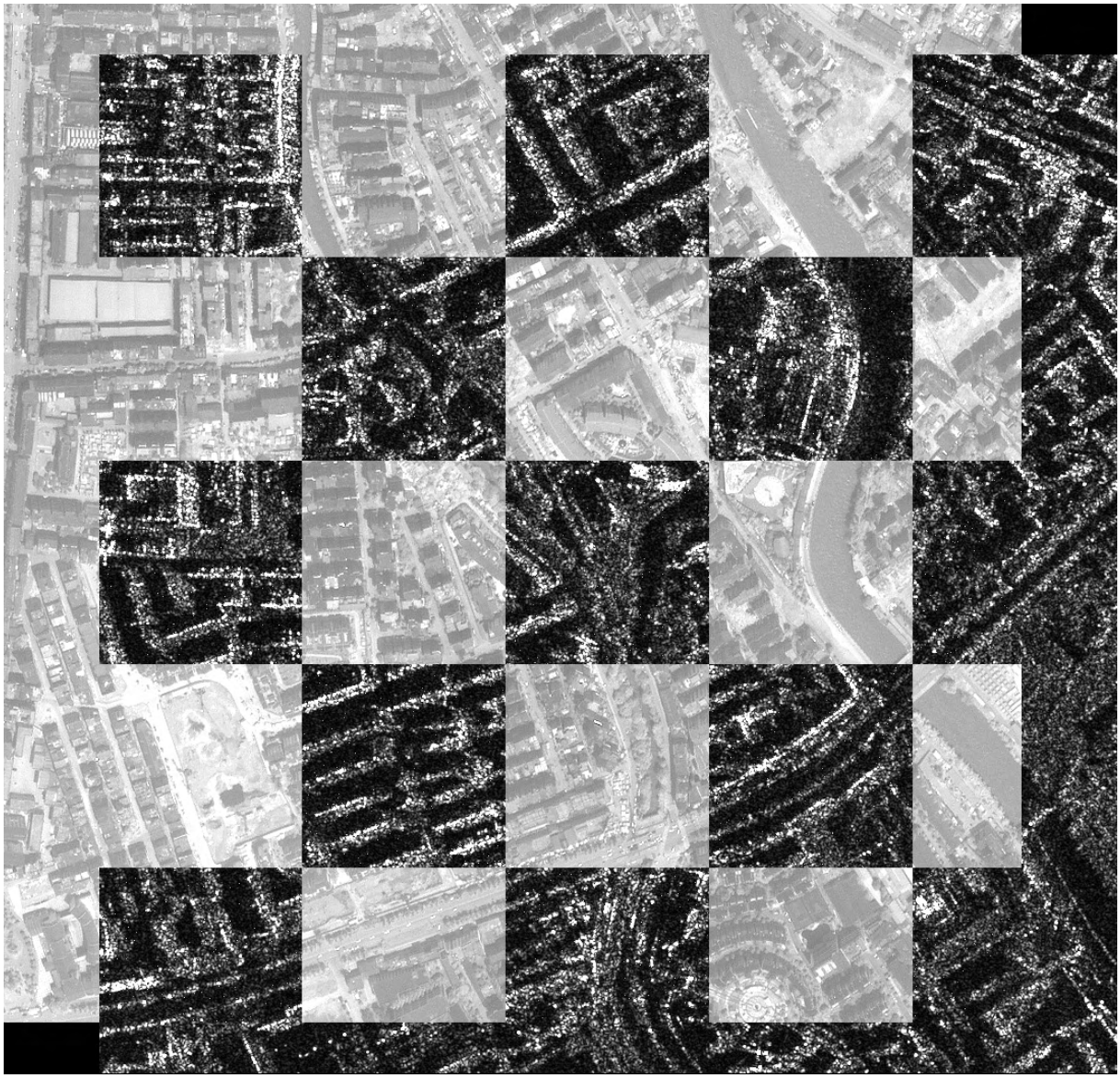
