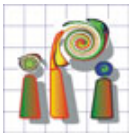
Institut für Photogrammetrie und GeoInformation
Leibniz Universität Hannover

Diplomarbeit

Automatic Fusion of SAR and Optical Imagery

Jan Dirk Wegner



September 2007

Tutor at the Institut für Photogrammetrie und Geoinformatik (IPI), Leibniz Universität Hannover:

Uwe Sörgel

Tutors at the Centre National d'Études Spatiales (CNES), Toulouse:

Jordi Inglada and Céline Tison

Tutor at Communcation & Systèmes (CS), Toulouse:

Thomas Feuvrier

Presentation and oral examination: Hannover, 10th September 2007

Statement

I, Jan Dirk Wegner, state that this Diplomarbeit has been written entirely by myself. No further sources besides the ones noted in the bibliography have been used.

Hiermit erkläre ich, Jan Dirk Wegner, dass die vorliegende Diplomarbeit selbstständig von mir verfasst wurde. Ich habe keine anderen als die im Literaturverzeichnis angegebenen Quellen und Hilfsmittel verwendet.

Hannover, 6th September 2007

Acknowledgements

This project has been accomplished on the basis of a French-German cooperation. The three partners are the Centre National d'Études Spatiales (CNES) and Communication & Systèmes (CS) on the French side and the Institut für Photogrammetrie und GeoInformation (IPI) at the Leibniz University of Hannover on the German side. It is the very first time that these three partners have worked together in research. While setting up this project, multiple bureaucratic cliffs had to be circumnavigated. Therefore, I wish to express my sincere thanks to my tutors Uwe Sörgel, Jordi Inglada, Céline Tison, Thomas Feuvrier and to the head of the CNES service "Analyse & Produits Image", Frédéric Adragna. It is due to their patience and commitment that I could spend inspiring ten months at the CNES and at CS. Additionally, their professional experience in remote sensing has deepened my understanding for the subject. They have sensitized me for current and future issues to think about.

My professional and personal life have greatly benefitted from the kindness and openness of my CNES colleagues Tarek Habib, Gwendoline Blanchet, Emmanuel Christophe, Stephane May, Vincent Poulain, Miarintsoa Ramanantsimiavona and Bernard Rougé. My motivation throughout the entire project is, not at last, due to the laid-back but none the less ambitious atmosphere they created.

I hope that future research will further benefit from this fruitful cooperation now established.

Contents

Abstract

The goal of this diploma thesis is the development and implementation of a fusion method for high resolution optical and SAR imagery. The developed approach is generic and hence it is not limited to cases when exact sensor parameters are known. The idea is to first reduce geometric distortions with rather generic transforms, applicable to different kinds of optical and SAR sensors, respectively. For example, the optical images are ortho-rectified using the collinearity equations. In the following, the ortho-rectified images are prepared for line extraction. Smoothing, anti-speckling and classification are conducted. The classification serves for partitioning the images into rectified ground and unrectified buildings. Next, lines are extracted because classical pixel based registration methods often fail for our very high resolution images. Furthermore, distance images are derived from the extracted lines. They are compared iteratively and resampled one onto the other. This iterative registration procedure is modularly constructed and thus allows for adaption to various input imagery. All algorithms are implemented in the open source software library ORFEO Toolbox (OTB). The diploma thesis was accomplished at the Centre National d'Études Spatiales (CNES) in Toulouse, section DCT/SI/AP, in cooperation with Communication & Systèmes (CS).

Résumé

L'objectif de ce rapport de stage est le developpement et l'implémentation d'un modèle de fusion pour les imageries optiques à haute résolution et les images SAR. La méthode développée est générique et par conséquence, elle n'est pas limitée aux cas où les paramètres des capteurs sont connus. L'idée est de commencer par diminuer les distorisions géométriques avec une transformation générique, applicable pour les différents types d'imageries. Par exemple, les images optiques sont ortho-rectifiées en utilisant les équations de colinéarité. Ensuite, les images orthorectifiées sont préparées pour l'extraction de lignes en utilisant plusieurs étapes de filtrage. Ces filtrages effectuent le lissage, débruitage et la classification. La classification sert à classifier les images en segments sols rectifiés et des bâtiments non-rectifiés. Ensuite, les lignes sont extraites car les méthodes basées sur uniquement les pixels ne sont pas assez performantes. Suite à l'extraction des lignes, les images distances sont obtenues. Ces images sont comparées itérativement et re-échantillonées l'une sur l'autre. Cette procédure d'enregistrement est modulaire et par conséquent, elle est adaptée à différents types d'imagerie. Tout les algorithmes sont implémentés en utlisant la librairie à code source libre ORFEO ToolBox (OTB). Ce travail a été mené au Centre National d'Études Spatiales (CNES) à Toulouse, service DCT/SI/AP, en coopération avec Communication & Systèmes (CS).

Kurzfassung

Das Ziel dieser Diplomarbeit ist die Entwicklung und Implementierung einer Fusionsmethode für optische und SAR Bilder hoher Auflösung. Die Methode ist generisch, da Sie nicht auf spezielle Sensorparameter angewiesen ist. Die Idee ist, zunächst mit generellen geometrischen Transformationen, die sich auf alle optischen und SAR Sensoren anwenden lassen, möglichst viele geometrische Verzerrungen in den Bildern zu beseitigen. Für den optischen Fall werden z.B. die Kollinearitätsgleichungen genutzt. Die so erzielten ortho-rektifizierten Bilder werden daraufhin separat weiteren Analyseschritten unterzogen, um die Unterschiede der Bilder in Geometrie und Radiometrie zu verringern. Nach der Vorverarbeitung der Bilder folgt eine Klassifikation, um für die spätere Registrierung zwischen rektifizierten Bodenflächen und nicht rektifizierten Gebäuden zu unterscheiden. Als nächster Bearbeitungsschritt wird eine Linienextraktion durchgeführt, gefolgt von der Berechnung von Distanzbildern. Die Distanzbilder der als rektifiziert klassifizierten Linien werden in einem iterativen Verfahren verglichen und aufeinander abgebildet. Dieses iterative Verfahren ist modular aufgebaut und erlaubt so die Anpassung auf die jeweiligen Eingabebilder. Alle entwickelten Algorithmen sind in der ORFEO Toolbox (OTB) implementiert. Die Diplomarbeit wurde durchgeführt am Centre National d'Études Spatiales (CNES) in Toulouse, Abteilung DCT/SI/AP, in Kooperation mit Communication & Systèmes (CS).

1 Introduction

1.1 Motivation

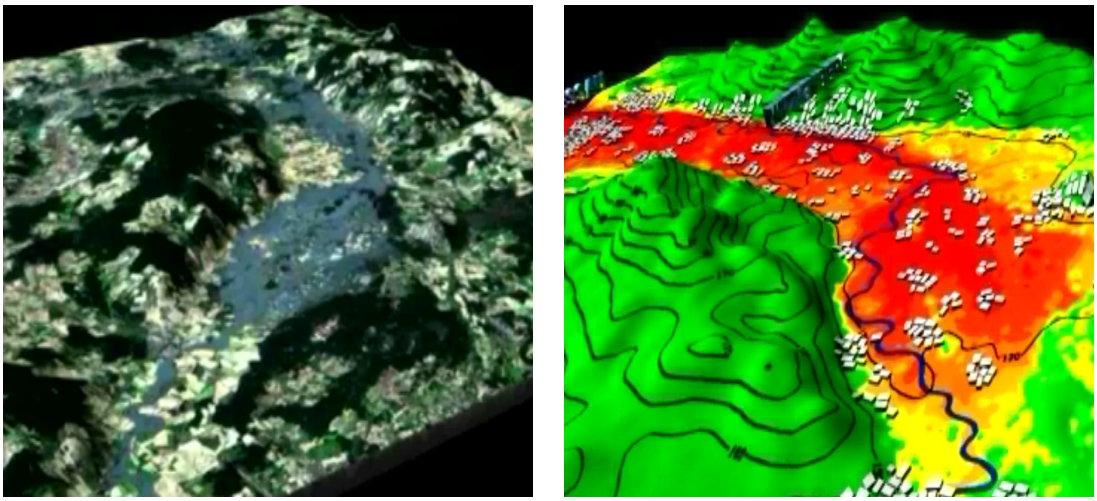
Within recent years, a variety of new high resolution airborne (Ramses, Pelican, PAMIR, DoSAR) and spaceborne (IKONOS2, ALOS, TerraSAR-X, CosmoSkyMed) imaging remote sensing sensors have been put into place. These sensors are either active (Synthetic Aperture Radar) or passive (optical). Further systems will be launched in near future (Radarsat2, Pléiades) to even extend today's potential of submetric imagery. Hence, new possibilities for the combined analysis of multi-sensor imagery of very high resolution arise. The combination of both radar and optical imagery enables the use of complementary information of the same scene and various application scenarios may be thought of.

An example will illustrate the need for combined optical and SAR imagery. In August 2002 a natural hazard of enormous extent hit parts of eastern and central Europe: the Elbe flooding (Fig. 1.1(a)). It caused 100 billion Euros of damage but fortunately almost no lifes were lost due to rapid emergency response. Satellite imagery supported the decision making process. The International Charter "Space and Major Disasters" [CNES, 2007a] was activated, facilitating the immediate exploitation of satellite images from international SAR and optical sensors. Radar images from ERS2 and Envisat were used to generate a virtual 3D landscape model (Fig. 1.1(b)). This model was overlaid with optical images in order to distinguish between farmland, forests, cities etc. (Fig. 1.2(a)). Furthermore, a risk map was derived, enabling emergency prediction. Areas which were in danger to be flooded are displayed in red while safe areas are shown in green (Fig. 1.2(b)). With the support of this risk map, much harm could be prevented. Roads were closed in time, windows and doors of houses within the risk zone were bricked up, entire areas could be evacuated before they were flooded. Since all data was continuously updated, changes within the flooded regions and the floods extent could be detected rapidly.

Hence, one major field calling for enhanced remotely gathered information is rapid change detection. It is conducted after the occurrence of natural or man-made hazards in order to facilitate emergency response. Satellites have the advantage of being operational globally while airborne sensors provide higher resolution data. Optical sensors provide relatively

(a) (b)

Figure 1.1: (a) Extent of the Elbe flood in eastern Germany in August 2002 (© ESA 2003), (b) ERS2 images overlaid with infrared images (© ESA 2003)



(a) (b)

Figure 1.2: (a) DTM generated from radar data overlaid with optical image (© ESA 2003), (b) Risk map derived from a DTM (© ESA 2003)

easy to interprete images since their viewing geometry and frequencies are closely related to the human perspective. With multispectral systems, color data is gathered. Color images of metric and even submetric (after pan-sharpening) ground resolution are available from diverse satellites. Although it usually takes several days for a satellite to fly again over

the same region, the relatively large number of satellites allows fast coverage of the area of interest. Some satellites also provide steering facilities, decreasing the time to reach a certain region. However, airborne sensors are more flexible. Optical data of the scene before the hazard stroke the region is usually available. Even in the case of developing countries basic optical data very often exists. Multiple automatic and semi-automatic algorithms have already been proposed to detect changes between two optical images. However, optical sensors need daylight and an unobstructed view of the scene of interest in order to produce meaningful data. This may not always be the case during or after hazardous events. Areas hit by hurricanes or floods are often located in tropical regions. Hence, dense cloud coverage limits the use of optical sensors. SAR sensors are not bound to the constraint of daylight and cloudless conditions. The radar technique actively transmits a signal and thus does not depend on daylight. A SAR sensor, either airborne or spaceborne, is capable of capturing data even at night, extending data collection time to 24 hours a day. The long wavelength compared to the optical wavelength facilitates the penetration of dense cloud layers. However, SAR images are not as easy to interprete as optical images since the viewing geometry and the wavelength domain are different. Different object features are measured and no color information is available. Additionally, effects like layover and shadowing, described in detail in 2.3.3, complicate information extraction. Considering the previously outlined advantages and disadvantages, it seems only logical to combine optical and SAR data. Details not detectable in optical images may appear in SAR images and vice versa. The fusion of optical and SAR images is usually conducted manually. However, manual fusion of multi-sensor imagery is a very time consuming process. In order to obtain accurate results, an experienced remote sensing expert is needed. Time critical applications call for easy to operate automated image processing. Therefore, one focus of CNES' (Centre National d'Études Spatiales) and IPI's (Institut für Photogrammetrie und GeoInformation) current research is on the automated fusion of optical and radar images as well as on information extraction of such fused data. In order to facilitate international cooperation and knowledge sharing, all developed algorithms are integrated into the open source software library ORFEO Toolbox (OTB[1]). OTB was introduced by CNES. This toolbox contains a set of algorithms for the exploitation of future submetric optical and radar images provided by the ORFEO program. ORFEO is a French-Italian high resolution Earth observation program. The CNES provides two optical satellites (Pléiades) and the ASI (Agenzia Spaziale Italiana) four radar satellites (CosmoSkyMed). Hence, one focus of ORFEO is on the joined exploitation of submetric resolution optical and radar imagery. Refer to Annex A for further details of the ORFEO program.

[1] see the OTB webpage `http://otb.cnes.fr`

1.2 State of the Art

The fusion of digital imagery is a valuable tool for the combined exploitation of imagery from multiple image sources or just different frequency bands. It is applied in such different fields like medical imagery, computer vision, close range photogrammetry and remote sensing. Hence, image fusion is not a technique limited to remote sensing imagery at all. For example, various algorithms have been introduced for medical imagery (e.g. the fusion of MRT and CT images) and proved to be very useful for this remote sensing project, too. As already mentioned in the very first sentences of the motivation, the field of high resolution imaging sensors, either airborne or spaceborne, has been enriched enormously during the past few years. This development facilitates the combined exploitation of multi-sensor, multi-resolution, multi-temporal and multi-frequency imagery. Additionally, fusion is not limited to merging one image with another image. Current research efforts also comprise the fusion of remote sensing imagery with GIS data and maps. Several different fusion approaches exist, usually based on pixel level or on feature level.

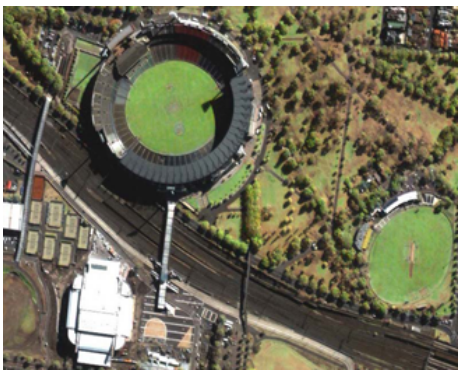


Figure 1.3: IKONOS 1 m pan-sharpened imagery showing the MCG and Tennis Center in Melbourne, Australia [Fraser, 2007b]

For example, the arithmetic combination of raster images on pixel level is commonly used for pan-sharpening (Fig. 1.3). Pan-sharpening is used in order to achieve both high spectral and spatial resolution in one image. It consists of the combination of high spectral resolution bands with a band of a high spatial resolution. It is conducted by multiplying the multi-spectral bands with the panchromatic band. A division by a synthetic intensity band follows up. Another technique commonly used in remote sensing is the principle component transformation (PCT) [Morrison, 1976] which may also be thought of as kind of a fusion technique. However, these fusion approaches usually call for data captured by the same kind of sensor. They fail as soon as great differences in both geometry and radiometry arise. This absolutely is the case for high resolution optical and SAR imagery. In order to achieve a fine

registration and fusion of submetric optical and SAR images, comprehensive modelling has to be done. Both geometry and radiometry have to be considered carefully. Therefore, another fusion approach, preferably on feature level, has to be developed. A variety of research efforts have already been made in order to merge optical and SAR data. One of the first papers from CNES, proposing a solution on feature level, is [Inglada and Adragna, 2001]. Therein, the registration issue is stated as finding the geometric transformation minimizing the distance between extracted features in both images. First, edges are extracted in both images. A distance image computing the Euclidean distances from the edges of the slave image to the edges of the master image is determined. The mean of this image is defined as the registration error. It is minimized by finding a rigid geometric transformation that well maps one image onto the other. A genetic algorithm is deployed for the minimization in order to avoid local minima. It is tested on rather low resolution ERS SAR and SPOT4 data. In [Inglada and Giros, 2004] several similarity measures, e.g. mutual information, are used for automatic fine registration of optical and SAR imagery. Local optimization of both similarity and deformation is proposed. After testing with SPOT4 and ERS2 images, deformation grids of sub-pixel accuracy are achieved. Furthermore, the automatic optical-SAR DEM estimation from a single satellite based on the deformation grid is proposed. Based on these results, [Inglada and Vadon, 2005] ortho-rectify and fine register SPOT5 and Envisat/ASAR images. Registration errors due to DEM errors are estimated. In [Galland et al., 2005] optical and SAR images are fine registered on feature level. Lines are extracted in both images. The registration process is used to refine comprehensive sensor models incorporating physically existing values (e.g. the sensor velocity and the terrain height). For the image transformation, a quadratic polynomial approach is chosen.

1.3 Objective of this Project

The objective of this project was the development of an image processing chain for the fusion of high resolution optical and SAR imagery. It was decided to develop a generic step by step approach, capable of registering images based on extracted features rather than on pixels. Indeed, at high resolution, pixel approaches fail on above ground structures. An approach on a higher semantic level is required. No detailed knowledge of the sensor parameters is needed.

Therefore, the fusion would later on be generally applicable to many kinds of remote sensing imagery. A registration strategy based on the projects outlined in the previous section was developed. The goal was to enable feature based registration of unrectified optical and SAR images with submetric resolution in urban areas. New aspects of this fusion approach are the very high resolution of the images, the introduction of a classification and the modular

construction of the entire processing chain based on the open source software library OTB. Modules not already existing were programmed in C++ and integrated into OTB. Each module of the processing chain was put to test. Usually, the results of several different techniques for each module were compared and the most promising was kept. However, the emphasis was not on the fine tuning of all processing parameters, but on the overall fusion strategy as well as on the implementation of the modules. This work is meant to serve as a basis and give new ideas for future research in the field of multi-modality image registration.

1.4 Structure of this Report

The structure of this report follows the developed image fusion process step by step. Chapter 2 provides the essential theory of remote sensing imagery. It consists of three main sections. Section 2.1 introduces the reader to the common physical background of optical and radar imagery: electro-magnetic waves. The two following sections describe in detail the most important properties of optical (section 2.2) and radar (section 2.3) sensors and their images. The radar section also explains the fundamentals of the *Synthetic Aperture Radar* technique which was used for capturing the test image.

Based on the essential theoretical background, the developed image registration strategy is introduced in chapter 3. First, the two test images, one optical image and one radar image, are displayed and their sensors are stated. Then, the reader is provided with an overview of all necessary registration steps. In the following sections and chapters, such image processing steps are explained in further detail. The first image processing step, the ortho-rectification of both images, is described in section 3.3 and the corresponding equations are given. Following up is section 3.4 explaining the preprocessing. In particular, the smoothing filters applied to the images and their corresponding mathematical modelling is outlined.

After the geometric rectification and preprocessing of both images, chapter 4 describes classification and feature extraction. Two classifications, the first one used as initialization for the second one, were applied to the images. The first classification is pixel-based deploying *Support Vector Machines* (section 4.1). Thereafter, a refinement of the results is achieved considering image statistics with a *Markov Random Field* classification (section 4.2). Sections 4.3 and 4.4 explain the algorithms implemented for feature extraction. The last section of this chapter briefly outlines the Danielsson technique for computing distance images from feature images (4.5).