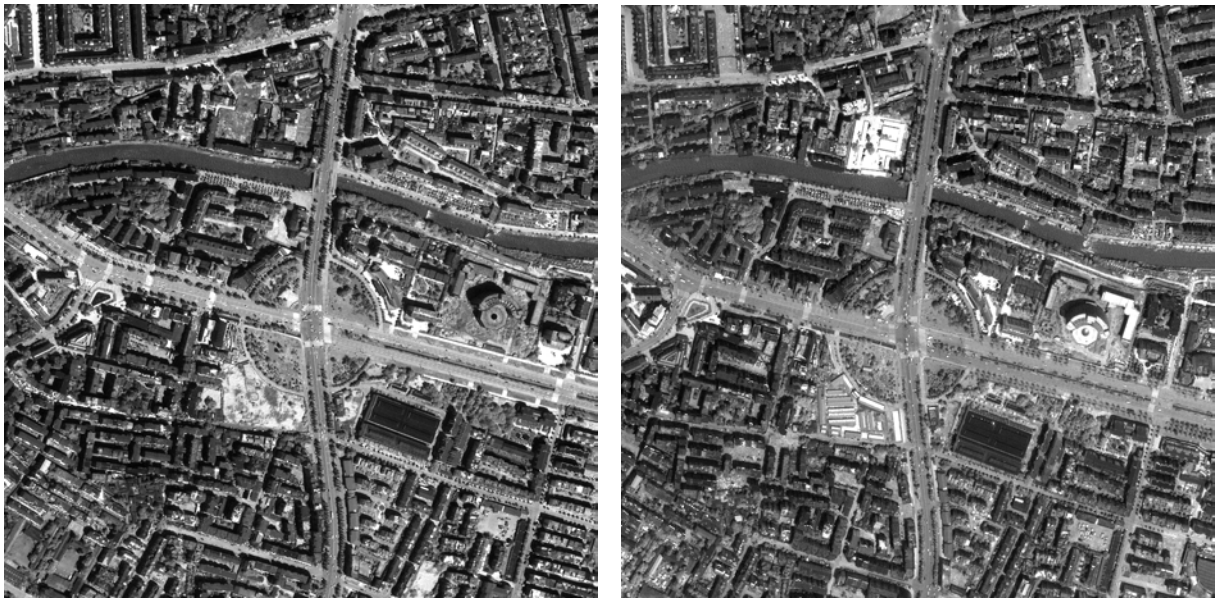