Chapter 5 describes the registration that is necessary for the image fusion. First, an overview is given, providing the reader with the layout of the image registration framework. Thereafter, the modules of the registration framework are described in detail. Section 5.2 explains the transform, section 5.3 the interpolation, section 5.4 the similarity measure and section 5.5 the optimizer. The following chapter 6 displays and discusses the results achieved with the image fusion approach proposed in this report. The final chapter 7 gives a conclusion and some future perspectives for further enhancements.

2 Theoretical Background

This chapter familiarizes the reader with particular problems that complicate the image fusion process. It begins with the fundamental physical "concept" that optical and SAR sensors have in common: electro-magnetic waves. A brief introduction to this widespread topic with the focus on the sensors is given (section 2.1). In the following, the essentials of optical and SAR imagery are explained. In order to implement an appropriate automatic fusion process for optical and SAR images, it is absolutely necessary to carefully account for the different properties of the sensors. Two main fields have to be considered:

- radiometric properties and
- geometric properties.

In the following sections 2.2 and 2.3, the image capturing technique of the sensors is explained. First, the focus is on the sensors (2.2.1, 2.3.1 and 2.3.2). The later on geometric deformation modelling in 3.3 is based on considerations of these sections. Then, the image properties are described (2.2.2 and 2.3.3). Differences between optical and SAR images are emphasized.

2.1 Propagation and Scattering of electro-magnetic Waves

Electro-magnetic waves are the most commonly measured information source in remote sensing. Their precise determination in terms of signal wavelength, amplitude, plarization, and phase with airborne or spaceborne sensors allows for the gathering of chemical and physical information of the object of interest. Imaging sensors for Earth observation capture electro-magnetic radiation either in the visible and near-infrared spectrum or in the micro-wave spectrum (Fig. 2.1). The visible and near-infrared radiation is captured with optical sensors. This spectrum allows for the detection of chemical properties of the ground objects. For example, near-infrared radiation is strongly emitted by plants. The more chlorophyll the plants contain, the higher is the amplitude of the emitted near-infrared radiation. Furthermore, the amount of chlorophyll within a particular plant tells about its fitness. This fitness unveils ingredients of the soil and so on. Hence, lots of optical sensors (e.g. the LANDSAT and SPOT satellites) measure radiation in the near-infrared spectrum for environmental and agricultural applications.

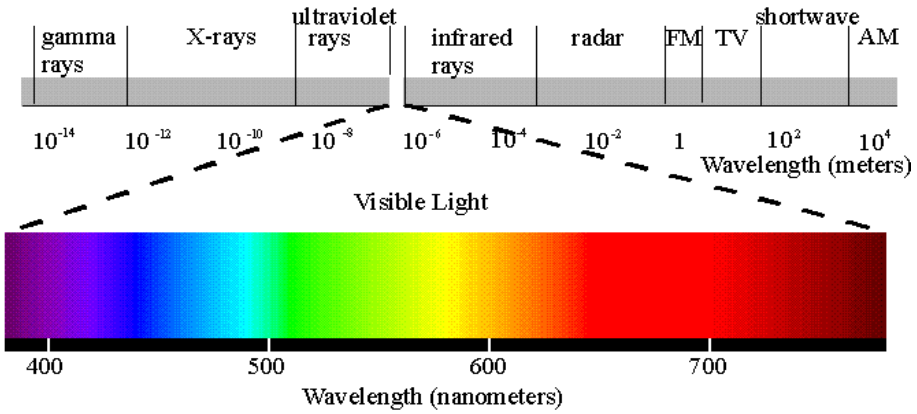


Figure 2.1: The spectrum of electro-magnetic waves (© University of South Carolina)

The microwave spectrum is captured with radar sensors, measuring mainly physical properties. Due to the radar's high sensitivity for water, humidity measures can be carried out. Furthermore, radar tells about the roughness and the electrical conductivity of objects. Constraints for the choice of the particular wavelength are imposed by the atmosphere. It limits the propagation of electro-magnetic waves to certain spectral bands within the optical and the microwave spectra. Very short waves with wavelengths smaller than 0.35 μm do not pass the atmosphere at all. Up to a wavelength of 14 μm, several spectral windows of different sizes exist. Between 14 μm and 1 mm, any radiation is completely absorbed by the atmosphere. The next spectral window opens up between 1 mm and 5 cm, enabling microwave techniques. Any radiation emitted by the sun (optical sensor) or the sensor itself (radar sensor) is bound to severe perturbations until it is finally captured. On its way to the Earth's surface, the electro-magnetic waves propagate through the Earth's atmosphere. They are reflected on the ground and propagate through the atmosphere a second time on their way to the sensor. The atmosphere as well as the ground have an impact on the properties of the waves and hence on the resulting image itself. In order to choose the appropriate wavelength for an application and to make use of occuring effects for image analysis, electro-magnetic wave propagation and diffusion have to be well understood.

While propagating through the atmosphere, electro-magnetic waves are affected by scattering. Scattering describes the effect of a direction change of the wave due to very small particles always present in the atmosphere. These particles are e.g. molecules and aerosols. The radiation is absorbed by these particles and immediatly emmitted again. While energy and wavelength stay constant, the propagation direction of the wave changes. Two main types of scattering are distinguished: Rayleigh Scattering and Mie Scattering. Rayleigh Scattering

occurs if the objects in the atmosphere are small compared to the wavelength [Sörgel, 2006]. The intensity I, affected by scattering, strongly depends on the wavelength λ.

$$I \sim \frac{1}{\lambda^4} \tag{2.1}$$

A well known example for Rayleigh Scattering in the optical spectrum is the blue color of the sky. Due to a smaller wavelength of blue light compared to red light, the blue waves are scattered ten times more than the red light. Hence, the sky occurs blue during the day. In the microwave spectrum of radar sensors, Rayleigh Scattering is due to rain drops. As a consequence, long wavelengths are desireable for weather independent remote sensing. It has to be considered that the molecules or atoms causing Rayleigh Scattering are rather regularly shaped. Hence, Rayleigh Scattering is almost isotropic.

Aerosol	Size $[\mu m]$
Vapour, fume, haze	0.001 - 0.5
Industrial fume	0.5 - 50
Fogg, clouds	2 - 30
Drizzle	50 - 200
Raindrops	200 - 2000

Table 2.1: Characteristic size of the most common aerosols [Kasser and Egels, 2001]

This is not the case for Mie Scattering. Mie Scattering occurs if the objects in the atmosphere are approximately of the same size as the wavelength. Within the optical spectrum, Mie Scattering is due to aerosols (see Tab. 2.1). For the microwave spectrum e.g. birds may cause Mie Scattering. Such objects are of rather irregular shape and therefore cause anisotropic scattering. Additionally, aerosols are distributed rather non-uniformly. Their density beneath five kilometers altitude varies over time, due to wind and human activities (e.g. industrial fumes).

2.2 Optical Imagery

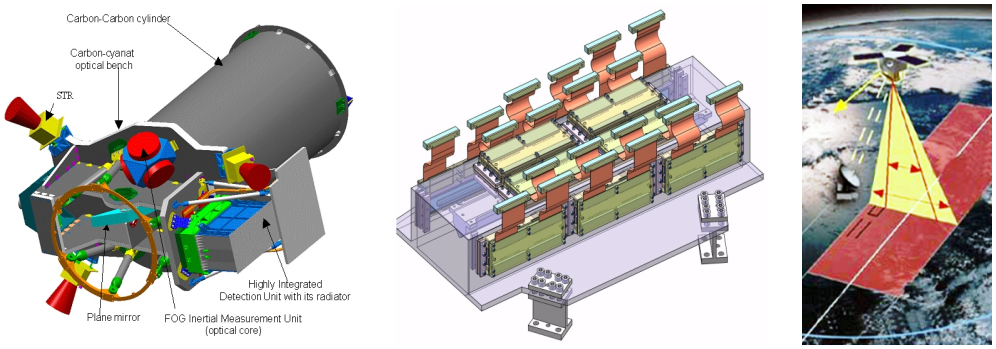
First remotely gathered optical images in large quantities date back to the first world war. Pilots, flying above the enemy's troops in propeller driven airplanes, leaned overboard and took pictures with analog cameras. Black and white images were developed from the exposed glass plates which were emulsified with a silver layer. Today's modern optical sensors capture digital images in various spectral bands. They determine the sensor's location in real-time, combining GPS (Global Positioning System) and IMU (Inertial Measurement Unit) measurements. Applications range from data acquisition for GIS (Geographic Information Systems)

and environmental monitoring to complete 3D modelling of cities (e.g. Google Earth, Microsoft Virtual Earth). A large variety of both airborne and spaceborne sensors exists, each of them particularly designed for a special application. The following sections will describe the basics of optical remote sensing imagery. Although analog airborne cameras are still in operation, the focus will be on digital sensors. Section 2.2.1 explains the sensors whereas section 2.2.2 describes the images.

2.2.1 Principle of Optical Sensors

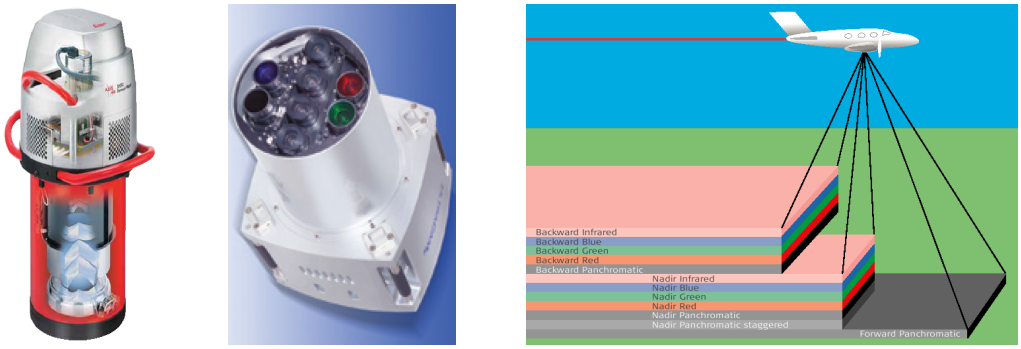
The imagery we use in this project is captured with digital sensors, either airborne or spaceborne. They are specifically designed to study the shape, details and contours of objects on the surface of the earth. Spectral bands of choice are the visible and near infrared spectra in rather broad spectral bandwidths. The panchromatic channel (bandwidth between 100 - 200 nm) around the visible spectrum (380 - 780 nm) enables the best spatial resolution. Usually, adding two or three narrower bands (bandwidths 50 - 100 nm) with a lower spatial resolution provides sufficient results for imaging purposes [Assemat et al., 2005]. The digital detectors inside the cameras are very sensitive, offering low noise and small pixels (often less than 10 x 10 μm^2). Two main configurations of the CCD detectors have to be distinguished: one-dimensional layouts with a linear array of several independent detectors and two-dimensional layouts that are composed of a mosaic of multiple detectors. Instruments with a linear detector configuration are so-called pushbroom sensors (Fig. 2.2(c)). Fig. 2.2(b) shows the linear CCD detector configuration of the Pléiades High Resolution (PHR) instrument of the Pléiades satellites while Fig. 2.2(a) shows the entire camera. Pushbroom detector layouts are also used in airborne sensors like the Leica ADS40 (Fig. 2.3(a)). An example for a detector composed of CCD mosaics, called frame camera, is the airborne sensor Vexcel ULTRACAM$_X$ (Fig. 2.3(b)).

For pushbroom scanning, a 2D image is created by moving a 1D line of p detectors over ground at high speed (Fig. 2.3(c)). In other words, the second dimension actually results from the fast forward motion. A linear CCD detector is situated perpendicularly to the velocity vector of the sensor. The time T_l, the sensor needs to pass over a pixel of size δx_{gr} (ground sample distance) on the ground, depends on its velocity on the ground V_{gr} ($T_l = \frac{\delta x_{gr}}{V_{gr}}$). For example, the HRG imaging instrument on board of the SPOT5 satellite, flying at an altitude of 832 km, samples 12,000 pixels every 5 m on the ground (6.5 μm pixel size in the focal plane) within 0.75 ms at a ground speed of 6.7 km/s [Assemat et al., 2005]. In order to enhance the ground sample distance, the SPOT5 satellite actually has two lines of CCD detectors. The second detector line is shifted by half a pixel thus decreasing the ground sample distance to 2.5 m. Frame camera images include p columns and n rows



(a) (b) (c)

Figure 2.2: (a) The layout of the Pléiades High Resolution (PHR) instrument with a primary mirror size of 650 mm (© CNES), (b) The focal plane assembly of the PHR instrument (© CNES), (c) Imaging principle of a satellite pushbroom line scanner [Fraser, 2007b]



(a) (b) (c)

Figure 2.3: (a) The airborne digital sensor Leica ADS40 2nd Generation is a line scanner (© Leica), (b) The airborne digital sensor Vexcel ULTRACAM$_X$ captures the light with a CCD mosaic of 13 arrays, 9 pan-chromatic and 4 color arrays (© Vexcel), (c) Imaging principle of the Leica pushbroom line scanner (© Leica)

of usually multiple CCD arrays. All columns are oriented in flight direction (parallel to the velocity vector) whereas the rows are oriented perpendicularly. Like amateur cameras, remote sensing cameras need a certain exposure time in order to capture images. This exposure time T_e has to be smaller than the time T_l the sensor needs to pass over the pixel on the ground. Between two successive image exposure centers there is a time delay T_d. It is used to read and to transfer the image which is still contained in the $n \cdot p$ registers of the CCD mosaic. For very high spatial resolution images the possible exposure time decreases due two low charge integration times of the CCD arrays. Smear effects can occur because T_e becomes

greater than T_l. Two different techniques exist in order to compensate for the smear effect. Their goal is to always keep the observed point on the ground immobile in the focal plane during exposure. The first possibility is to mechanically correct for the sensor motion. On some satellites a moving mirror is located near the perspective center and the CCD detectors capture only light reflected into the instrument by the mirror. This mirror compensates the satellite forward motion by slightly rotating backwards , i.e. in the opposite direction of the velocity vector during exposure. The second technique corrects the smear effect electronically within the CCD array. Charges within the CCD array are transferred to the neighboring row n in the direction opposed to the flight direction. The velocity of this charge transfer has to be synchronized with the actual sensor velocity. For airborne sensors the velocity is measured with navigational equipment, in particular with GPS. The parameters for satellites are derived from the orbit parameters which themselves are constantly refined using GPS and startracker measurements. The trajectory of satellites can be approximated locally very well with a Keplerian orbit. However, satellites are equipped with solar panels which are automatically adjusted with small electric motors from time to time. This adjustment results in small attitude changes of the platform and in micro-vibrations which can not completely be measured. Thus, the reduction of these residuals introduced to the images has to be conducted during post processing. Airborne sensors are exposed to more severe and abrupt attitude changes than satellites. Due to permanent turbulences in the air, the aircraft shows roll, pitch and yaw motions. The amplitude and frequency of such attitude changes is directly related to the turbulences in the atmosphere. Hence, attitude changes along the three axis of the sensor have to be measured continuously deploying an inertial measurement unit (IMU). These high frequency measurements are complementary to the rather low frequency GPS measurements. IMU measurements fill the gap between the GPS positions. However, GPS provides a much better long term stability than the IMU. Usually, an approach based on Kalman filtering is applied in order to combine both IMU and GPS measurements. Thus, the IMU drift is corrected with the very stable GPS data. However, very high frequency turbulences due to thermal lift, in particular above urban areas, cannot be measured directly and would result in image distortions. This effect can be prevented by installing the sensor on a gyro stabilized platform which corrects for high frequency attitude changes in real time.

Besides image distortions due to the sensor motion, further distortions have to be considered: small errors are introduced by its optics, mainly decentering distortion (dx_{dist-d}, dy_{dist-d}) and radial lens distortion (dx_{dist-r}, dy_{dist-r}) (refer to Annex C for the corresponding equations), distorsions of the focal plane ($dx_{unflatness}$, $dy_{unflatness}$), errors in the interior orientation elements (dx_{IO}, dy_{IO}), atmospheric refraction ($dx_{refraction}$, $dy_{refraction}$) and the earth curvature ($dx_{earthcurvature}$, $dy_{earthcurvature}$) [Fraser, 2007a]. Decentering effects are due

to lenses that are not symmetrically mounted on a straight line. The radial lens distortion, usually small but typical for wide angle airborne cameras, results in either barrel distortion or pincushion distortion (Fig. 2.4) of the image.

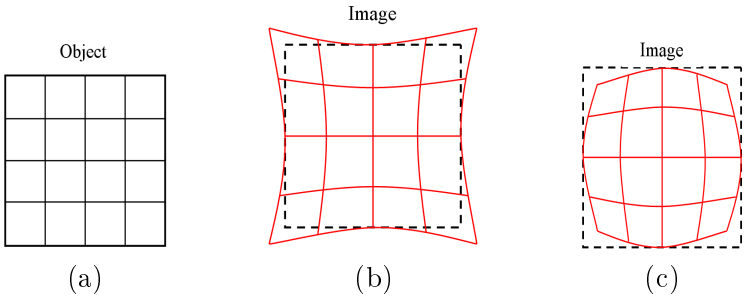


Figure 2.4: (a) Object, (b) Barrel distortion, (c) Pincushion distortion

Focal plane unflatness effects in CCD arrays, although usually small, lead to displacement effects of image points and have to be accounted for. Calibration of the sensor may not always be conducted under laboratory conditions because the focal plane assembly of some digital frame cameras consisting of several 2D CCD arrays becomes unstable under flight conditions [Jacobsen, 2007]. This distortion is highly correlated with residuals in the interior orientation parameters. In case of the classical frame camera the parameters of the interior orientation are the offset of the principle point from the CCD array center $(\delta x, \delta y)$ in the image plane and the focal length (f). Atmospheric refraction occurs because light follows a curved path as it passes through layers of different atmospheric pressure. The refraction correction is radially inwards for near-vertical airborne sensors. It is made based on a standard atmosphere for both airborne and spaceborne sensors. Earth curvature has to be corrected for if the swath width of the sensor becomes significantly large and non-cartesian reference coordinate systems (e.g. UTM) are used. All previously described physical deviations from a straight line between the object point, the perspective center and the principle point of the sensor give rise to a perturbation (dx, dy) in the image point location. The correction for these perturbations is modelled with Eq. 2.2. Refer to [Fraser, 2007a] for the detailed calculation of the perturbations.

$$\begin{aligned} dx &= dx_{dist-r} + dx_{dist-d} + dx_{IO} + dx_{unflatness} + dx_{refraction} + dx_{earthcurvature} \\ dy &= dy_{dist-r} + dy_{dist-d} + dy_{IO} + dy_{unflatness} + dy_{refraction} + dy_{earthcurvature} \end{aligned} \tag{2.2}$$

Although airborne and spaceborne sensors show many similarities, important differences exist. The main difference between airborne and spaceborne optical sensors is their field of

view. It depends on the size of the CCD array in the focal plane and the focal length. In order to capture images from space with a reasonably high resolution, the focal length is chosen very long and hence the field of view is small. For example, the IKONOS satellite launched in 1999 has a field of view of 0,9°. In consequence, the light rays captured by the spaceborne sensor are almost parallel which leads to less occluded areas in cities. In other words, the spaceborne sensor can look down to the street level next to skyscrapers whereas airborne sensors will only see the building facades.

2.2.2 Properties of Optical Imagery

One of the inputs for this project is a digital, optical, grey scale image. It can be interpreted as a simple matrix, which represents the grey levels of a scene. In Fig. 2.5 we can clearly see that each grey value of the image actually is a digital number within the matrix. Due to quantization, this grey value is always an approximation of the radiometry. Its spatial location is provided by the image coordinates (x, y) of the matrix element (pixel). In mathematical terms we can say that an image consists of a regularly sampled two-dimensional function. It provides a value proportional to the brightness emitted by the scene for each sample point [Assemat et al., 2005].

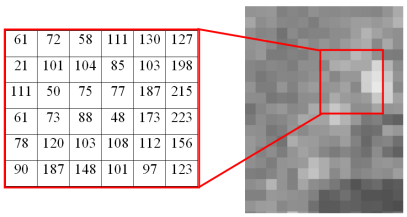


Figure 2.5: Image matrix of a grey scale image (grey values between 0 and 255)

The brightness of a pixel displays the reflectivity of an object on the ground in the spectral band of the sensor. Most sensors are capable of capturing data in several spectral bands (see also 2.2.1). Each band can be displayed as an own two-dimensional image. Therein, any pixel is assigned a reflectivity value of the corresponding wavelength. Hence, images captured by sensors with multiple spectral bands may consist of several layers of subimages (which do not necessarely have the same pixel size). The curve of the reflectivity magnitude of objects over the entire spectrum is called their spectral signature. It is distinct for many land cover types and thus very useful for classification purposes. In order to specify certain criteria for image product requirements, a number of standard, quantifiable quality measures for optical imagery are necessary. They ensure that an image product meets user needs. In terms of

resolution, three different categories exist for evaluating digital imagery: the spectral, the spatial and the radiometric resolution. The spectral resolution is the ability of the sensor to discriminate different wavelengths of the electro-magnetic spectrum. It incorporates both the number of spectral bands as well as the corresponding spectral bandwidths. Spatial resolution describes the smallest angular or linear object separation on the ground that can be resolved. This measure has to be carefully distinguished from the ground sample distance (GSD) which is the size of an image pixel projected onto the ground. In fact, the spatial resolution can be higher than the GSD since interpolation techniques enable subpixel resolution. Additionally, bright features that are much smaller than the GSD may spread over one or several pixels. Another important quality descriptor is the radiometric resolution. It measures the sensitivity to differences in signal strength of the radiant flux received by the sensor [Fraser, 2007b]. For example, within a panchromatic image the sensitivity translates to the number of grey levels between black and white. Another important quality measure related to radiometry is the noise level. It is the ability of the sensor to obtain a uniform image for a uniform landscape. For cartographic mapping purposes, planimetric and elevation accuracy of the imagery are also important.

In order to achieve high quality images in terms of geometry and radiometry, much attention has to be paid to its generation. A mathematical model is introduced in order to prevent disturbing effects, e.g. aliasing. The appropriate model varies with the sensor it is applied to. The linear model briefly outlined in the following was developed for imagery taken by the French SPOT1 - SPOT4 satellites (refer to [Assemat et al., 2005] for a more detailed description). The raw image captured by a SPOT satellite is assumed to consist of two summands, the ideal image folded with a filter modelling the sensor device (the so-called instrument's impulse response) and noise. A two-dimensional Dirac comb discretizes the continuous input signal. The discrete image is transfered to a spectral representation involving a Fourier transform of the raw image. A Fourier transform is also applied to the instrument's impulse response, which is then called a Modulation Transfer Function (MTF). The MTF expresses the attenuation factor for spatial frequencies of the imaged scene on the ground. In order to capture high spatial frequencies, the MTF has to be high. In case the MTF is too low, undersampling of the image occurs. Three main parameters have to be specified for the entire mathematical model: One geometric parameter, the sampling grid, and two radiometric parameters, the MTF and noise. The sampling interval of the orthogonal grid is the product of the satellite's velocity on the ground and the sampling time (see also section 2.2.1) and the noise is specified in terms of the mean signal-to-noise ratio (SNR). The MTF is defined as the product of the optics MTF, the detector MTF and the image motion MTF along a column (which is parallel to the velocity vector).

2.3 Radar Imagery

This section will first explain the basic radar theory for the detection of targets as well as the fundamentals of imaging radar sensors (section 2.3.1). The development of the Synthetic Aperture Radar (SAR) technique led to a significant improvement of the resolution in azimuth direction of imaging radar sensors (2.3.2). However, the SAR technique incorporates some constraints which have to be taken into consideration when analysing such images. Radar sensor properties and SAR characteristics in particular will be explained in section 2.3.3.

2.3.1 Radar Principle

The acronym radar stands for *Radio Detection and Ranging*. It was originally developed by the military for the detection of ships and aircrafts. First radar developments date back to the time between the first and the second world war. Radar enables the detection of the direction of an object in relation to the radar sensor as well as the range between the radar sensor and the object. Hence, the position of an object detected by a radar sensor can be determined if the position of the sensor itself is known with sufficient accuracy. The radar usually deploys electro-magnetic waves in a frequency band of 0.225 GHz to 36.0 GHz which translates to wavelengths between 133 cm and 0.83 cm (refer to Tab. 2.2 for radar frequency bands). However, radar sensors for particular applications with lower frequencies (e.g. over-the-horizon coastal radar systems, 3 - 30 MHz) or higher frequencies exist.

Band	Frequency Interval	Wavelength
P	0.225 - 0.390 GHz	133 - 76.9 cm
L	0.39 - 1.55 GHz	76.9 - 19.3 cm
S	1.55 - 4.20 GHz	19.3 - 7.1 cm
C	4.20 - 5.75 GHz	7.1 - 5.2 cm
X	5.75 - 10.90 GHz	5.2 - 2.7 cm
K_u	10.90 - 22.0 GHz	2.7 - 1.36 cm
K_a	22.0 - 36.0 GHz	1.36 - 0.83 cm

Table 2.2: Radar frequency bands

Radar sensors have two main advantages compared to optical sensors. The first advantage is the longer wavelength of the radar, which enables ground imaging even through dense cloud coverage. The second advantage is the radar's capability of collecting data at both day and night time. Radar sensors, emitting and receiving signals, are called "active". Optical sensors capture the ground reflection of the sun light and are called "passive" sensors. Thus, optical sensors are only capable of capturing meaningful data at daytime while radar sensors are daylight independent.

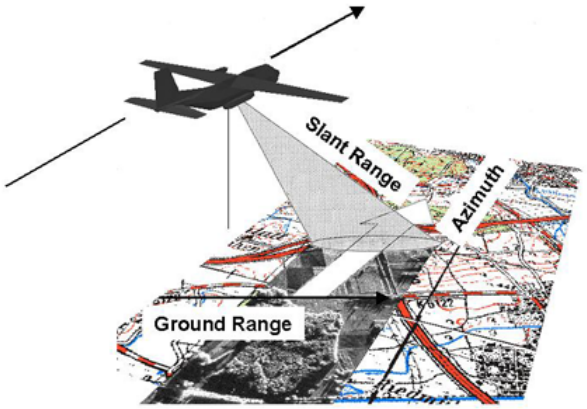


Figure 2.6: Principle of Side Looking Airborne Radar (SLAR) (© Institut für Photogrammetrie (ifp), Universität Stuttgart)

While the first applications consisted of target detection for military applications, first imaging radar sensors were introduced in the 1950s. Radar with real aperture (RAR) was installed on aerial platforms for mapping purposes. This technique is called Side Looking Airborne Radar (SLAR), since the sensor is installed on one side of the aircraft (Fig. 2.6). The SLAR signal consists of a series of short, coherent micro-wave pulses, transmitted aslant to the ground, perpendicularly to the flight direction. The signal reflections from the ground are captured by the sensor. Thus, the distance between sensor and object (slant range) is determined by multiplying the time of flight of the signal with the speed of light. It has to be halfed since the signal propagates forth and back the same distance. The power of the backscattered signal depends on the sensor design, the backscattering properties of the object (geometric shape, directivity, reflectivity) and the radar equation (Eq. 2.3).

$$P_E = \frac{P_S \cdot G^2 \cdot \lambda^2 \cdot \sigma}{(4\pi)^3 \cdot r^4 \cdot L_v} \tag{2.3}$$

P_E: received power $[W]$
P_S: transmitted power $[W]$
G: antenna gain of the receiving antenna $[dB]$
L_V: dimensionless factor subsumming the overall system loss
λ: wavelength $[m]$
σ: radar cross section $[m^2]$
r: range distance $[m^2]$

The time of flight of the signal and its intensity are captured by the sensor. The time of flight determines the distance whereas the intensity leads to the image grey value. Length, width and orientation of the imaged area on the ground depend on the antennas position

and orientation as well as on the transmission direction of the micro-waves. The transmission direction varies due to roll, pitch and yaw motions of the sensor platform (aircraft or satellite). The geometric resolution of a radar sensor is usually anisotropic. With resolution, we mean the minimal distance between two objects on the ground that still allows for their distinction. The resolution in flight direction (along-track, azimuth resolution) differs from the resolution perpendicular to the flight direction (across-track, range resolution). The geometric slant range resolution δ_{sr} depends on the duration of the transmitted pulse τ and the speed of light c (Eq. 2.4). The pulse duration itself derives from the signal's bandwidth B. The geometric resolution δ_{gr} on the ground depends on c, τ and on the local look angle θ_L (Eq. 2.5).

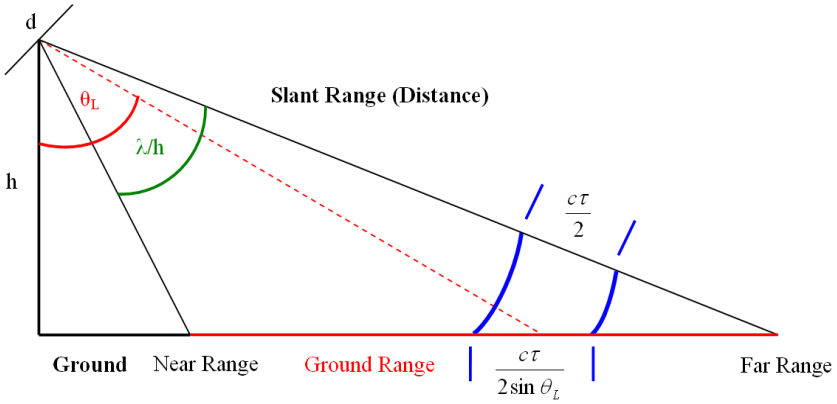


Figure 2.7: Across-track resolution of a RAR sensor; d: antenna length, h: altitude of the sensor, λ: wavelength and λ/h: angular resolution

$$\delta_{sr} = \frac{c\tau}{2} \approx \frac{c}{2 \cdot B} \tag{2.4}$$

$$\delta_{gr} = \frac{c\tau}{2sin\theta_L} \tag{2.5}$$

Regarding Fig. 2.7, it becomes obvious that the ground range resolution degrades by moving the sensor's inclination to the nadir (decreasing look angle). A more aslant perspective improves the ground resolution. Usually, the across-track diffraction angle (in elevation direction) is chosen large in order to achieve an enhanced swath width. On the other hand, the angular resolution θ_a in along-track (azimuth) direction is chosen very small because it leads to the corresponding ground resolution. The geometric ground resolution in azimuth direction can be approximated by multiplying θ_a with the distance to the imaged object (Eq. 2.6).

$$\delta_{ra} = \theta_a R \approx \frac{\lambda R}{d} \tag{2.6}$$

Therefore, an enhancement of the geometric ground resolution in azimuth direction with a RAR sensor requires a small wavelength and a long antenna. Additionally, the very long distance between sensor and object, in particular for spaceborne sensors, calls for extremely long antennas in order to achieve a sufficient resolution. However, neither the construction of infinitely long antennas (the spacecraft limits the satellites extent and weight) nor infinitely small wavelengths (a smaller wavelength increases the losses within the atmosphere) are possible. Hence, another radar technique is usually used for high resolution Earth imaging applications: *Synthetic Aperture Radar (SAR).*

2.3.2 SAR Technique

The SAR technique improves the spatial resolution in azimuth direction. It is based on the idea of simulating one large antenna out of several measurements. Measurements are taken along the trajectory of the sensor with a certain pulse repetition frequency (PRF). Multiple radar pulses are emitted and received. Hence, an object on the ground is illuminated by multiple radar pulses as long as it is located inside the footprint of the sensor. The reflected pulses, send from different antenna positions at different moments, are combined for the simulation of one long antenna. This is a contrast to the RAR approach. The RAR technique considers a ground object to be illuminated only once. No combination of the reflected pulses is conducted. In fact, the length of the synthesized SAR antenna equals the distance between the first and the last antenna position from which a ground object is illuminated. The mathematical model, well explaining the synthesis of one long antenna, is based on the Doppler shift f_D (Eq. 2.7).

$$f_D = \pm\frac{2 \cdot v_{rel}}{\lambda} \tag{2.7}$$

A Doppler shift between different sensor antenna positions occurs. It is due to the change of the relative velocity along the line-of-sight v_{rel} between antenna and object. v_{rel} changes because the distance between sensor and object changes whereas the absolute sensor velocity is constant. In fact, the distance curve is a parabola. Its apex is located at the shortest distance r_0 between sensor and object. This also is the point of the lowest relative velocity, i.e. the Doppler frequency is zero f_{D0}. In order to combine all radar echos for the synthesis of one long antenna, the distance variation has to be accounted for. This goal is achieved by evaluating the Doppler shift. The great advantage of SAR over RAR is the SAR's improved spatial azimuth resolution. In contrast to RAR, it is completely independent from

the sensors distance to an object and its wavelength (compare Eq. 2.6 and Eq. 2.8). It can be approximated by the half length D of the synthesized antenna:

$$\delta_{SARaz} = \frac{D}{2} \tag{2.8}$$

As a matter of fact, the azimuth resolution of SAR improves with increasing antenna length D whereas the RAR's azimuth resolution deteriorates. The SAR signal u itself is complex and consists of a real part u_i and an imaginary part u_q (Eq. 2.9). Both complex components may be displayed as cartesian coordinates of the signal (Fig. 2.8). Additionally, the pixel value of a SAR image is the sum of multiple coherent signal reflections on the ground (Eq. 2.10). Various independent scatterers within one resolution cell contribute to the final signal received by the sensor [Goodman, 1985; Sörgel, 2006]. The fact that the pixel value is the coherent sum of a large number of complex signals also leads to the speckle effect. It is described in further detail in the following section 2.3.3.

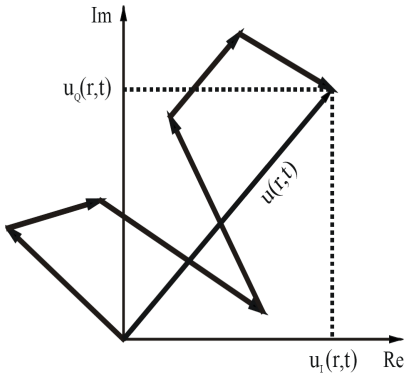


Figure 2.8: Complex cartesian representation of the SAR signal

$$u = u_i + ju_q \tag{2.9}$$

$$u_i = Re\{u\} = \frac{1}{N}\sum_{n=1}^{N} a_n cos\phi_n \ , \quad u_q = Im\{u\} = \frac{1}{N}\sum_{n=1}^{N} a_n sin\phi_n \tag{2.10}$$

In order to model SAR images appropriately, e.g. for classification purposes, the image statistics have to be well understood. Statistics of singlelook intensity images follow an exponential distribution. Multilook intensity images are χ^2-distributed. This is not the case for amplitude images. Singlelook amplitude images are Rayleigh distributed while multilook images follow a χ-distribution.

2.3.3 Properties of SAR Imagery

SAR images have certain properties that may cause difficulties during image analysis for unexperienced interpreters. A mapping of a plane from the ground to the image is non-linear [Sörgel, 2003] since the radar principle consists of measuring distances between the sensor and the object (Fig. 2.9(a)). For human interpreters this fact appears somehow disturbing at first because it does not correspond to our eye perception. The radar principle of distance measures causes layover, shadowing and foreshortening effects.

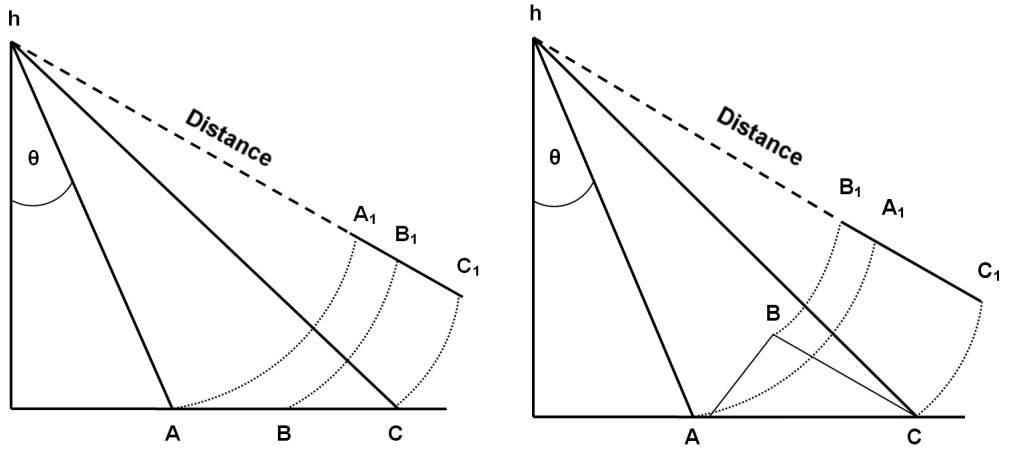


Figure 2.9: (a) Mapping of flat ground, (b) Layover

Layover appears if the inclination of a plane is higher than the look angle of the radar sensor (Fig. 2.9(b)). For example, the distance between the highest point of the mountain (B) and the radar sensor is smaller than the distance between the lowest point (A) and the sensor. Hence, the inclination of the plane on the ground is higher than the look angle θ. The highest point is mapped closer to the sensor (B_1) than the lowest point (A_1). This effect also appears in urban areas at buildings because vertical building facades in general have a higher inclination than the look angle. In conclusion, the buildings' facades appear upside down in the image. The very bright lines (due to double bounce effects of the signal) where walls meet the ground are mapped onto the roof.

Shadowing occurs if the inclination of a plane facing away from the sensor is bigger than the corresponding look angle (Fig. 2.10(a)). Point D is not mapped into the radar image since it is located within a shadowed area. It is occluded by the mountain top B. Shadowed regions appear dark in radar images. Shadowing poses serious image analysis problems, in particular in urban areas. An increasing look angle results in more severe shadowing effects, i.e. layover

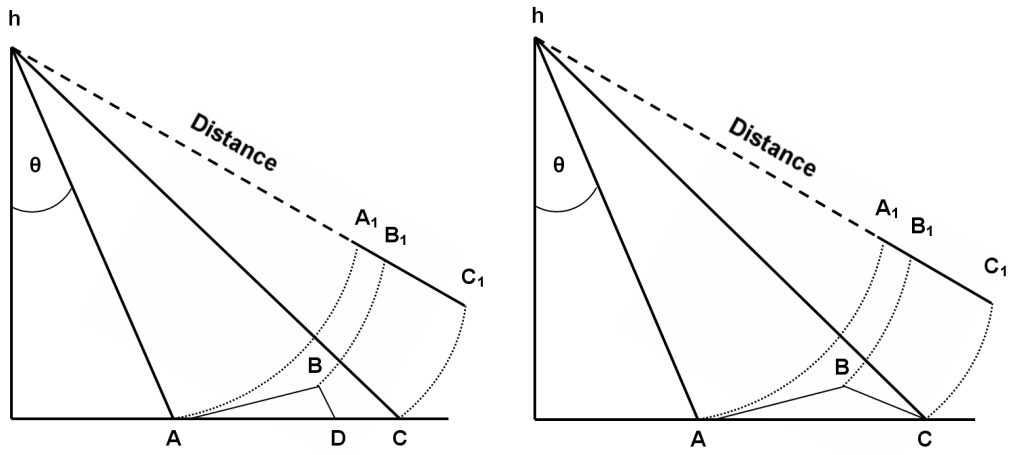


Figure 2.10: (a) Shadowing, (b) Foreshortening

declines and the resolution is enhanced. As a trade off, more areas are dark in the image. In optical imagery objects may also be obstructed by higher objects on the ground due to a not pure nadir view of the sensor. However, the best resolution in case of optical imagery is achieved in nadir direction which also results in the least obstructed regions. Hence, in optical imagery the opposite relation of resolution and occluded regions is valid. This fact is a very important difference between radar and optical sensors. It constraints both imaging techniques to certain applications.