Figure 5.17: Check squared post disaster IKONOS-2 and TerraSAR-X image after registration (Dataset I). The translation parameters are taken from cross-correlation matching (-84, -45).

5.2.2 Dataset II

Figure 5.18 and 5.19 visualize the compressed and enhanced 1000 x1000 pixels images of dataset II.

Figure 5.18: Compressed and enhanced TerraSAR-X image (Dataset II)



(a) pre-disaster (b) post-disaster

Figure 5.19: Compressed and enhanced IKONOS-2 image (Dataset II)

The detected regions of datasetII are shown in figure 5.20 and 5.21.

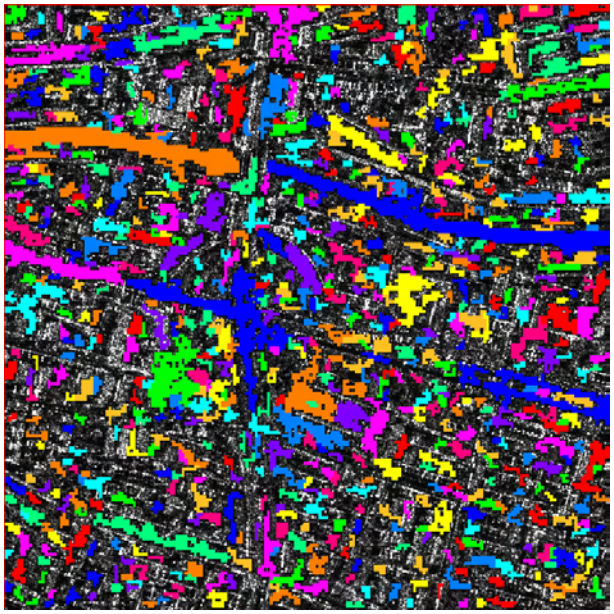
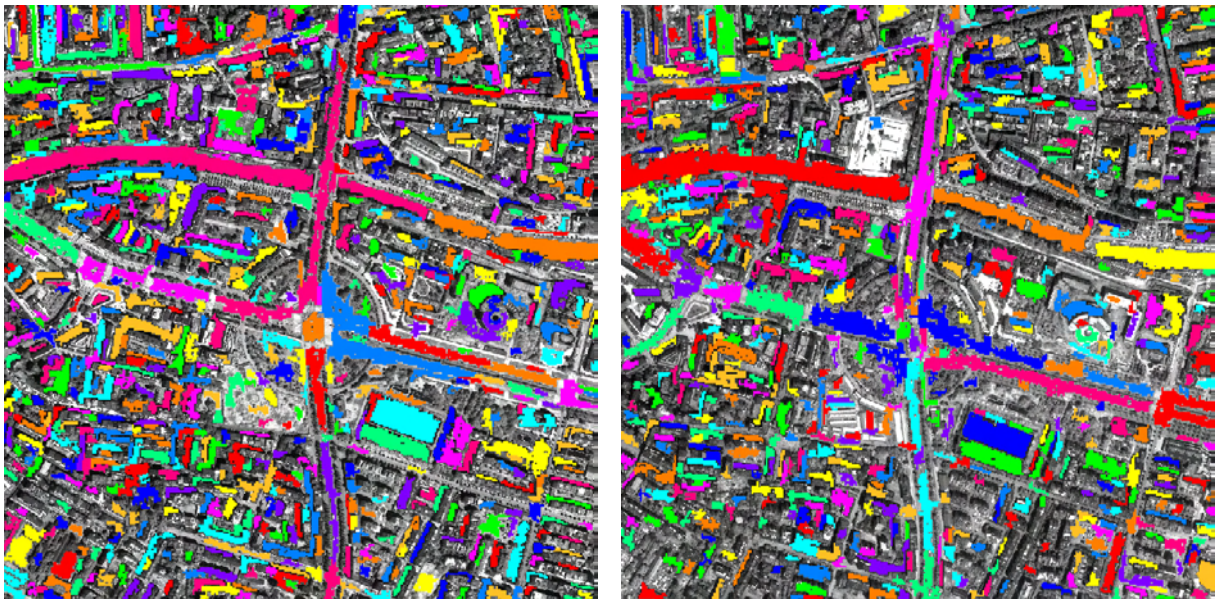


Figure 5.20: Regions TerraSAR-X image (Dataset II)



(a) pre-disaster (b) post-disaster

Figure 5.21: Regions IKONOS-2 image (Dataset II)

The final regions after the selection step are displayed in figure 5.22 and 5.23.

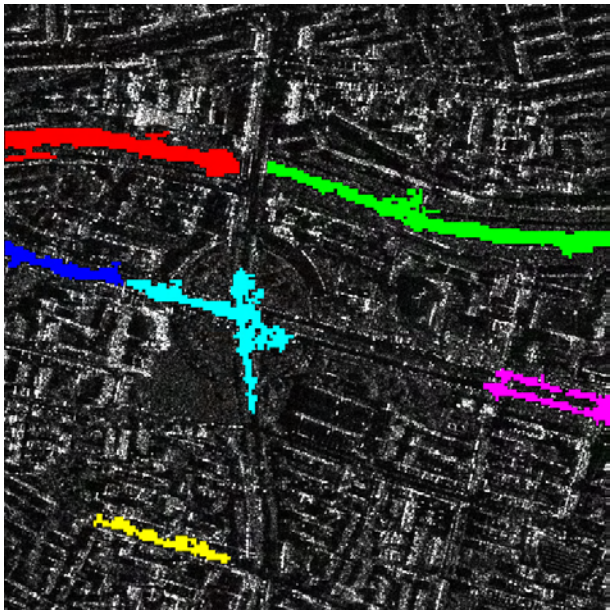
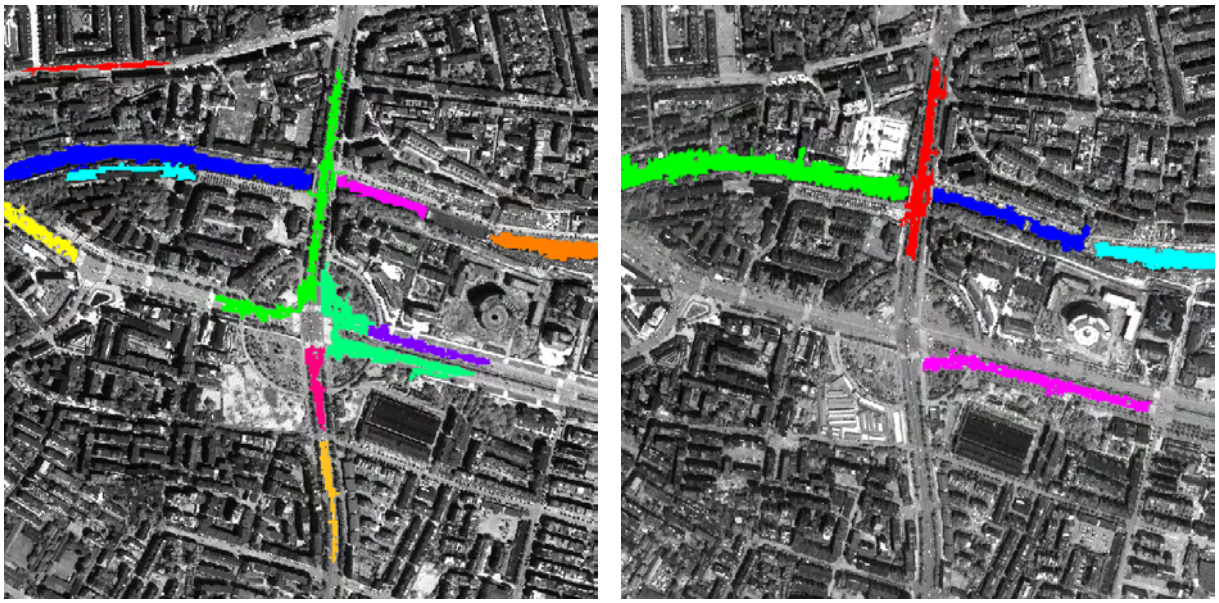


Figure 5.22: Selected regions TerraSAR-X image (Dataset II)



(a) pre-disaster (b) post-disaster

Figure 5.23: Selected regions IKONOS-2 image (Dataset II)

The three similarity matrices for the matching of post disaster IKONOS and TerraSAR-X are visualized in figure 5.24, 5.25 and 5.26. The focus is on the peak of the graphs which is desired without ambiguities.

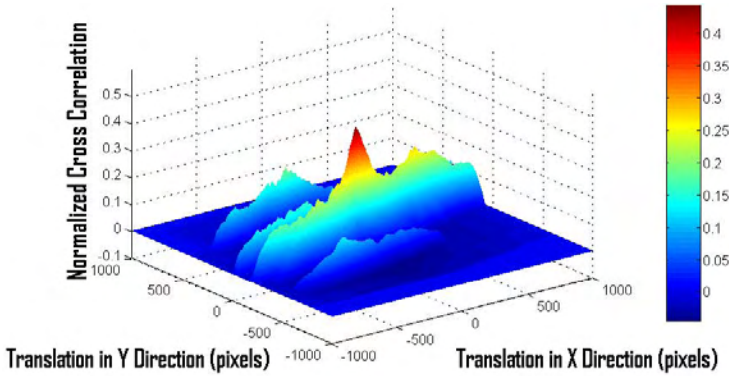


Figure 5.24: Normalized Cross-correlation (Dataset II)

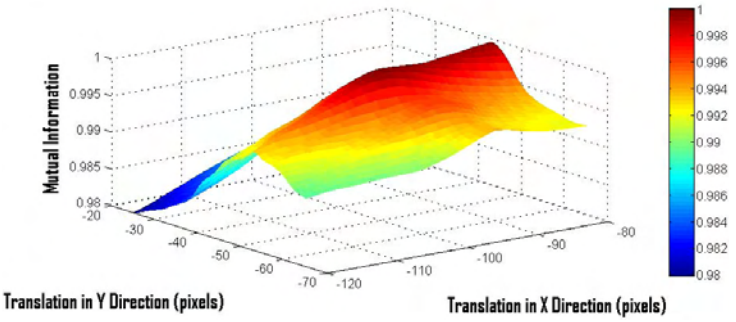


Figure 5.25: Mutual Information with distance map (Dataset II)