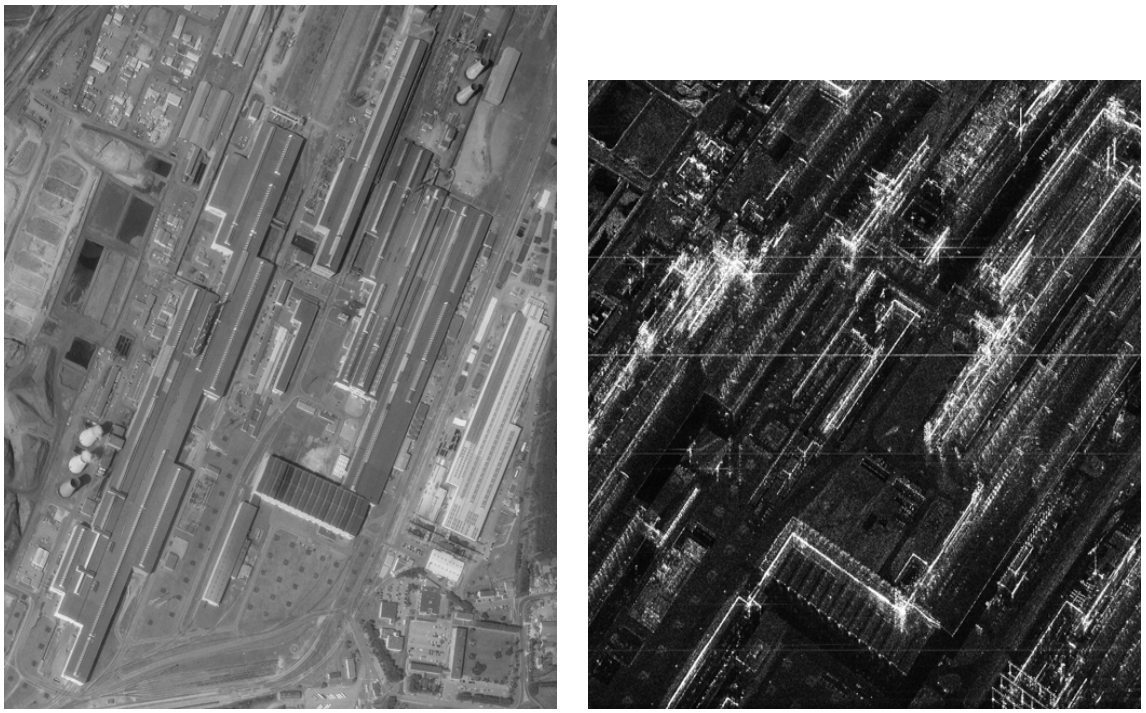
The third effect, typical for radar images is foreshortening (Fig. 2.10(b)). It appears because the ground terrain usually is not flat. Inclined planes facing towards the sensor are mapped shorter in the image than they appear on the ground. They appear brighter in the image since the entire energy of the ground area sums up in a relatively small area in the image [Sörgel, 2003]. Another effect disturbing SAR images is speckle. It appears because the resolution cell size of the SAR sensor is usually bigger than the wavelength (refer to Eq. 2.10). Hence, the captured signal intensity within an resolution cell is in fact the coherent sum of multiple interfering signal responses. In case the ground object's surface is rough compared to the wavelength, many signals contribute to the measured overall sum. This sometimes results in a very high amplitude and in a low amplitude at other locations. Affected areas show high contrast between neighboring pixels (salt-and-pepper). Additonally, speckle is distributed non-uniformly. Hence, any kind of image analysis has to consider a particular statistical speckle and signal model. A detailed description of the speckle filter chosen for this project is given in chapter 3.4.2.

3 Image Registration

The image registration approach of this project applies a general model to the input remote sensing images. This approach was chosen because precise sensor parameters may not always be distributed with the images. Instead of modeling each sensor model separately, only two models are used in total. The first model is applied to all optical sensors while the second model is taylored for all SAR sensors. Each model is set up in order to account for both aerial and spaceborne sensors. This fact implies that both models are not capable of geometrically rectifying all existing sensors with equal precision. Obviously, residuals will always rest if such different sensors as e.g. pushbroom and frame optical sensors are treated with an identical geometric model. Therefore, residuals remaining after the first registration step are treated in the following component. Hence, the registration precision is refined graduately. This chapter introduces the entire image processing chain. Its final result is a deformed SAR image, registered onto the optical image. The first section of this chapter (3.1) introduces the test data. Then, section 3.4 outlines the complete registration strategy. It gives an overview of all necessary computation steps for the final goal of fused images. The following sections describe the processing that is needed in order to prepare for feature extraction. Geometric distortions have to be accounted for first. This image rectification step is explained in detail in section 3.3. As soon as the images have been projected from camera space to the ground, preprocessing is carried out. The smoothing and despeckling filters of choice are described in section 3.4. Thereafter, the optical and the SAR image are sufficiently prepared for classification and feature extraction, which is explained in the following chapter (4).

3.1 Test Data

Images from airborne platforms are used as test data. Our developed registration strategy is tested on two images, one optical and one SAR image. Both images cover approximately the same area. They were captured over an industrial zone of the city of Dunkerque in the north of France. The optical image was taken with an airborne sensor of the Institut Géographique National France (IGN). Its spatial resolution (pixel size on the ground) is about 0.3 m and its original size is 3033 by 2559 pixels (Fig. 3.1(a)).



(a) (b)

Figure 3.1: (a) Optical aerial image taken by IGN (© IGN), (b) SAR aerial image in slant range geometry taken by ONERA with the RAMSES sensor in X-Band (© DGA)

The SAR image was taken by the French Aerospace Lab (ONERA) in X-Band (Fig. 3.1(b)). It was captured with the airborne SAR sensor RAMSES and thus is courtesy of the French Délégation Générale pour l'Armement (DGA). Its original size is 2048x2048 pixels. The real part Re and the imaginary part Im of the signal were separately registered in two different layers of a layer stack image. In order to visualize the SAR image as a grey value image, it is possible to either compute intensity values (Eq. 3.2), amplitude values (Eq. 3.1) or decibel values. The intensity image is exponentially distributed (Fig. 3.2(b)). It has the disadvantage that the actuarial expectation $E[I]$ of the image intensity equals its variance $\sigma(I)$. Hence, the speckle effect has a strong and multiplicative impact on the image. The amplitude image shows a Rayleigh Distribution of its grey values (Fig. 3.2(a)).

$$A = \sqrt{Re^2 + Im^2} \tag{3.1}$$

$$I = Re^2 + Im^2 \tag{3.2}$$

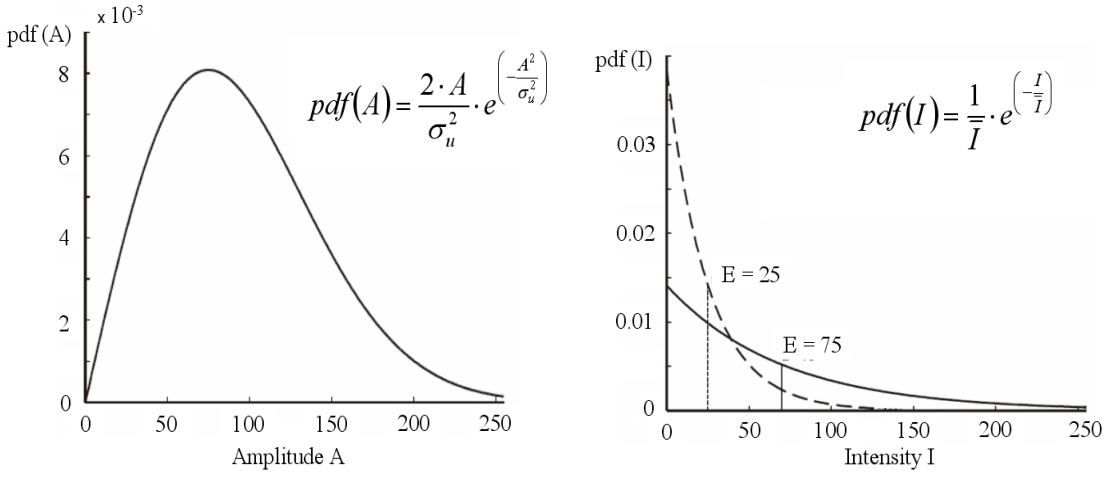
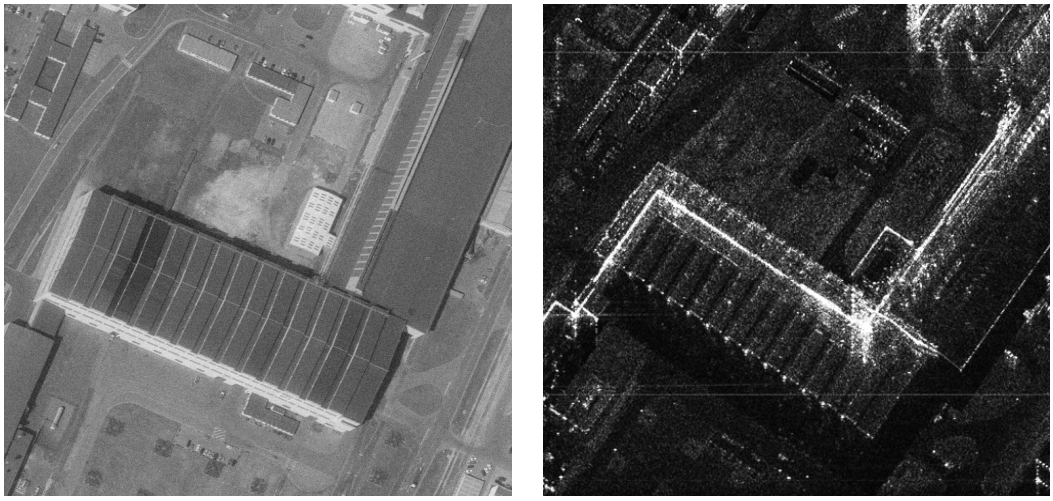


Figure 3.2: (a) Rayleigh distribution of the SAR amplitude, (b) Exponential distribution of the SAR intensity for two different actuarial expectations

SAR Image	Minimum	Maximum	Mean	Std. Dev.
Before rescaling	0.000057	261.29	0.341369	0.969359
After rescaling	0	255	53.3	55.7

Table 3.1: Parameters of the SAR amplitude image Fig. 3.1(b) before and after rescaling

However, the resulting amplitude image showed a very large dynamic range (more than 40 dB) of the grey values, compared to the optical one. Most regions of interest were displayed with extremely low contrast almost black (see Tab. 3.1). Hence, thresholding and rescaling of the SAR amplitude image were conducted. In order to determine the appropriate threshold, the next step was to calculate mean and standard deviation of the amplitude image. All amplitude values above the mean plus the standard deviation were thresholded. Pixel grey values above the threshold were set to the threshold itself. Thereafter, the image was linearly rescaled (between 0 and 255). The resulting amplitude image (Fig. 3.1(b)) shows sufficiently contrast and thus facilitates interpretation. Both images were found to be too large for testing purposes because the implemented algorithms would have been computationally too expensive. None the less, such algorithms can be optimized for treating very large images under operational circumstances. However, this was not in the focus of this project. Hence, two smaller regions were extracted (Fig. 3.3(a) and 3.3(b)). The optical test region has a size of 750 by 736 pixels whereas the SAR test regions' size is 990 by 920 pixels. All implemented algorithms were tested on these two regions.



(a) (b)

Figure 3.3: (a) Test region of the optical image, (b) Test region of the SAR image

3.2 Registration Strategy

The developed registration strategy is displayed in Fig. 3.4. Its main idea is to register corresponding features that are extracted from both the optical and the SAR image. Feature extraction is necessary because we deal with very high resolution imagery. Different imaging properties of the sensors develop their full extent. Hence, fine details look very different in the optical and in the SAR image. Classical approaches based on pixel level, e.g. normalized cross-correlation, were developed for the direct registration of rather low resolution images. Usually, those approaches also require images taken by the same kind of sensor. They will fail to provide subpixel registration accuracy in our case of multi-modality imagery because the images do hardly show any similarity. Line features enable good registration results because our images cover an urban area. Due to man-made structures like roads and buildings, they can be found in high quantities. The first part of the registration strategy thus consists of preparing the images for feature extraction and line extraction. Optical and SAR image are treated separately during ortho-rectification, preprocessing, classification, feature extraction and the computation of distance images. Thereafter, the second part treats both images jointly. The SAR image is registered onto the optical image deploying an algorithm provided by the open source software library ITK (National Library of Medicine Insight Segmentation and Registration Toolkit)[1].

[1] see the ITK webpage `http://www.itk.org`

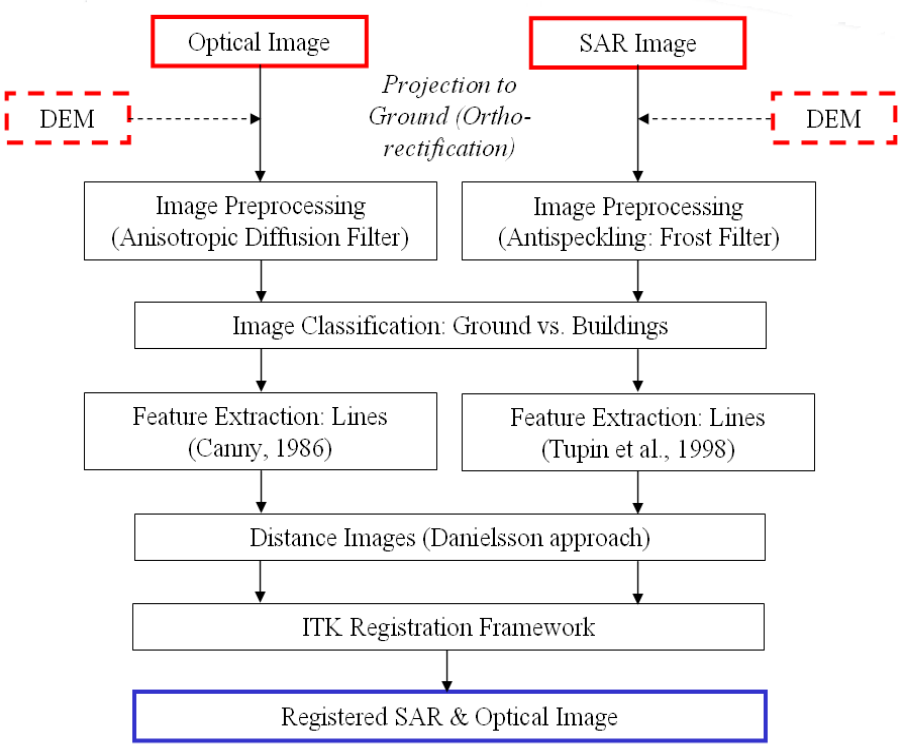


Figure 3.4: Image registration strategy

1. The first registration step consists of an ortho-rectification of both optical and SAR image (section 3.3). Both images are projected from sensor space to object space. An external digital elevation model (DEM) is used in order to rectify distortions introduced by the terrain. Objects included in the DEM are rectified whereas objects which are not included stay distorted, in particular buildings. The transformations also account for sensor specific residuals.

2. The second registration step applies appropriate preprocessing filters to the rectified images (section 3.4). This registration component reduces the noise level and prepares the images for further image analysis. An anisotropic diffusion filter is applied to the optical image. For despeckling reasons, the SAR image is filtered with the Frost filter [Frost et al., 1982] (section 3.4.2).

3. The third image registration component applies a classification to both images (sections 4.1 and 4.2). It distinguishes between rectified ground regions and unrectified buildings. This step is necessary because only rectified areas of the images can be considered for the registration of the images.

4. After all rectified image regions have been identified, features are extracted (chapter 4). The outcome of this fourth registration step are two images displaying the extracted features.

5. Distance images are calculated from the feature images in the fifth registration component (section 4.5). They are registered using the ITK registration framework.

6. The ITK registration framework's architecture is modular (section 5.1). It calls for a particular combination of metric, transform, optimizer and interpolator in order to adapt it to the input images.

7. After convergence of the registration towards an optimal solution, the final registration parameters are used to register the SAR image onto the optical image.

3.3 Geometric Deformation Modelling

Modelling has to be conducted in order to account for the different viewing geometries of SAR and optical sensors (Fig. 3.5). Distortions due to the terrain have to be taken care of, too. Both images are projected from sensor space to the ground as a first geometric registration step. A general a priori model-based approach was chosen, capable of rectifying the images without in depth knowledge of sensor parameters. However, current work also comprises the integration of the software library OSSIM into OTB. Once integrated, OSSIM will enable the usage of the precise geometric sensor model for each sensor. An external DEM is used to reduce distortions introduced by rough terrain.

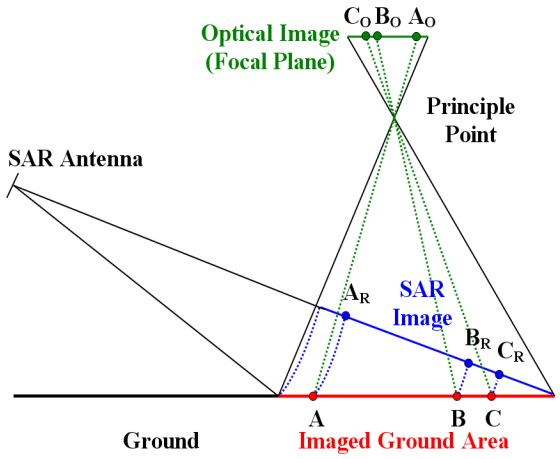


Figure 3.5: Comparison of optical and SAR viewing geometry

3.3.1 Geometric Optical Model

Usually, deformation modelling of residuals due to the particular architecture of the instruments is conducted. However, this registration approach was designed in order to deal with multiple kinds of sensors (see section 2.2.1). Since knowledge of the particular parameters of each sensor is lacking, a general geometric sensor model is adapted. In case of optical imagery, the inverse 3D collinearity equations (object to image) are used in order to project the image to the ground (Fig. 3.6(a)).

$$x = x_0 - f * \frac{r_{11} * (X - X_C) + r_{21} * (Y - Y_C) + r_{31} * (Z - Z_C)}{r_{13} * (X - X_C) + r_{23} * (Y - Y_C) + r_{33} * (Z - Z_C)} \tag{3.3}$$

$$y = y_0 - f * \frac{r_{12} * (X - X_C) + r_{22} * (Y - Y_C) + r_{32} * (Z - Z_C)}{r_{13} * (X - X_C) + r_{23} * (Y - Y_C) + r_{33} * (Z - Z_C)} \tag{3.4}$$

- X, Y, and Z are the ground coordinates,
- x_0, y_0, and f are the parameters of the interior orientation of the sensor,
- X_C, Y_C, and Z_C are the coordinates of the sensor's principle point,
- r_{ij} are the elements of the rotation matrix at line i and column j. The elements of the rotation matrix are based on the three rotation angles ω, φ and κ (see Tab. 7.2 in 7.2).