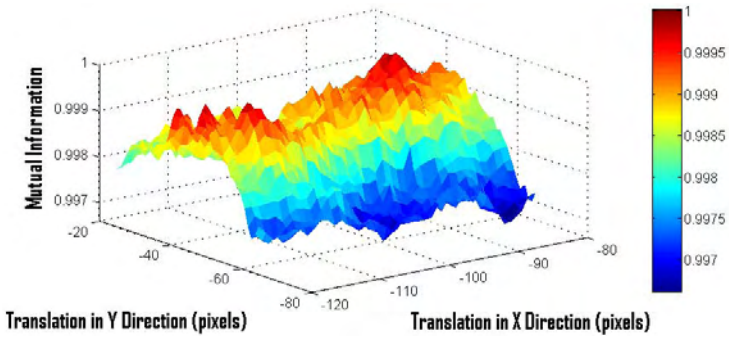
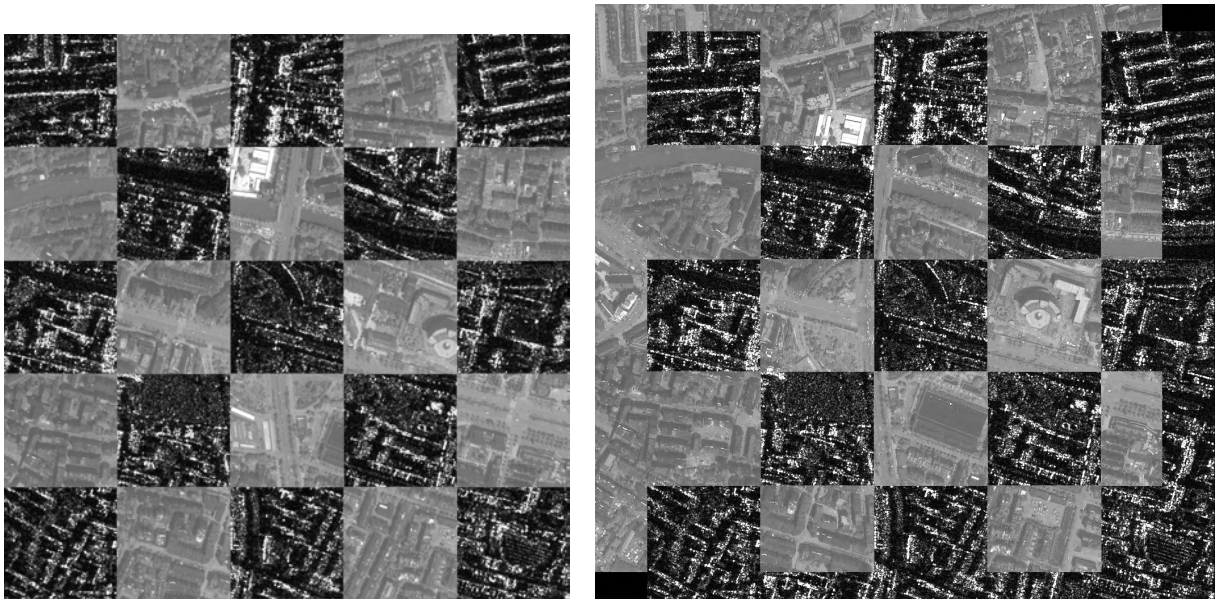


Figure 5.26: Mutual Information with skeleton (Dataset II)

Figure 5.27 shows the dataset II before registration (a) and after accomplished registration (b).



(a) Not registered (b) Parameters of translation (-93, -48)

Figure 5.27: Check squared original TerraSAR-X and geocode improved IKONOS-2 post-disaster image (Dataset II)

5.3 Discussion

The previous sections showed some attempts to register SAR and optical imagery, now the results will be discussed. Due to the very different nature of the images involved, interest-point based registration schemata are not expected to offer a robust solution.
Another possibility of registration provides the next group of operators, the edge based ones. But as section 5.1.1 (Canny Edge Detector Results) shows, the edges of SAR and optical imagery in urban areas are very incompatible. Generally, it would be quite hard to match these edges and it is also difficult to establish automatically that the extracted lines belong to the same feature, in both the images.
While observing the chapter 5.1.2 (Steger Operator Results) it becomes obvious that the Steger operator is more robust, but only in a rural area. The extracted road is not surrounded by any buildings. When the Steger operator is applied in urban areas with a high building density it produces similar results to the Canny detector.
The conclusion could be that edge based applications are not able to handle the requirements of high resolution images with dense building structures. The simple reason is the complete different appearance of objects in SAR and optical images. The contours in the SAR image mostly surround the shadows or double bounce reflections of the buildings

and there is no solution to avoid this behavior.
The next section 5.2 presenting the results of the explored registration strategies shows that there are other ways that have shown encouraging results. As the proposed strategy shows the region growing based method has much more potential concerning high resolution in urban areas than previous methods. Major reasons are the robustness of the region detection which is able to handle even very noisy SAR images and the creditableness of matching only parts of ground features. On the other hand, the limitation of this method is obviously the requirement of homogeneous on ground regions like roads or rivers. But if there are no other ground features that can be used, automatic registration with image processing algorithms is impossible without additional information.
The matching step after region detection showed a result that has to be differentiated. All three matching methods seem to work, but every method has its own drawback. In the case of cross-correlation it is the lack of robustness against wrong detected regions, but it is the fastest algorithm compared to other two techniques. Figure 5.13 and 5.24 show the peak of the similarity metric, which looks very solid, but this is due to good extracted regions, if there are more defect detections the cross-correlation matching based on maximum likelihood fails. Results from distance maps have not been very encouraging. As results in table 5.2 and 5.3 show it provides least consistency and figures 5.14 and 5.25 support the thesis due to a very fuzzy peak. A possible reason is that the distance maps put too much emphasis on wrong detected regions that are still included in the final region image. And the last strategy, the skeleton based strategy, showed the best compromise between runtime and robustness which is evident by observing tables 5.2 and 5.3. However the similarity metrics in figure 5.15 and 5.26 do not show a desirable peak. This strategy needs further investigations on similarity metric behavior with respect to the amount of information collected from the SAR image and this has not been accomplished in this work. Finally the results of region based strategy can be compared to the MI Urban operator which is also listed in tables 5.2 and 5.3. The consistency shown in the error column has the best values for the MI Urban, but the difference to the other methods is only a few pixels. And that is the point, in reality it is not easy to register radar and optical imagery with a accuracy of one pixel. Also a manual processed registration by determining corresponding points visual can not claim the accuracy of 1 pixel. Further all these algorithms have only made attempts to achieve an automated rough global registration, fine local variations are still very much expected to be present in the aligned images depicted in figure 5.17 and 5.27.

6 Conclusions and Scope of Future Work

The report began by highlighting the importance and motivation of the presented work. Thereafter general characteristics about optical and SAR imaging along with the details about the used data products from IKONOS-2 and TerraSAR-X were explained. Further the necessary theoretical background drew attention to present common registration methods. The next chapter showed the developed registration strategy followed by the results discussed and presented with the help of text, tables and figures. Finally this chapter gives a summary of the entire project along with that draw conclusion and highlights the scope of future work.

6.1 Conclusion

This thesis points out the problems which surface in the SAR - optical registration process and suggested solutions for the same. We show that many problems combine and make the problem of registration quite complex. Major of these problems are the different geometry and the very different radiometry which is inherent because of the very different sensor nature.
Due to high resolution and the urban area of interest, classical approaches like interest point based or edge based methods do not offer viable solutions. Hence the major idea was to develop a registration strategy based on region growing. The advantage of region growing is the low sensitiveness in the case of noisy SAR data. As a preprocessing step, high frequencies in the images have been reduced to simplify the region extraction. The extracted regions have to run through a selection chain and finally matched using two very different similarity metrics namely normalized cross-correlation and mutual information. The first strategy using cross-correlation is very fast, but has a problem with wrong detected regions and this results in a low robustness. Next strategy using distance maps takes a long computation time and has also not been found to be very robust. Alternatively the skeleton based variant offered some positive results with its short run time and acceptable robustness, but still this technique needs rigorous testing on various datasets. The testing with utilized datasets showed that the presented method has potential for an automatic rough alignment of SAR and optical imagery, but needs on all accounts a testing on other scenes as well. The test, included in this thesis, only proofs the consistency, as due to the lack of ground control points a verification about the absolute

geometric accuracy is not feasible and as already mentioned visual verification becomes very subjective to individual interpretation.