Each pixel of the image on the ground is transformed via the previously displayed equations to the original image. An indirect geometric image transformation for each pixel *P(X, Y)* of the ortho-image is conducted (Fig. 3.6(b)). The pixel size of the ortho-image is selected corresponding to the ground resolution of the sensor. A raster pixel size of the DEM does not necessarily have to be the same as the ortho-image pixel size. For all raster points of the ortho-image on the ground, the corresponding height values have to be interpolated within the DEM. Under the assumption that both interior and exterior orientation are known, inserting the ground coordinates *(X, Y, Z)* into the inverse collinearity equations (Eq. 3.3 and Eq. 3.4) will lead to the point *P'* in the image with the corresponding image coordinates (x', y'). The grey value for the pixel of interest in the ortho-image can now be interpolated in the original unrectified image. This process is known as image resampling. The entire geometric modelling process is conducted in physical coordinates. The interpolation of the grey value within the original image in sensor geometry is a simple bilinear interpolation technique in this case (explained in detail in section 5.3). This rather simple interpolation was chosen in order to decrease computation time for the testing of the entire algorithm. Much more sophisticated interpolation methods exist, leading to improved interpolation results. However, most of them have the disadvantage of being computationally expensive. As soon

as the entire registration process will be fully developed and tested, they will be introduced (e.g. B-Spline interpolation).

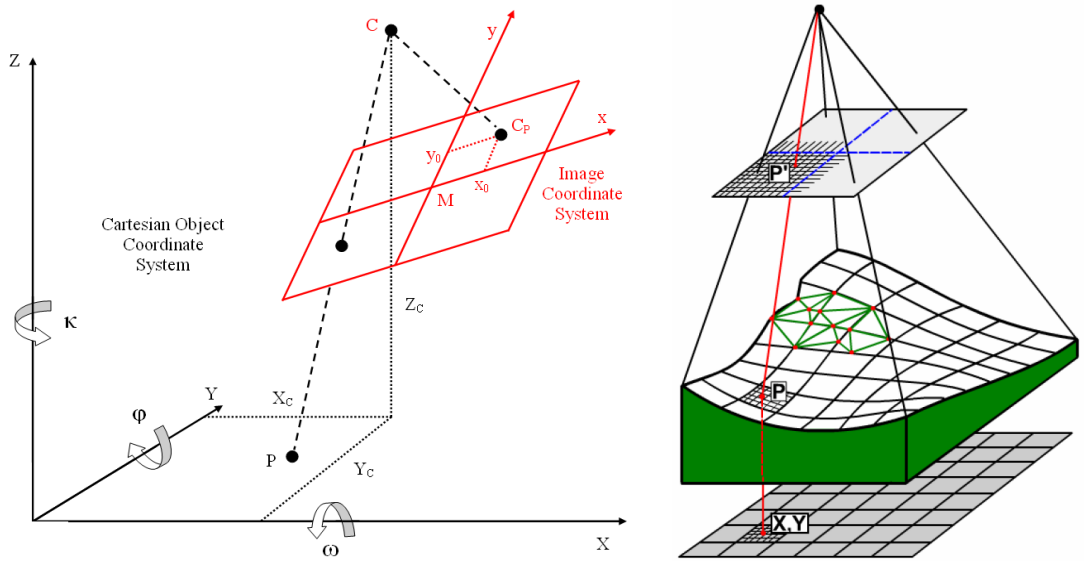


Figure 3.6: (a) Geometric model of the collinearity equations with $C\,(X_C,\ Y_C,\ Z_C)$: perspective center, $P\,(X,\ Y,\ Z)$: object point, $C_P\,(x_0,\ y_0)$: principle point, M: center of the CCD array, (b) Ortho-rectification of the image with a digital elevation model (DEM) [Heipke, 2007]

Residuals rest although the main distortions due to the terrain are reduced significantly by the ortho-rectification of the optical image.

- Displacement effects rest because buildings are not included in the DEM which represents the height of the terrain. These effect could be treated by the introduction of a DSM, provided from an external source, for example LIDAR. However, LIDAR data are not widely available. Further research conducted at CNES and IPI will show if an internal DSM, derived with InSAR techniques from the same SAR data set will lead to sufficient results in urban areas.

- Errors in the parameters of the interior and exterior orientation have an effect on the ortho-photo. For example, errors may be introduced by residuals of the ground control point determination in particular if the exterior orientation is derived from a space resection.

- Errors of the DEM propagate through to the final ortho-photo. The final accuracy is almost independent of the sensor if both DEM and ortho-photo are derived from the

same data set. However, in case the DEM and the rectified image are derived from different data sets or even different types of sensors (e.g. LIDAR and optical aerial imagery) residuals rest in the DEM.

- The approximation of the curved surface by a raster leads to residuals in particular in case of high frequency curvature of the terrain.

3.3.2 Geometric SAR Model

The SAR image is projected to the ground with the inverse equations from [Toutin et al., 1992], originally derived from the collinearity equations. This approach models the residual errors still present after the image has been generated from raw data. For example, residuals rest in the estimation of slant range, Doppler frequency, ephemeris and the ellipsoid. It incorporates three different models: the motion model, the sensor model and the earth model. Hence, three coordinate systems are used: the image coordinate system, the intermediate coordinate system and the ground cartographic coordinate system. The first step is a transformation of the ground coordinates to the intermediate coordinate system. It simply applies one translation and one rotation. Furthermore, the coordinates of the intermediate system (x, y, h) are transformed to the image coordinates (p, q) with the equations shown below. Image coordinate p corresponds to the azimuth while q corresponds to the distance.

$$p = \frac{-y * (1 + \delta\gamma * X) + \tau * h}{P} \tag{3.5}$$

$$q = \frac{-X - \frac{\theta * h}{cos\chi}}{\alpha * \left(Q + \theta * X - \frac{h}{cos\chi}\right)} \tag{3.6}$$

$$X = (x - a * y) * \left(1 + \frac{h}{N_0}\right) + b * y^2 + c * x * y + \delta h * h \tag{3.7}$$

- with N_0, the normal distance between the sensor and the ellipsoid,
- a, a function of the non-perpendicularity of axes,
- α, the field-of-view of an image pixel,
- P, a scaling factor in along-track direction,
- Q, a scaling factor in across-track direction,
- τ and θ, functions of the leveling angles in along-track and across-track direction,
- b, c, χ, $\delta\gamma$, and δh, second order parameters.

This approach was found appropriate because it implies several desirable properties. It models the complete viewing geometry of the sensor and works with both ground control points and a DEM. In contrast to polynomial rectification methods, all variables and factors are directly related to physical quantities. As a matter of fact, parameters, which translate directly to physical properties of the sensor, make the equations somehow easier understandable and interpretable. Appropriate parameter values may be found quite easily.

3.4 Image Preprocessing

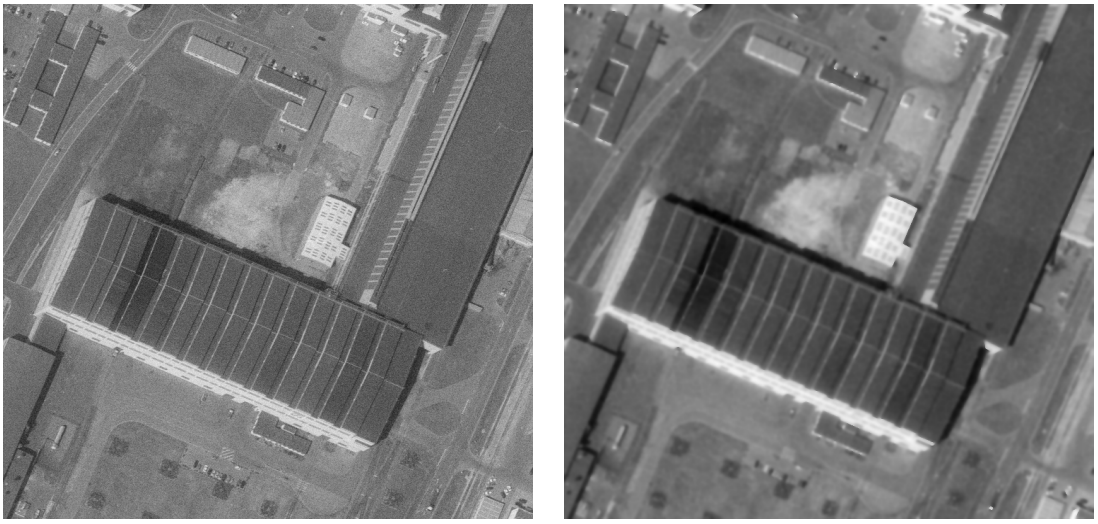
In order to reduce the noise level and the speckle, SAR and optical images first have to be processed with edge preserving smoothing filters. This step facilitates the extraction of lines, contours and regions. Reconsider, that lines will be the input to the following image registration because the very high image resolution prohibits pixel based fusion techniques. Preparing the images for further processing is crucial, since high quality image smoothing is necessary for feature extraction algorithms. It depends on three major constraints.

- Since both geometric and radiometric properties of SAR and optical sensors are different, applying the same smoothing filter to both SAR and optical images results in insufficient outcomes. Images captured by different sensors hence call for adapted smoothing filters.
- The desired image analysis strategy has an impact on the choice of the preprocessing algorithm. For example, edge preserving smoothing filters have to be used if line detection algorithms are applied to the image in further processing steps. In case of the application of mutual information techniques to the image, edge preservation may not be necessary. Therefore, faster to compute filtering techniques (e.g. mean, median) can be applied.
- Preprocessing has to be tailored to fit the type of scene displayed in the imagery, too. Dense urban areas with lots of metallic features result in a higher dynamic range due to more frequent dominant scattering than agricultural scenes. Hence, the filtering parameters have to be adjusted carefully for each image.

3.4.1 Preprocessing of Optical Imagery

For the optical image, an anisotropic diffusion filter, already existing in OTB, was chosen. This filter implements an N-dimensional version of the anisotropic diffusion equation for scalar-valued images proposed by [Perona and Malik, 1990]. The basic idea is derived from nature. Natural surfaces are composed hierarchically of a small discrete number of scale

levels. For example, a coarse scale would be an entire forest. The next step in scale-space would be a particular tree, the following steps a branch, a leaf, the substructure of the leaf etc.. The translation of this scale-space technique to our case in imagery is: the original image $I_0(x, y)$ is embedded in a set of derived images *I(x,y,t)*. This set has been derived by convolving the original image with a Gaussian kernel *G(x,y;t)*. t is the chosen variance of the Gaussian kernel. The greater the variance becomes, the smoother the image and thus the coarser the resolution.



(a) (b)

Figure 3.7: (a) Test region of the original optical image, (b) The optical image after preprocessing with the anisotropic diffusion filter

$$I(x, y, t) = I_0(x, y) * G(x, y; t) \tag{3.8}$$

Three criteria must be met [Perona and Malik, 1990]:

1. Any feature at a coarse level of resolution is required to have a cause at a finer level of resolution.

2. The region boundaries should be sharp at any level of resolution. Additionally, they should always coincide with semantically meaningful boundaries.

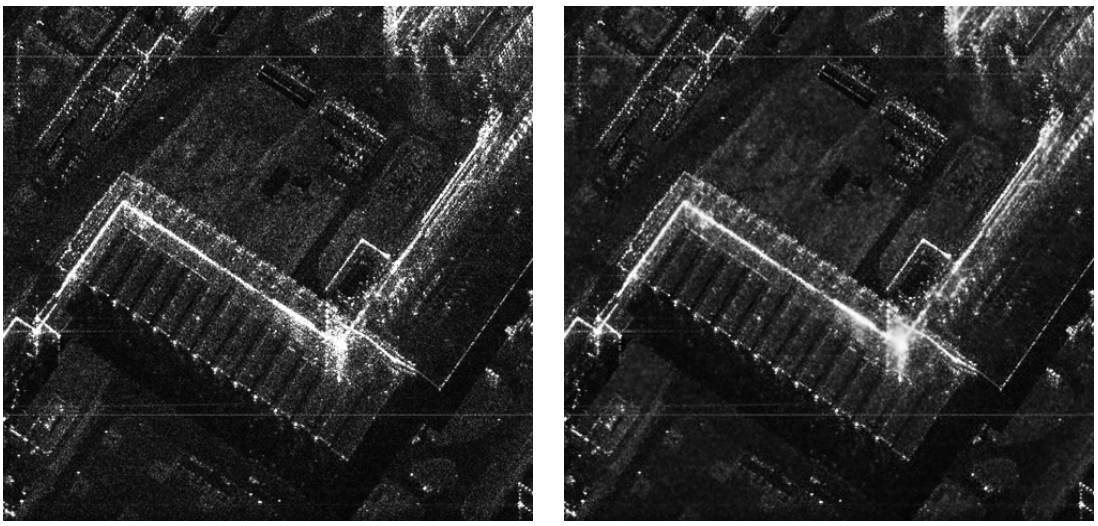
3. Intraregion smoothing should always occur previously to interregion smoothing at all scales. Regarding the tree example previously introduced, branches merge to a tree and trees merge to a forest. No branches merge directly to make up the forest.

The main issue is that Gaussian smoothing does not consider region boundaries. Hence, region boundaries have to be estimated. A variable conductance coefficient is introduced to set smoothing at region boundaries to zero and inside regions to one. This conductance term was chosen as a function of the gradient magnitude of the optical image i.e. region boundaries are estimated from the gradient image. Fig. 3.7(b) shows the optical image as shown in Fig. 3.7(a) after smoothing has been conducted with the anisotropic diffusion filter.

3.4.2 Preprocessing of SAR Imagery

In SAR imagery, both noise reduction and the treatment of the speckle effect takes place. The speckle effect causes severe perturbations within SAR images (refer to section 2.3.3 for the theory of speckle). First and foremost, it prevents the direct application of optical image analysis algorithms to SAR images. Thus, specific algorithms, which take into account the physical nature of SAR images, have to be developed. In order to prepare for image analysis, e.g. feature extraction, the speckle effect is reduced as far as possible. Although our line extraction algorithm (see details in section 4.4) was specifically developed for SAR images, it seems to provide improved results on despeckled images. For good line detection results, the preservation of edges by the anti-speckling filter is imperative. Edges separate image regions and therefore may also be thought of as region boundaries. Hence, a filter has to be chosen that does not smooth the image globally across such region boundaries. An approach has to be found which restricts smoothing to rather homogenous regiones. Local statistics (e.g. mean, standard deviation) and texture parameters (e.g. contrast, entropy) can be calculated within the filter matrix for deciding whether a pixel belongs to a region or not. A filter that takes into account local properties of an image in order to fine tune its parameters is called adaptive. It is desireable to apply such an adaptive filter to the image for edge preservation. Hence, edge preserving speckle reduction was conducted with the Frost filter [Frost et al., 1982]. The input image from Fig. 3.8(a), after the application of the Frost filter, is displayed in Fig. 3.8(b).

This filter is adaptive because it considers the local mean and the local standard deviation of the input image. In other words, the Frost filter smoothes inside regions and not across region boundaries (like the anisotropic diffusion filter). It applies an exponential weight factor, which depends on the local statistics, in order to adjust the smoothing. This is a difference to the Lee filter [Lee, 1981], which does not use such a weight factor. Therefore, the Frost filter preserves edges better than the Lee filter. The Frost filter considers the desired information, the terrain backscatter $r(x,y)$ (the ideal image we want to estimate), to be multiplied (i.e. to be perturbed) with a stationary random process $n(x,y)$ (the speckle effect). Equation 3.9 displays the relation previously described. Additionaly, it integrates a



(a) (b)

Figure 3.8: (a) Test region of the original SAR image, (b) The SAR image test region after the entire original SAR image was Frost filtered

third function *h(x,y)* (the operator * describes convolution). This third function expresses the spatial correlation of pixel values introduced by the SAR system components such as the antenna and the receiver. *I(x,y)* is the observed image.

$$I(x,y) = [r(x,y) \cdot n(x,y)] * h(x,y) \tag{3.9}$$

Modelling a SAR image like done by Frost in equation 3.9 is referred to as a multiplicative speckle model. According to this particular multiplicative speckle model, the Frost filter minimizes the mean-square error in order to estimate the ideal image *r(x,y)* from the observed image *I(x,y)*. The speckle filter always should be applied to the entire original image. Filtering only an image region would lead to a slightly different outcome. Since the extracted region shows another mean and standard deviation, the filter would produce another result, too. Hence, the SAR image of the original size (Fig. 3.1(b)) was Frost filtered. Thereafter, the test image region (Fig. 3.8) was extracted.

4 Classification and Feature Extraction

The first step of the image registration strategy was to project the optical and the SAR image to the ground. Both images were ortho-rectified in order to account for the different sensor geometries. Distortions due to the terrain are also decreased by this step. Thereafter, the noise level of the optical image was reduced and the speckle effect of the SAR image was treated. A classification of both images is the following step described in the first two sections of this chapter (sections 4.1 and 4.2). The main reason for the introduction of a classification is that the image fusion has to take place in rectified regions of the images. All objects both present in the images and the DEM have been rectified. However, objects present in the image but not in the DEM have not been accounted for. In particular, buildings are not contained within a DEM, i.e. buildings in the images stay distorted. Hence, we have to distinguish between buildings and the ground in order to register the images only on the rectified ground level. Registering the images in building regions would immediately lead to severe perturbations. Additionally, a classification will prove to be useful for the following feature extraction and the final registration. For example, very bright lines in the SAR images are one additional class besides the two main classes roof and ground. Such bright lines often result from double bounce effects of the radar signal. It usually occurs where the walls of buildings meet the ground. Thus, we know that these bright lines are on ground level, i.e. they can be used for registration purposes. A large variety of image classification techniques exists and new approaches are permanently developed. Within this project, we tested a classification subdivided into two steps. The first classification uses Support Vector Machines (SVM) for a pixel based classification. Results are refined with a Markov Random Field classification which is based on global and local image statistics.

After the classification has been accomplished, features are extracted from the images. Features to be extracted can be regions, lines, or points. Later on, the images will be registered on feature level. This abstraction from the original images illiminates all radiometrical differences between the optical and the SAR images. It is necessary because we deal with images of very high resolution. A classical fusion based on pixel level would fail in our case. Both, the extraction of point and line features were tested. The point detection within the SAR images was conducted with the Lopès point detector [Lopès et al., 1993]. However, line

detection proved to give more promising results and hence this approach was followed. The line extraction algorithm applied to the optical image is explained in section 4.3 whereas the one applied to the SAR image is described in section 4.4.

4.1 Classification with Support Vector Machines (SVM)

Support vector machines (SVM) belong to the family of kernel based learning methods. They deploy learning theory for classification and regression tasks and are a generic tool [Cortes and Vapnik, 1995]. For example, a first application for SVM was text categorization, a subject with only a slight relation to image classification. However, within the last few years, SVM proved to successfully classify remote sensing imagery. This section will give a brief introduction to SVM classification since classification is only one out of several parts of the registration strategy. Sources [Schölkopf and Smola, 2002] and [Shawe-Taylor and Cristianini, 2004] are recommanded in order to gain further in depth understanding of kernel methods in general and SVM in particular.

- Our fundamental problem to be solved within the context of SVM classification is: *Two classes of objects are given and we have to assign a new object to one of the two classes.*
- The basic idea to deal with this issue is rather simple and can be imagined in a geometric way: *The particular surface (called hyperplane) in feature space has to be determined that optimally separates two feature sets derived from the objects (Fig. 4.1).*

In the case of imagery the objects are samples taken from the image. The new object is assigned a class depending on which side of the hyperplane it lies. In order to achieve optimal separation, the SVM algorithm searches for the subset of training samples which best describes the optimal surface. The distance of the closest vectors to the hyperplane is maximized. These are the so-called support vectors and the minimal distance is called margin [Inglada et al., 2006].

The input for the SVM classification are N samples which are taken from training regions. These training regions have to be specified by the user within the images we want to classifiy (see Fig. 4.3(a) and Fig. 4.3(c)). Since the entire classification is based on feature vectors from the training regions, they have to sufficiently represent the classes we want to distinguish. The basic version of SVM solves two-class problems and thus the following considerations are based on a classification into two classes ω_1 and ω_2. Each single sample taken from a training region consists of the class label y_i with $i = 1, 2, ..., N$ and the corresponding feature vector $\vec{x}_i$. The class label y_i is either -1 if $\vec{x}_i$ belongs to ω_1 or $+1$ if $\vec{x}_i$ belongs to ω_2. The feature vector $\vec{x}_i$ consists of real numbers and has the dimension n. The dimension n depends on

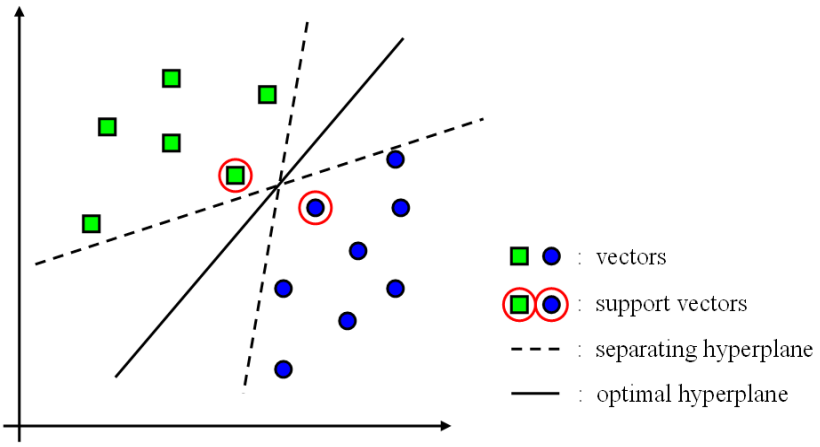


Figure 4.1: Basic principle of a SVM classification of two vector sets in feature space (drawn after [Tourneret, 2003])

the amount of features the vector incorporates. For our project we used seven statistical and texture features ($n = 7$), i.e. each image pixel contains not only one grey-value but a vector of seven features. The hyperplane has the equation

$$\vec{w} \cdot \vec{x} + b = 0 \tag{4.1}$$

with its normal vector $\vec{w}$ and $\vec{x}$ being any point on the hyperplane in feature space. All feature vectors that are not located directly in the hyperplane do not fulfill equation 4.1 because the left side of the equation is either smaller or greater than zero. Hence, the classifier function can be written as in equation 4.2.

$$f\left(\vec{x}, \vec{w}, b\right) = sgn\left(\vec{w} \cdot \vec{x} + b\right) \tag{4.2}$$

Two new hyperplanes are constructed which are parallel to the optimal separating hyperplane. The normal vector $\vec{w}$ may be interpreted as a weight vector and b can be regarded as a threshold. They are rescaled such that the left side of equation 4.1 becomes either $+1$ or -1 (Fig. 4.2). Equation 4.3 expresses this constraint for feature sets which are linearly separable. In case the feature sets are not linearly separable, the constraints (Eq. 4.3) can be modified thus generating a soft margin classifier.

$$\begin{aligned} \vec{w} \cdot \vec{x}_i + b \geq +1 \quad if \quad y_i = +1 \\ \vec{w} \cdot \vec{x}_i + b \leq -1 \quad if \quad y_i = -1 \end{aligned} \tag{4.3}$$

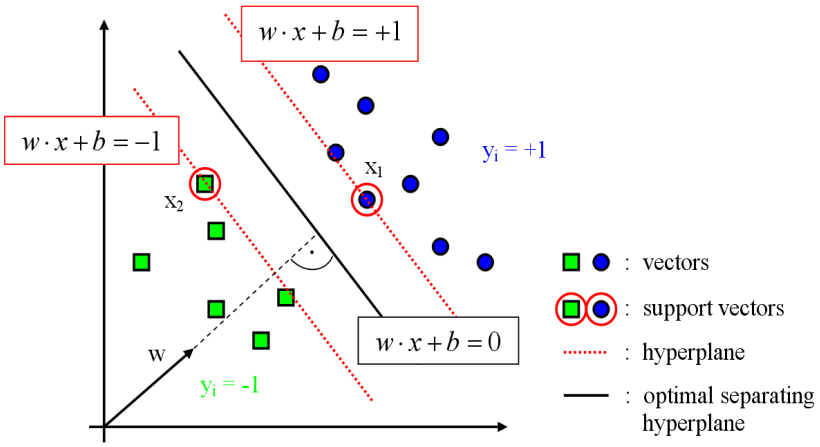


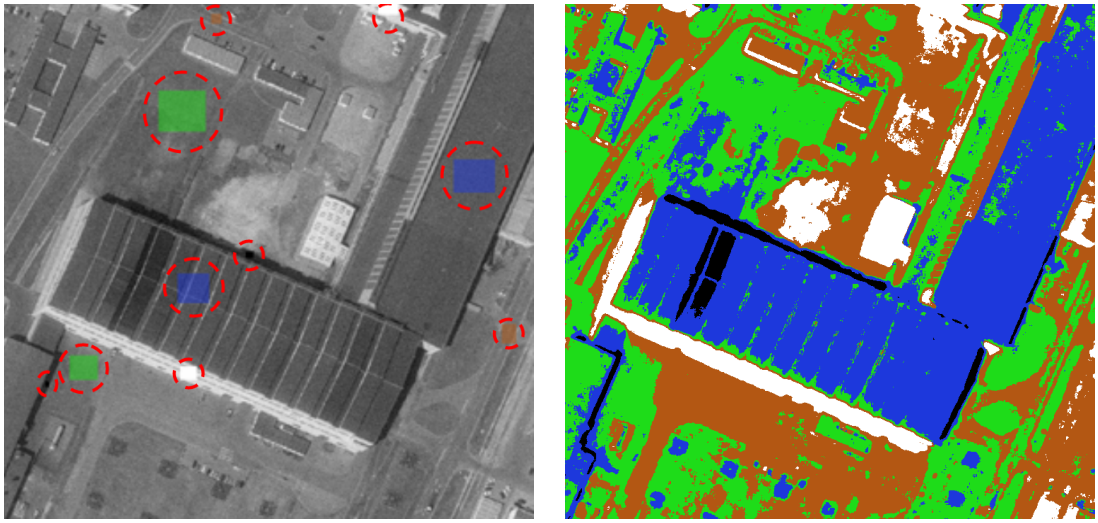
Figure 4.2: A binary SVM classification of two vector sets in feature space (drawn after [Schölkopf and Smola, 2002]); the orthogonal distance between the optimal hyperplane and the two red dotted lines is $|1-b| \,/\, \|\vec{w}\|$ and $|-1-b| \,/\, \|\vec{w}\|$ respectively; the margin which has to be maximized for optimal separation thus becomes $2/\,\|\vec{w}\|$

The goal of SVM is to optimally separate the two feature sets. In order to achieve this aim the margin of $\frac{2}{\|\vec{w}\|}$ has to be maximized. Therefore, $\vec{w}$ and b have to be rescaled in order to minimize the expression $\frac{1}{2}\|\vec{w}\|^2$. This minimization task has to fulfill the constraint $y_i\,(\vec{w}\cdot\vec{x}_i+b) \geq 1, \quad i = 1,\ 2, ...,\ N$. A so-called constrained optimization problem can be solved by introducing Lagrange multipliers $\alpha_i \geq 0$ and the set-up of a Lagrangian L (Eq. 4.4). The Lagrangian L has to be minimized with respect to the primal variables $\vec{w}$ and b and maximized with respect to the dual variables α_i. In other words, a saddle point of the Lagrangian equation has to be found [Schölkopf and Smola, 2002]. It can mathematically be proven that only the support vectors have positive Lagrangian multipliers α_i ([Inglada et al., 2006], p.225).

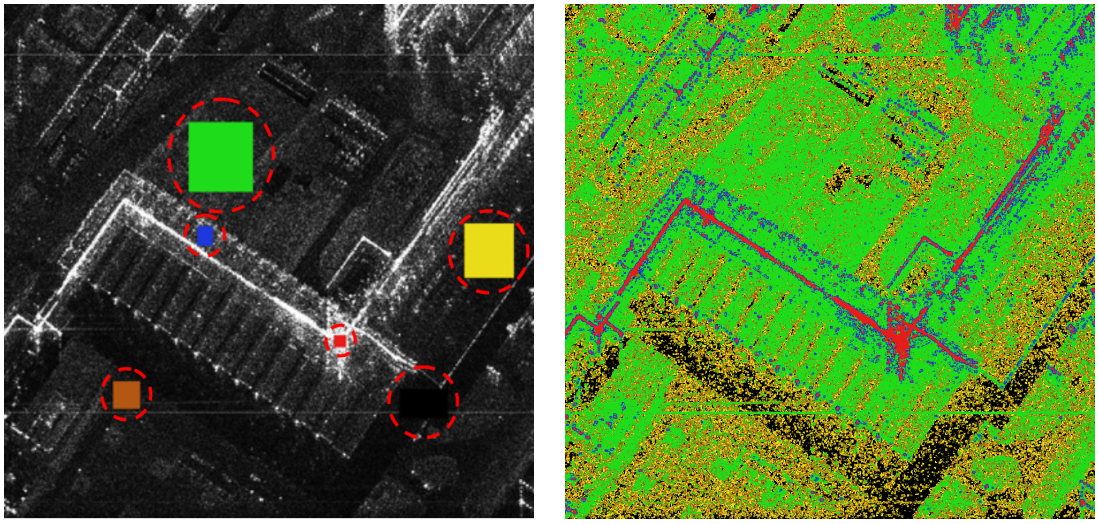
$$L\,(\vec{w},\ b,\ \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{N}\alpha_i\,(y_i\,(\vec{w}\cdot\vec{x}_i+b)-1) \tag{4.4}$$

While SVM proves to be a powerful tool for classification tasks some drawbacks exist. The SVM classification was originally developed to solve two class problems. However, in image classification we usually have more than two classes. Two main theoretical approaches exist to deal with this issue. The first possibility is to train each single feature set against any other feature set. Another possibility is to train each feature set against the rest. For this project the first approach was selected. Another drawback is that the classification is not entirely

automated because the user has to provide training regions. The quality of the classification highly depends on the selection of those training regions (Fig. 4.3(a) and Fig. 4.3(c)). Additionally, the features we use as input have a influence on the result. In our case, an image stack with several layers is the input for the SVM classification. Each layer is derived from the original rectified image. Diverse texture parameters as well as stochastic values were determined in order to achieve a meaningfull classification: mean, median, entropy, energy, standard deviation, skewness and kurtosis. Refer to Annex B for the corresponding equations. The classified optical and the classified SAR image are shown in figures 4.3(b) and 4.3(d) respectively.



(a) (b)



(c) (d)

Figure 4.3: (a) Training regions for the SVM algorithm in the optical image, (b) Classification into five classes (ground, vegetation, roof, shadow, facade) of the optical image, (c) Training regions for the SVM algorithm in the SAR image, (d) Classification into six classes (ground, vegetation, dark roof, light roof, shadow, bright lines) of the SAR image

4.2 Classification with Markov Random Fields (MRF)

After the images have been classified with the SVM method, a second classification is necessary in order to refine the SVM results. The SVM technique is a pixel based classification approach. Vectors for each pixel are first determined and then optimally separated. Classification with Markov Random Fields (MRF) takes global and local statistics of the input image into consideration. The image is regarded as the result of a random process with a corresponding probability density function. This chapter will give a short introduction to MRF. It is based on [Sigelle and Tupin, 1999] and further details may be found there.

- The definition of a Markov field applied to imagery is: *x is a Markov field if and only if the local conditional probability is exclusively a function of the neighborhood within the considered region.* In other words, the grey value of a pixel solely depends on the grey values of its neighboring pixels.

In the context of MRF a digital image is seen as a bidimensional (or n-dimensional) quantified variable. It can be subdivided into zones, contours and structures due to parameters like contrast, texture etc.. Hence, a single pixel grey value may not be significant itself but the relation and interaction with neighboring pixels can lead to significance. The MRF approach uses local grey value differences within a specified pixel neighborhood in order to distinguish between regions. Any pixel s is part of a discrete finite network S (the entire image) and it always has a certain property, usually its grey value. Cliques C_i of pixels are derived from the local neighborhood V_j. Index i stands for the number of pixels within the clique whereas j is the number of pixels within the entire neighborhood (Fig. 4.4).

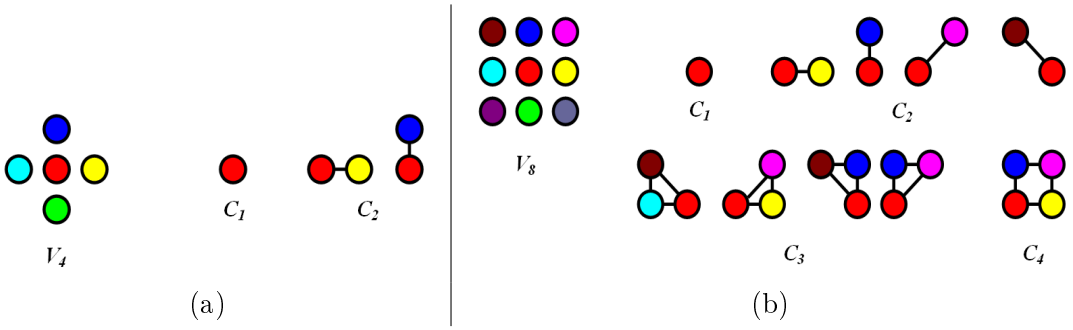


Figure 4.4: (a) Cliques C_i within a 4-connectivity neighborhood V_4, (b) cliques C_i within a 8-connectivity neighborhood V_8 (figures drawn after [Sigelle and Tupin, 1999], p.10)

The local interaction between the grey values of a single clique c_i within a neighborhood is called the potential of the clique U_{c_i}. The sum of all potentials U_{c_i} of all cliques c_i of an image is called the global energy of the image U_g. The local energy U_l is the sum of the potentials of cliques U_{c_i} within a certain image region. In order to apply MRF, statistics of the image have to be defined. Hence, any image is considered a realization x of a random field. The global image probability $P(x)$ is used to determine the relation between a local region and the rest of the image. In order to compute local conditional probabilities, it is necessary to introduce Gibbs fields. The Gibbs measure is an energy function. The global energy U_g of a Gibbs field can be decomposed into the local energies U_l of the cliques (Eq. 4.5).

$$\begin{aligned} P(x) &= \frac{1}{Z} exp\left(-U_g(x)\right) = \frac{1}{Z} exp\left(-\sum_{c \in C} U_{c_i}(x)\right) \\ Z &= \sum_{x \in \Omega} exp\left(-U_g(x)\right) \end{aligned} \tag{4.5}$$

Z is a normalization term, defined over all possible energy relations Ω. It can be mathematically proven that Markov fields and Gibbs fields are equivalent (Hamilton theorem). In order to summarize the so far developed steps we can say that the energy function U is based on the potentials of cliques C_i of pixels, defined inside specific neighborhoods V_i. The potentials of the cliques allow for the evaluation of the global probability as well as the local conditional probability. An issue that rests to be dealt with is the determination of the global Gibbs probability of a configuration. For example, for a binary image of size 512 x 512 the number of possible configurations is 2^{262144}. This huge number of possibilities makes the direct determination of the global probability computationally very expensive. Hence, the idea is to take samples of the image and to calculate the global Gibbs probability based only on such samples. The goal is to find a sampling algorithm that satisfies the Gibbs probability (Eq. 4.5). Two classical approaches are usually deployed for the extraction of these image samples: the Gibbs samples and the Metropolis algorithm. The latest version of the programmed OTB classes uses the Gibbs sampler. It is based on an iterative construction of a sequence of images. After a sufficient number of iterations the image sequence converges and will satisfy the global Gibbs probability. The MRF model used in this project, defined with a neighborhood and a particular energy function, is a Gaussian Markovian model. While assigning pixels to certain classes (within a Bayesian framework), this model prefers small grey value differences for neighboring pixels.

Two images are the input to the MRF classification carried out in this project. The first image is the optical image after rectification respectively the SAR image after rectification. It is important to use the rectified images without any smoothing in order to leave the image statistics unchanged. The MRF classification considers image statistics and thus changed image statistics will lead to errors. The second input image is the labelled image from the SVM classification. It was tested for initialization purposes of the MRF algorithm. However, a maximum likelihood classification as implemented in the MRF algorithm is usually suffiecient. The initialization with the SVM classification was introduced only for testing reasons. A MRF classification is characterized by certain functions that define its properties. Such functions to be specified are: a likelihood term, a regularization term, the clique within a neighborhood, the optimization algorithm and the regularization coefficient. Likelihood term, regularization term and the regularization coefficient β define the maximum a posteriori energy $U(x/y)$ as implemented in the used software (see Eq. 4.6). The first two summands describe the likelihood term (a Gaussian distribution with mean and standard deviation as parameters) whereas the last summand is the regularization term. β acts as a weight factor for the regularization term thus defining the relative impact of likelihood term and regularization term on the maximum a posteriori energy.

$$U(x/y) = \sum_{s} \frac{(y_s - \mu_{x_s})^2}{2 \cdot \sigma_{x_s}^2} + \log \sqrt{2 \cdot \pi} \cdot \sigma_{x_s} + \beta \cdot \sum_{(s,t) \in C_2} \Phi(x_s, x_t) \tag{4.6}$$

The regularization function Φ is necessary for modelling the potential $U_{c=(s,t)}$ of the cliques. Our programm applies the Potts model [Wu, 1982] as shown in Eq. 4.7. This model is capable of dealing with several grey values and class labels. The chosen clique is of second order (C_2) and defined on an 8-connectivity neighborhood (see Fig. 4.4(b)).

$$\begin{aligned} U_{c=(s,t)}(x_s, x_t) &= -\alpha \quad if \quad x_s = x_t \\ U_{c=(s,t)}(x_s, x_t) &= +\alpha \quad if \quad x_s \neq x_t \end{aligned} \tag{4.7}$$

As optimizer, the ICM algorithm developed in [Besag, 1986] was deployed for fast convergence. As the initialization is pretty good, we can be relatively certain that the global minima is reached. Since a Gaussian model is used, the grey-value means and the corresponding standard deviations of the desired classes have to be specified, too. They were taken manually from the rectified images.