6.2 Scope of Future Work

Registration of high resolution SAR and optical images in urban areas is an ambitious aim. To look ahead the explored region based strategy offers the possibility to handle imagery with much higher resolution as used in this thesis. There is no doubt, that high resolution satellites of the future would have a much higher geometric resolution than 1 m, this poses new problems, but also offers new prospects.
Further advantage could be the use of more information, e. g. several polarizations of the SAR signal these act like bands in the electromagnetic spectrum and would simplify the classification of on-ground features and off-ground features.
The aim of this thesis was the exploration of a strategy that is able to handle the difficulties of registering SAR and optical satellite data. Here, a global registration has been achieved fully automatically. Further work can be done to achieve a fine local registration using MI as the similarity metric but it might also require a DSM and certain shadow detection techniques for fruitful results.

Bibliography

Ayoub, F., Leprince, S., Binet, R., Lewis, K., and Aharonson, O. (2008). Influence of camera distortions on satellite image registration and change detection applications. *IGARSS*. 2, 6

Breu, H., Gil, J., Kirkpatrick, D., and Werman, M. (1995). Linear Time Euclidean Distance Transform Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):529–533. 23

Brown, L. (1992). A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4). 15

Canny (1986). A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 679–698. 19

Castleman, K. (1996). *Digital Image Processing*. Prentice Hall. 22

Chanda, B. and Majumder, D. (2006). *Digital Image Processing and Analysis*. Prentice Hall of India. 17

Chen, H., Varshney, P., and Arora, M. (2003). Mutual information based image registration for remote sensing data. *International Journal of Remote Sensing*, 24(18):3701–3706. 24

Eckardt, U. (1988). Verdünnung mit perfekten Punkten. In *10.DAGM Symposium, IFB 180*, Zürich. 22

Eineder, M. and Fritz, T. (2008). TerraSAR-X Ground Segment, SAR Basic Product Specification Document (TX-GS-DD-3302). Technical Report Release 1.5, DLR. vii, ix, 10, 11, 13

Elachi, C. (1988). *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE Press. 6

European Space Imaging (2009). IKONOS price list for Europe and North Africa. Website. http://www.euspaceimaging.com/sime.asp?page=about, access 20.2.2009. ix, 10

Fabri, R., Costa, L., Torelli, J., and Bruno, O. (2008). 2D Euclidean distance transform algorithms: A comparative survey. *ACM Computing Surveys*, 40(1):1–44. 23

Foresti, G., Regazzoni, C., and Varshney, P. (2003). *Multisensor Surveillance Systems: The Fusion Perspective*. Springer Press. 5

Frost, V., Stiles, J., Josephine, A., and Shanmugan, K. (1982). A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4(2). 16

Förstner, W. and Gülch, E. (1987). A fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In *Proceedings Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken. 18

Goshtasby, A. (2005). *2-D and 3-D Image Registration for Medical, Remote Sensing and Industrial Applications*. Wiley Press. 18

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151. 18

Hayes, M. (1996). *Statistical Digital Signal Processing and Modeling*. Wiley Press. 16

Inglada, J. (2002). Similarity measures for multi sensor remote sensing images. In *Proceedings of IGARSS*, Toronto. 2, 23, 25

Kramer, H. (2002). *Observation of the Earth and Its Environment: Survey of Missions and Sensors*. Springer Verlag, 4th edition. ix, 9

Lee, J. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(2):165–168. 16

Lehureau, G., Tupin, F., Tison, C., Oller, G., and Petit, D. (2008). Registration of metric resolution SAR and Optical images in urban areas. In *7th European Conference on Synthetic Aperture Radar*, Friedrichshafen. 2

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–154. 21

Lopes, A., Touzi, R., and Nezry, E. (1990). Adaptive Speckle Filters and Scene Heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, 28(6):992–1000. 17

Lowe, D. (1999). Object recognition from local scale invariant features. In *Proceedings of the International Conference on Computer Vision ICCV*, pages 1150–1157, Corfu. 18

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110. 18

Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*. 15

Meinel, G. and Reder, J. (2001). IKONOS Satellitenbilddaten - Ein erster Erfahrungsbericht. *Kartographische Nachrichten, Kirschbaumverlag*, 1:40–46. 1, 9