4.3 Feature Extraction in Optical Imagery

So far, the optical image and the SAR image are ortho-rectified, smoothed and classified. The next step consists of extracting features. Feature extraction in the optical image is necessary in order to register the results with the feature image of the SAR image. The best results for the optical image were achieved with the Canny edge detection algorithm [Canny, 1986]. Five major steps are used in the edge detection scheme:

1. The input image is smoothed with a Gaussian filter.
2. The second directional derivatives (Hesse matrix) of the smoothed image are computed.
3. Non-maximum suppression is applied: the zero-crossings of the second derivative are found and the sign of the third derivative is used to determine the appropriate extrema.
4. The zero-crossings are multiplied with the gradients of the image.
5. Hysteresis thresholding is applied to the gradient magnitude of the smoothed image in order to find and link edges.

The result of the Canny-Operator applied to our smoothed optical test image is shown in 4.5. It highly depends on the previously conducted preprocessing. More smoothing results in less detected edges. Thus, the parameter choice of the anisotropic diffusion filter has a high influence on the outcome of the line detection.

Figure 4.5: Extracted lines from the optical image using the Canny line detector

4.4 Feature Extraction in SAR Imagery

Three different categories of features visible in SAR imagery may be useful for the fusion of SAR and the optical imagery: linear features, regions and point targets. A point detection filter proposed by [Lopès et al., 1993] was implemented. Results are promising but still need further refinement. The best extraction results for the SAR image are achieved with linear feature extraction. Visible linear features in SAR images of urban areas are:

- Very strong reflections occur where building walls meet the ground. This is due to double bounce effects of the signal.
- The regularly structured surface of factory hall roofs, if oriented perpendicularly to the sensor's flight direction, is visible in the image.
- Power lines and any conductive ground feature appear bright in the SAR image.
- Very often, the contrast between roofs and soil is clearly visible.

An assymetric fusion of lines approach was chosen for the line extraction in SAR imagery. It was originally developed for the detection of road networks [Tupin et al., 1998]. The result of this algorithm, applied to the SAR image (after thresholding), may be seen in Fig. 4.6. Its strategy is the fusion of the outcome of two separate line detectors D1 and D2. Both line detectors consist of two edge detectors. D1 is based on the ratio edge detector proposed in [Touzi et al., 1988]. D2 uses the normalized centered correlation of two pixel regions. The following two paragraphs will explain these two line detectors in further detail.

Line detector D1 consists of two ratio edge detectors, one on each side of the region of interest. The ratio edge detector is a filter with a fixed false alarm rate. It is appropriate for SAR images because the speckle effect is considered as multiplicative (in contrast to optical images where noise is considered additive). D1 calculates the ratio of local means μ of the regions i, j on both sides of an edge (Fig. 4.7(a)). The edge detectors response $r_{i,j}$ is defined as:

$$r_{ij} = 1 - min\left(\frac{\mu_i}{\mu_j}, \frac{\mu_j}{\mu_i}\right) \tag{4.8}$$

The line detector D1 minimizes the responses of two edge detectors. For line detection, we consider three regions: 1, 2 and 3 (Fig. 4.7(b)). Region 2 is the probable line. 1 and 3 are the neighboring regions. Hence, the response r to the line detector D1 is:

$$r = min\left(r_{12}, r_{23}\right) \tag{4.9}$$

Diverse widths of region 2 are tried since the width of linear features may vary. Additionally, eight directions are tested for each pixel. Only the best response is kept. A pixel is considered

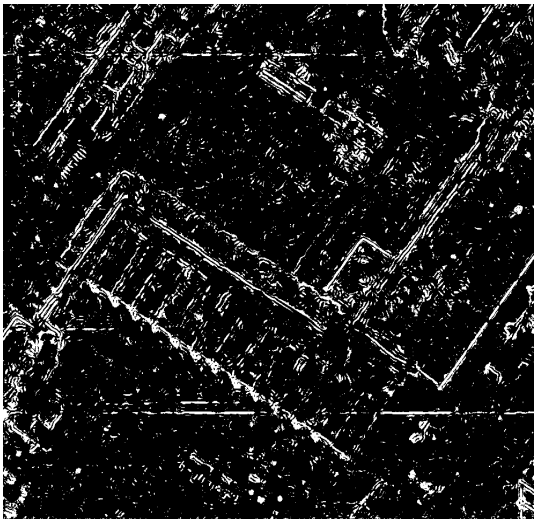


Figure 4.6: Extracted lines from the SAR image

a line pixel, as soon as the response r exceeds a previously chosen threshold. Lowering the threshold results in more detected lines but also in a higher false-alarm rate. Therefore, this decision threshold is a compromise between the chosen false-alarm rate and the minimum detectable contrast. However, the false-alarm rate may also be decreased by increasing the regions size. The more pixels contribute to the empirical mean of the region, the less false-alarms occur. It has to be considered that larger regions increase computation time.

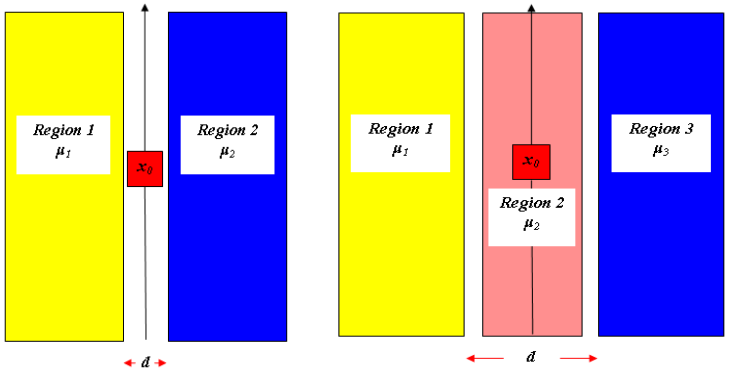


Figure 4.7: (a) Vertical edge model, (b) Vertical line model; μ_i is the empirical mean of region i (figures drawn after [Tupin et al., 1998], p.436)

The second line detector D2 is also based on two edge detectors. D2's edge detector is based on the normalized-centered cross correlation coefficient ρ_{ij}^2. An edge consists of two regions i and j with their corresponding means μ, the pixel number inside the region n and the ratio of standard deviation and mean γ. In the following equation the ratio of the regions' means $\bar{c}_{ij}$ is interpreted as the empirical contrast between the two regions i and j. The closer to one this variation coefficient is, the more homogenuous is the area.

$$\rho_{ij}^2 = \frac{1}{1 + (n_i + n_j) * \frac{n_i * \gamma_i^2 * \bar{c}_{ij}^2 + n_j * \gamma_j^2}{n_i * n_j * (\bar{c}_{ij} - 1)^2}} \tag{4.10}$$

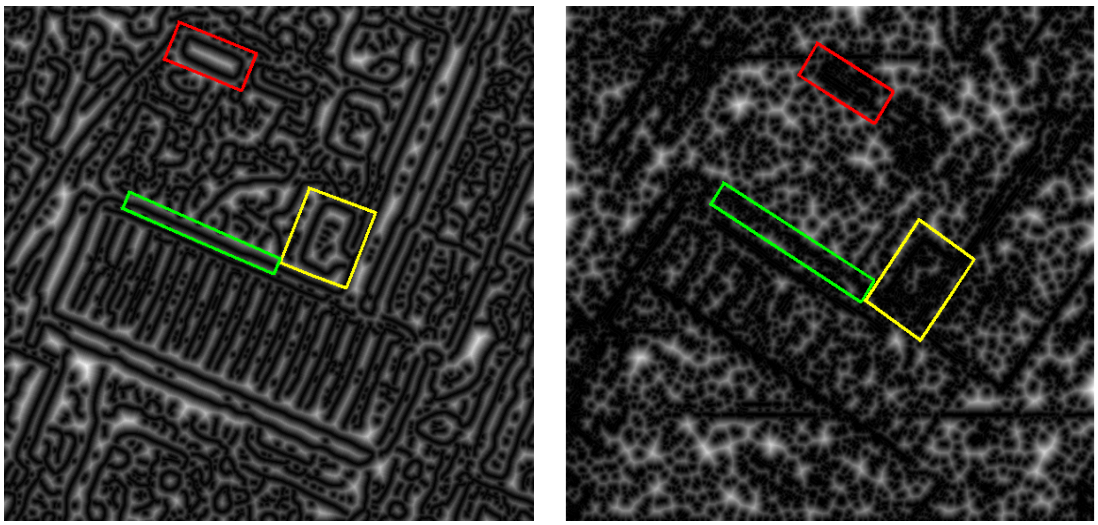
The advantage of this edge detector is its dependence on both the contrast between the two regions $\bar{c}_{ij}$ and inside each region γ. A disadvantage is that may be influenced by outliers contained in the regions. Again, the line detector strives for the minimum response $\rho = min\,(\rho_{12}, \rho_{23})$ of the edge detectors neighboring the potential line. Finally, the responses from both line detectors D1 and D2 are fused. A so-called associative symmetrical sum $\sigma\,(x, y)$ is computed (Eq. 4.11). x and y are the responses from the line detectors.

$$\sigma\,(x, y) = \frac{x \cdot y}{1 - x - y + 2 \cdot x \cdot y}\ , \quad with \quad x, y \in [0, 1] \tag{4.11}$$

4.5 Calculation of Distance Images

Distance images are calculated from the feature images derived in section 4.3 and section 4.4. This step was introduced in order to reduce remaining geometric residuals that result in different absolute positions of the extracted lines.

The differences between corresponding lines in the images are assumed to be more similar than the absolute line positions. Hence, distance images will increase the similarity between the optical image and the SAR image. This process is also referred to as distance mapping. Fig. 4.8(a) and Fig. 4.8(b) show the distance maps of the feature images. The colored boxes frame corresponding parts of the images. Bright values display longer distances while darker values display pixels close to a line pixel. A distance transformation is the algorithm producing a distance map from a binary image that contains objects and background. In our case, the objects are the previously extracted lines. Such a distance map displays the Euclidean distance between a background pixel and the nearest line pixel. This distance is translated to a grey value. An approach developed by [Danielsson, 1980] was chosen for the computation of the Euclidean distances. This approach writes into each pixel the vector of the relative position of the nearest line pixel instead of solely noting the distance. It leads to a representation of the Voronoi division of the object pixels.



(a) (b)

Figure 4.8: (a) Distance map of the optical image, (b) distance map of the SAR image

5 Image Fusion

So far, the optical and the SAR image have been ortho-rectified, preprocessed, classified, lines have been extracted and distance images have been computed. All such steps can be considered as preparations for the registration of the images following up. Whenever possible, geometric and radiometric differences between the images, due to the different sensors, have been reduced (reconsider the ortho-rectification and the feature extraction, respectively). Up to this stage, both images thus have always been treated separately. This chapter will now introduce the reader to the registration of the images. For the first time within the processing chain, both images are treated simultaneously and a relation between them is established. This relation consists of an image comparison which is conducted with a similarity measure. We measure, to which extent we can find corresponding information in both the optical and the SAR image. Since we have reduced the information contained within the images to line features, corresponding lines are considered corresponding information. The first section briefly introduces open source software library ITK and outlines the architecture of its registration framework. It consists of four modules which are described in further detail in the sections following thereafter.

5.1 Registration Framework

The image registration process is embedded into a registration framework, originally provided by ITK (Fig. 5.1). ITK is the core component of the software library OTB. Remember that OTB is the software library, where the fusion approach developed in this project is integrated in. ITK was originally developed for the exploitation of medical images. It provides sophisticated algorithms for image analysis tasks.

The inputs to this framework are two images. While one image is called the fixed image (it acts as the reference), the other one is called the moving image. The goal is to find the optimum spatial mapping parameters that align the moving image with the fixed image. Hence, the moving image has to be deformed. In this project, the fixed image is the optical image while the SAR image is the moving image. We have chosen the optical image as the fixed image because we consider it to contain less residuals than the SAR image. In the following, the optical image will allways be called fixed image and the SAR image will be

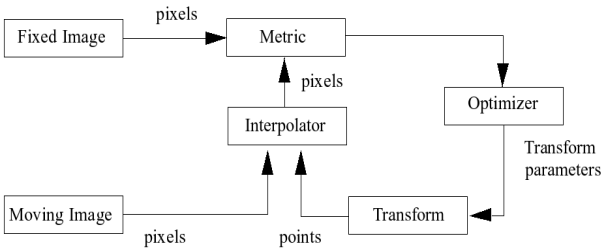


Figure 5.1: The ITK registration framework [Ibánez et al., 2005]

called moving image. The registration framework treats image registration as an iterative optimization problem and consists of four components:

- the transform component applies a geometric transformation to the fixed image points in physical space in order to map them to the moving image,
- the interpolator evaluates intensities in the moving image at non-grid positions,
- the metric measures the similarity between the deformed moving image and the fixed image,
- and the optimizer optimizes the similarity value.

A new set of parameters for the transformation of the following iteration is determined after each optimization step. The major advantage of this modular conception of the registration framework is easy compatibility of a large variety of optimizers, geometric transformations and similarity measure techniques. A detailed description of the chosen algorithms for each registration component is given in the following sections.

5.2 The Transformation

Different two-dimensional and three-dimensional geometric transformations exist e.g. affine transformation, projective transformation or polynomial transformation. We also have to distinguish between rigid and non-rigid transformations. Rigid transformations act globally on the image i.e. the transformation parameters stay the same for each image point. Non-rigid transformations act locally and thus allow for different transformations of the image points. A rather simple two-dimensional rigid transformation was chosen for this project in order to facilitate quick parameter estimation. It consists of a clockwise rotation α around

the geometric image center (C_X, C_Y) and of translations in x and y direction, T_X and T_Y respectively (Eq. 5.1). The rotation is applied first, followed by the translation.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} cos\ \alpha & -sin\ \alpha \\ sin\ \alpha & cos\ \alpha \end{bmatrix} \cdot \begin{bmatrix} X - C_X \\ Y - C_Y \end{bmatrix} + \begin{bmatrix} T_X + C_X \\ T_Y + C_Y \end{bmatrix} \tag{5.1}$$

Two ways of applying the a transformation to the image exist: direct and indirect (Fig. 5.2). While the direct method is commonly used for transforming data, the indirect method is almost always used for transforming imagery. Applying an indirect transformation prevents holes and guarantees that each pixel receives one and only one new grey value.

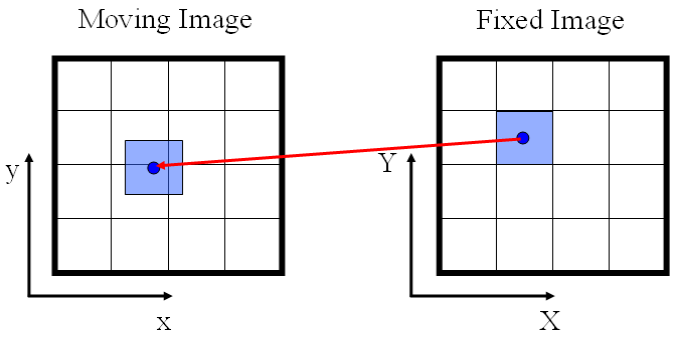


Figure 5.2: Indirect transformation technique

Hence, we use the indirect transformation method. It starts with the result and transforms back to the original image. The result in this case is the deformed moving image with the grid of the fixed image. The idea is

1. to start iterating through the grid of the fixed image (in physical space),
2. to transform each point to the moving image,
3. to interpolate the grey value within the moving image,
4. and to assign this grey value to the current fixed image grid position.

The result of this transformation is an image with the grid of the fixed image and the interpolated grey values of the moving image. This new image is the deformed moving image. Our aim is to find the transformation parameters that optimally deforms the moving image in order to maximize the similarity measure.

5.3 The Interpolator

Grey value interpolation is necessary since the transformation into the moving image leads to positions that usually do not coincide with the exact grid positions. In order to obtain a grey value at a fractional location in the moving image several interpolation techniques may be thought of. In this case, the goal is to implement an interpolation function that is rather simple understand, fast to compute and that results in a continuous grey value surface. Hence, the interpolator used in this project is based on bilinear interpolation. Bilinear interpolation is the extension of a one-dimensional linear interpolation for interpolating functions of two variables on a regular grid. The grey value of the point of interest within the image is calculated from the weighted average of the four surrounding grid points (Fig. 5.3). First, two intermediate grey values g_A and g_B are determined in y direction (Eq. 5.2). In the following the grey value of the point of interest g_P is determined by linearly interpolating between g_A and g_B in x direction. It has to be considered that all computations are done in physical image coordinates.

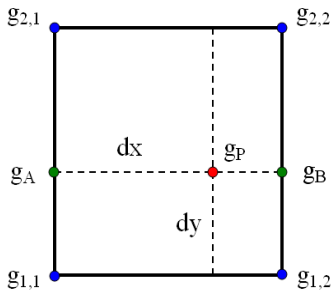


Figure 5.3: Bilinear interpolation of a grey value incoporating four neigboring grey values

$$\begin{aligned} g_A &= g_{1,1} + dy \cdot (g_{2,1} - g_{1,1}) \\ g_B &= g_{1,2} + dy \cdot (g_{2,2} - g_{1,2}) \\ g_P &= g_A + dx \cdot (g_B - g_A) \end{aligned} \tag{5.2}$$

The generated output grey value surface is continuous but not necessarely smooth because only four neighboring grey values are included for the interpolation. Smoother interpolation results can be obtained by using more sophisticated interpolation functions such as B-Spline interpolation. However, the relatively simple approach of bilinear interpolation was found sufficient in this case.

5.4 The Metrics

A metric is a tool for measuring the similarity of two images. It measures how well the transformed moving image fits the fixed image. Therefore, it compares the grey-scale intensity of the images [Ibánez et al., 2005]. Considering two different images as input to the framework, corresponding objects in both images have to be detected. For example, a particular point in the fixed image has to be found in the moving image as well. Only in case we find such corresponding point, a transformation may be determined that maps one point onto the other. More generally we can say: a metric evaluates the amount of overlapping information of both images. It provides a measure of their similarity. Completely identical images, i.e. images taken by the same sensor under the same circumstances, having identicial radiometric properties as well as geometric properties, will thus lead to optimal metric values. Usually, the images are not completely identical. In fact, very often they show only few similarities at first sight. This lack of similarity is a serious issue, particularly in the case of optical and SAR imagery comparison. In this project, the performance of three different metrics was tested, based on least-squares adjustment, normalized correlation and mutual information.