Modersitzki, J. (2004). Numerical methods for image registration. *Oxford University Press*. 15

Nonaka, T., Ishizuka, Y., Yamane, N., Shibayama, T., Takagishi, S., and Sasagawa, T. (2008). Evaluation of the Geometric Accuracy of TerraSAR-X. Beijing. Proceedings of 21th ISPRS Congress. 1, 11

Roth, A., Hoffmann, J., and Esch, T. (2008). TerraSAR-X: How can High Resolution SAR Data Support the Observation of Urban Areas? In *ISPRS Congress*, pages 840–844, Istanbul. 12

Schmidt, N., Koppe, W., and Hennig, S. (2008). *Optimisation of the TS-X ORI Processor and test of the prototype*. DeSecure, d420.2 edition. ix, 11, 13

Schwind, P. (2008). Critical Evaluation of the SIFT Operator for Remote Sensing Imagery. Master's thesis, FH Landshut. 2, 6, 18, 19, 38

Schwind, P., Suri, S., Reinartz, P., and Siebert, A. (2009). Applicability of the SIFT operator to geometric SAR image registration. *International Journal of Remote Sensing*. (inpress). 19

Shan, J. and Stilla, U. (2008). Foreword remote sensing data fusions. *PE&RS*, 74(2). 15

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656. 24

Sobel, I. and Feldman, G. (1968). A 3x3 isotropic gradient operator for image processing. presented at a talk at the Stanford Artificial Project (unpublished). 20

Steger, C. (1996). Extracting curvilinear structures: A differential geometric approach. *B. Buxton, R. Cipolla, eds., Fourth European Conference on Computer Vision, Lecture Notes in Computer Science, Springer Verlag*, 1064:630–641. 21

Stilla, U. (2007). High resolution radar imaging of urban areas. In *Fritsch D. (ed) Photogrammetric Week, 07*, pages 149–158, Wichmann: Heidelberg. (invited). 7

Stilla, U., Soergel, U., and Thoennessen, U. (2003). Potentials and limits of InSAR data for building reconstruction for built-up areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58:113–123. vii, 8

Studholme, C., Hill, D., and Hawkes, D. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86. 25

Suri, S. and Reinartz, P. (2008). Application of Generalized Partial Volume Estimation for Mutual Information based Registration of High Resolution SAR and Optical Imagery. Hannover. 11th International Conference on Information Fusion. 25

Suri, S. and Reinartz, P. (2009). On the Possibility of Intensity Based Registration for Metric Resolution SAR and Optical Imagery. In *AGILE Conference*, Hannover. 2, 33, 41

Suri, S., Türmer, S., Reinartz, P., and Stilla, U. (2009). Registration of High Resolution SAR and Optical Satellite Imagery in Urban Areas. In *ISPRS Congress*, Hannover. 41

TerraSAR-X (2009a). Mission statement. Website. http://wwwserv2.go.t-systems-sfr.com/tsx/start_en.htm, access 17.2.2009. ix, 12

TerraSAR-X (2009b). Price list commercial. Website. http://www.dlr.de/Portaldata/1/Resources/raumfahrt/weltraum/pricelist_commercial.pdf, access 18.3.09. ix, 12

Wegner, J. (2007). Automatic Fusion of SAR and Optical Imagery. Master's thesis, Leibnitz University Hannover. 6, 7

Weidner, U. (1997). *Gebäudeerfassung aus digitalen Oberflächenmodellen*. Deutsche Geodätische Kommission, München, Band 474 der Reihe C edition. 6

Wessel, B. (2006). *Automatische Extraktion von Straßen aus SAR-Bilddaten*. PhD thesis, Technical University Munich. 18, 21

Wessel, B., Huber, M., and Roth, A. (2007). Registration of Near Real-Time SAR Images by Image-to-Image Matching. *Stilla U et al. (Eds) PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36. 18

Wong, A. and Clausi, D. (2006). ARRSI: Automatic registration of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1483–1493. 15

Zitová, B. and Flusser, J. (2003). Image registration methods: A survey. *Image Vision and Computing*. 15

Declaration of Authorship

I certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other University.

Sebastian Türmer
Munich, 30th April 2009