The mean-squares metric calculates the quadratic difference between two images A and B (Eq.5.3). The difference is determined pixel-wise over a user defined region. A_i and B_i are the grey-values at the i^{th} pixel of the corresponding image, N is the number of pixels inside the considered region and $MS\,(A, B)$ is the metrics value. The optimal value of this metric is zero and poor image matches hence result in high metrics values. The mean-squares metric is restricted to images of the same spectral band since it does not allow for intensity differences between the two input images. This metric is useful to compare the distance images since they have no intensity difference. Its capture radius is large i.e. the metric stays robust for large misalignments of the images and does not need very precise initial parameters.

$$MS\,(A, B) = \frac{1}{N} \sum_{i=1}^{N} (A_i - B_i)^2 \tag{5.3}$$

Another very useful metric is based on normalized cross-correlation. It calculates the cross-correlation between the two input images A and B (Eq.5.4). Furthermore, the cross-correlation is normalized by the square root of the autocorrelation of the images. Again, the metric is limited to images obtained using the same modality (identical spectral band). However, it is insensitive to multiplicative factors between the input images due to its normalization. Compared to the mean-squares metric, the capture radius is relatively small.

$$NC\,(A,B) = \frac{\sum_{i=1}^{N}(A_i \cdot B_i)}{\sqrt{\sum_{i=1}^{N} A_i^2 \cdot \sum_{i=1}^{N} B_i^2}} \tag{5.4}$$

The third metric tested is based on mutual information. Mutual information provides a measure to show how much the image intensity of the first image tells about the image intensity of the second image. Since the actual form of dependency of the images does not have to be specified, this metric is very useful for the comparison of multi-modality imagery. It is defined in terms of entropy E of the images A and B (Eq. 5.5). In imagery we deal with discrete data and hence the entropy is described by a sum (integral for continuous data). The input to the entropy are the probability density functions (pdf) p_A and p_B. In case of imagery, the pdf simply is the image histogram of A and B respectively. Usually, the pdf is estimated by superimposing histograms derived from image samples (a process called Parzen windowing).

$$E\,(A) = -\sum p_A(a) \cdot \log p_A(a) \tag{5.5}$$

The joint entropy of the images is defined as

$$E\,(A,B) = \sum p_{AB}(a,b) \cdot \log p_{AB}(a,b). \tag{5.6}$$

The sum of both individual entropies *E(A)* and *E(B)* equals the entropy *E(A,B)* if we consider both images to be completely independent one from another. In other words, no similar information exists (e.g. images captured above two completely different regions with different sensors). However, if there is any equal information, the joint entropy *E(A,B)* apparently must be smaller than the sum of the individual entropies. The difference between the joint entropy and the sum of the individual entropies is the mutual information *I(A,B)* 5.7.

$$I\,(A,B) = E\,(A) + E\,(B) - E\,(A,B) \tag{5.7}$$

An in detail description of the deployed algorithm can be found in [Viola and Wells, 1995].

5.5 The Optimizers

The optimizer's goal is to find the set of transformation parameters that maximizes the metric value (i.e. the similarity) of the fixed and the moving image. The search for optimal parameters takes place in parameter space. The dimension of the parameter space equals the number of parameters. Hence, the more parameters, the more complicated and unstable the

optimization issue becomes (particularly if parameters are correlated). Thus, it is desirable to use transformations with a small number of uncorrelated parameters. Practical testing showed that optimization of more than seven parameters becomes very difficult. More than ten parameters lead to a completely unstable optimization process. However, transformations with more than seven parameters exist. In such cases, the optimization has to be split up in several steps (e.g. rotation, translation, scaling). Different optimizers with various properties were tested. A demanding task is to account for the various parameter scales. Different parameter types have different total values. An optimization step in parameter space may have almost no impact on the registration for one paramter while the same change applied to another parameter shifts the moving image far away from the fixed image. For example, translations on the ground amount to several meters while rotation angles (in radiance) are usually very small. A substraction of 1.6 will shift the moving image by only 1.6 meters on the ground but result in a 90° rotation. Therefore, the most challenging part of the entire registration process proved to be the fine tuning of the parameter scales.

A fairly simple and thus easily understandable optimizer is based on a regular step gradient descent. It is a deterministic optimizer and advances the transformation parameters in the gradient direction. The step size is determined using a bipartition scheme, known from the mathematical field of graph theory. Input parameters for the computation of the current step are the gradient and the direction of the previous step. This approach allows for the reduction of oscillation around local minima ([Bignalet-Cazalet, 2004], p.37).

Another optimizer tested uses a so-called (1+1) evolution strategy [Styner et al., 2000], originally developed for the analysis of medical imagery. It is a stochastic, nonlinear optimizer and belongs to the family of evolutionary algorithms. The basic idea is that a parameter vector (containing the transformation parameters) represents an individual. The individual is asigned a certain energy value displaying its fitness (probability of survival). All individuals of one iteration step form a population. The following optimization step mutates this population (parent population) and create a new population (children). Both generations are added and the size of the combined population is reduced to the size of the parent population. Only the fittest individuals survive. The mutation is carried out by a random vector. Its random values are derived from a normal distribution with the dimension of the parameter space. The mean and the covariance matrix are computed from the current parent population. In case the newly generated population consists of fitter individuals than the previous one, the covariance matrix is increased by multiplication with a coefficient. The covariance matrix is multiplied with a shrink factor if the new generation is less fit.

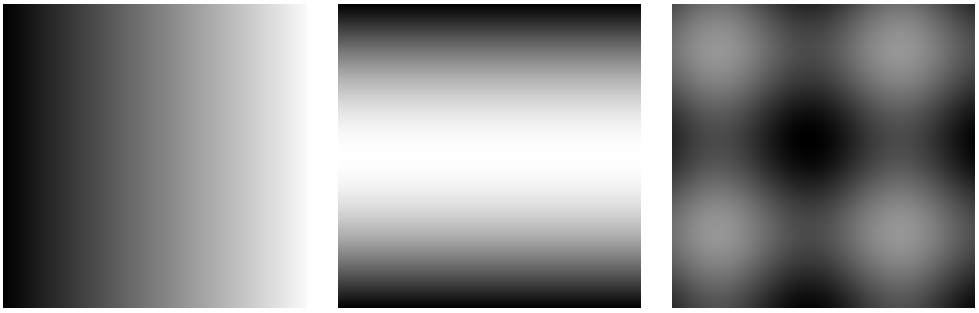
6 Results and Discussion

The previous chapters have explained in detail the components of the image fusion process. Two main components characterize our approach: a preparation for line extraction component and a registration component. Initially, the optical and the SAR image are prepared separately for line extraction. Therefore, optical and SAR image are first ortho-rectified with rather general transformations. Once projected to the ground, preprocessing follows up. The optical image is smoothed with an anisotropic diffusion filter and the SAR image is Frost filtered. Both filters consider local properties of the images. Thus, intra-region smoothing is conducted and contours are preserved. In order to distinguish between rectified ground and unrectified buildings, a classification takes place. It consists of two classifications, an initial one with Support Vector Machines and a final one with Markov Random Fields. In the next step, lines are extracted and distance maps calculated. Remember that line extraction is necessary because we deal with very high resolution imagery. Classical pixel based approaches fail, due to the high level of detail in the images and their multi-modality. The second component of our approach treats the images simultaneously. Two distance maps are input two a modular registration framework. A registration is accomplished iteratively, registering the SAR image onto the optical image. In the following sections, results are provided for the main steps of the developed fusion approach. Finally, the outcomes of for different fusion test programs are shown and evaluated. Reconsider that the emphasis of this project is on the overall development of the registration procedure. Thus, each component result may be further improved.

6.1 Ortho-rectification

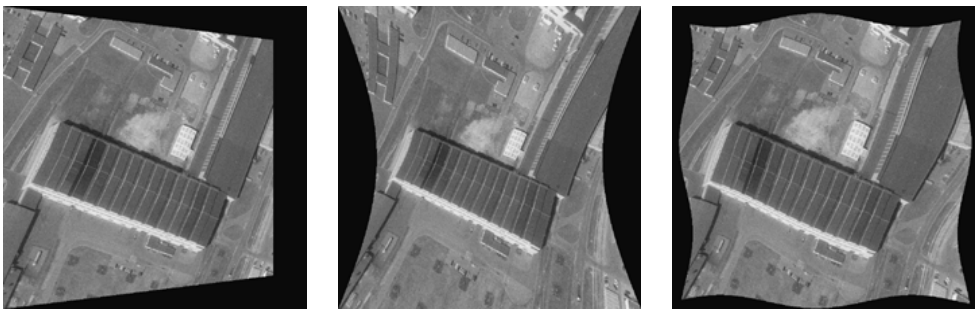
The first step of the image fusion consists of an ortho-rectification of both optical and SAR image. The programmed algorithms were tested using simulated DEMs (Fig. 6.1(a), (b), (c)) since no real DEM was provided for the test images. The DEMs show the same size as the optical and the SAR image, the grey values display the height and their unit is meters. Zero elevation is displayed in black. The brighter the grey value becomes, the more elevated is the DEM. Since the image format (.png) only allows for integer grey values, the height difference between two neighboring pixels of different color is at least one meter. As a consequence, the

DEMs were only used for debugging the source code of the transformations. Fig. 6.2 and Fig. 6.3 show the two test images as seen in Fig. 3.3 after deformation with DEMs from Fig. 6.1. It has to be reconsidered that inverse transforms have been used. In other words, the rectified image in image space is projected to the ground. In a real world application, the image in image space is the deformed one. Thus, the image in image space would be rectified. Here, the images are swapped for testing purposes.



(a) (b) (c)

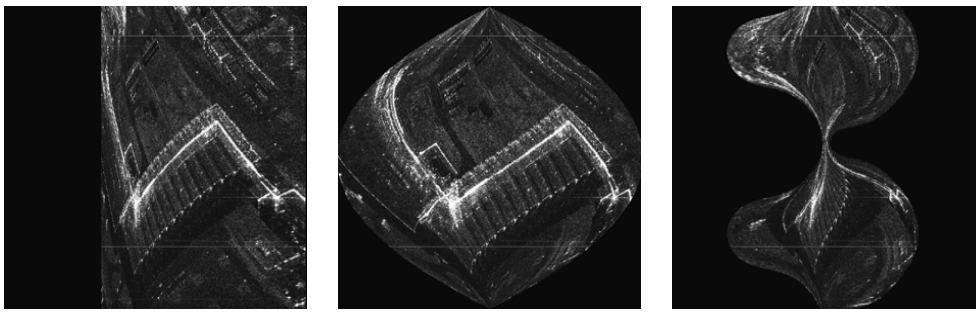
Figure 6.1: (a) DEM inclined towards the right side with height values between 0 m and 255 m, (b) DEM with height values between 0 m and 255 m showing a summit in its middle horizontal axis, (c) DEM with height values between 0 m and 150 m simulating a hilly landscape



(a) (b) (c)

Figure 6.2: (a) Optical image deformed with DEM 6.1(a), (b) Optical image deformed with DEM 6.1(b), (c) Optical image deformed with DEM 6.1(c)

Since the transformation is indirect (refer to section 5.2 for further details), the rectified image has to be determined iteratively. The height value for the first iteration has to be estimated. Our programm simply uses the mean height of the entire DEM. With this initial height value, a first transformation is conducted. For the following iteration, the mean height is calculated in a window, centered on the previously computed coordinate. It leads to a



(a) (b) (c)

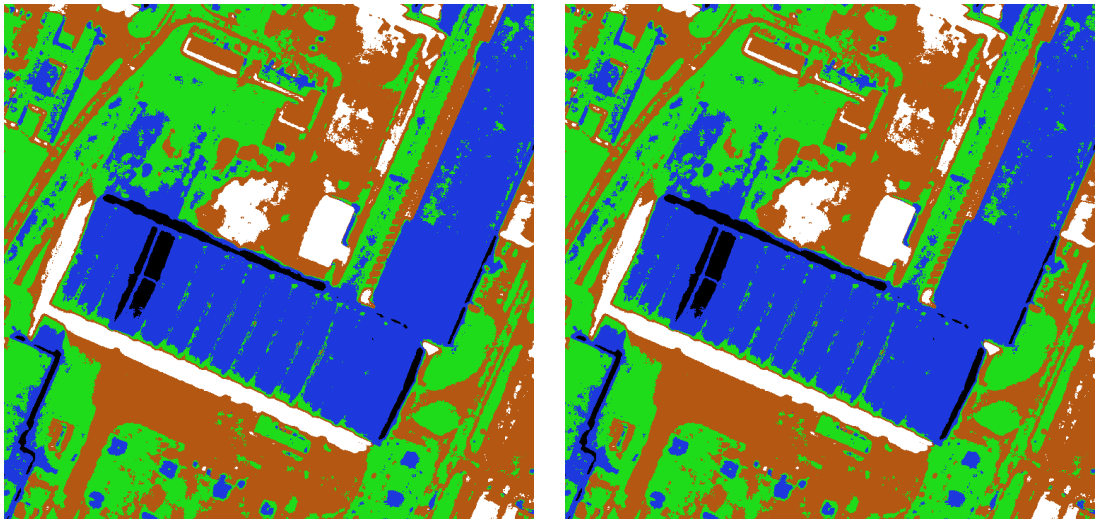
Figure 6.3: (a) SAR image deformed with DEM 6.1(a), (b) SAR image deformed with DEM 6.1(b), (c) SAR image deformed with DEM 6.1(c)

refined transformation into the original image. This procedure is now continued iteratively until the DEM window contains only one height value or the position change in the original image is below a specified threshold. For the optical image, the inverse collinearity equations were used (Eq. 3.3 and Eq. 3.4). The SAR image was ortho-rectified with the inverse Toutin equations (Eq. 3.5, Eq. 3.6 and Eq. 3.7). In contrast to the rectification of the optical image, the SAR image calls for an additional processing step. The unrectified optical image shows only regions the sensor can actually "see". This is not the case for SAR images due to their slant range geometry. Occluded areas are shown in the image and appear as dark regions. Additionally, layover effects appear. With a DEM and the knowledge of the sensor's trajectory and attitude, affected regions can be determined. An algorithm proposed by [Meier et al., 1993] was implemented.

6.2 Results of the Classification

The optical image was classified into five classes, displayed by five different colors (Fig. 6.4). The main goal is to classify the image into the rectified ground level and objects above ground that have not been rectified. Therefore, the classes chosen for the optical image are vegetation (green), roof (blue), soil (brown), shadow (black) and facade (white). Vegetation, soil and shadow are considered as ground. Considering the class vegetation as ground can be justified for the test image because no trees or bushes exist. Soil incorporates all roads, parking lots and other non-elevated man made structures. The shadow class is considered as ground because no elevated objects appear next to the buildings in our particular test image. Usually, the classification of shadow as ground has to be thought over carefully since elevated objects may be hidden in shadowed areas. The SAR image was classified into six classes (Fig. 6.5): occluded area (black), soil (brown), dark roof (yellow), vegetation (green), light

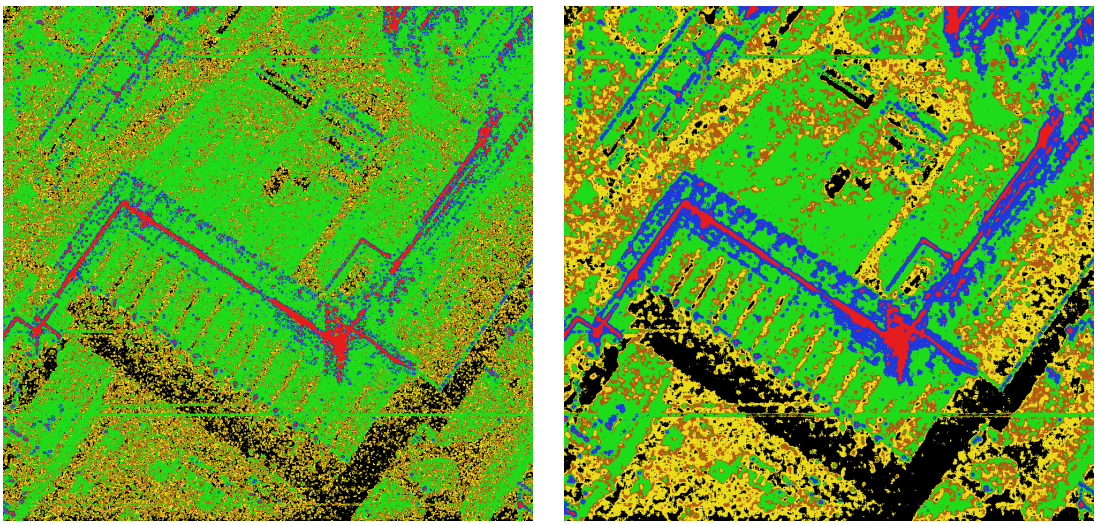
roof (blue) and strong reflectors (red). Occluded area, soil, vegetation and strong reflectors are considered as classes on the rectified ground level. As already mentioned for the shadow class of the optical image, the occlusion class has to be reconsidered since elevated objects may be hidden inside. Strong reflections in urban areas occur due to double bounce effects of the signal. Usually, this double bounce effect appears were building walls meet the ground. Bright L-shaped lines thus are good indicators for the ground level and can be used for the registration. However, the double bounce effect may also occur if a building substructure exists on a roof.



(a) (b)

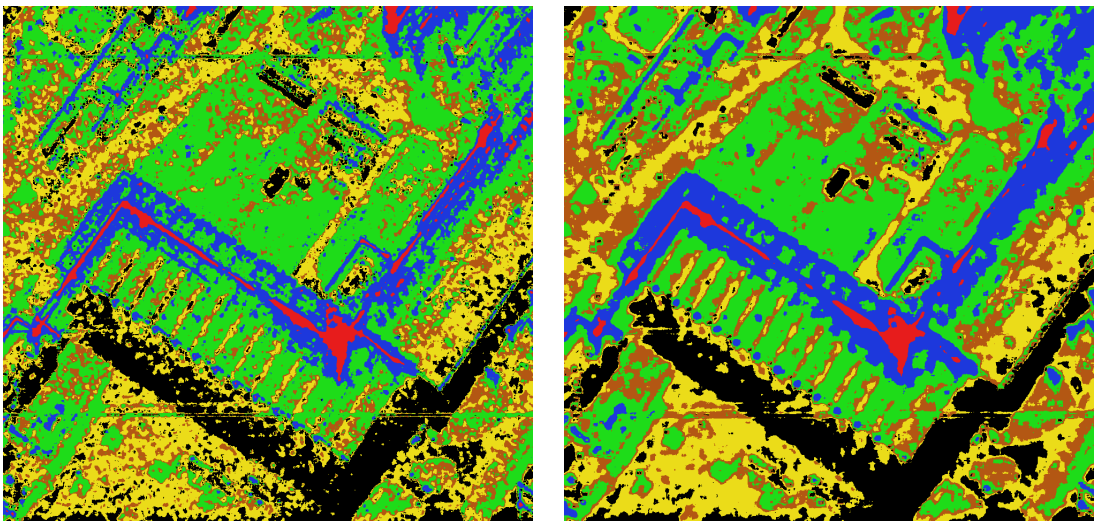
Figure 6.4: (a) Optical image classified into five classes based on image statistics (3x3 filter matrix), (b) Optical image classified into five classes based on image statistics (7x7 filter matrix)

The input images for the SVM classification were computed within two neighborhoods. Mean, median, standard variation, energy, entropy, skewness and kurtosis were calculated within a 3x3 matrix and a 7x7 matrix. Since the 7x7 matrix enlarges the neighborhood, the classification results are smoother. In particular, the classification results obtained from the SAR image are a lot smoother if a larger neighborhood is used. This effect becomes obvious if Fig. 6.5(a) and Fig. 6.5(b) are compared. As already explained in chapter 4, the input to the classification should be the original image. However, tests were also conducted with the Frost filtered SAR image as input to the SVM classification. This approach can be justified as long as consistency is kept during the entire classification process. Therefore, it is not possible to use the SVM classification of the Frost filtered image as initialization for the MRF classification of the original SAR image and vice versa.



(a) (b)

Figure 6.5: (a) SAR image classified into six classes based on image statistics (3x3 filter matrix), (b) SAR image classified into six classes based on image statistics (7x7 filter matrix)



(a) (b)

Figure 6.6: (a) Frost filtered SAR image classified into six classes based on image statistics (3x3 filter matrix), (b) Frost filtered SAR image classified into six classes based on image statistics (7x7 filter matrix)

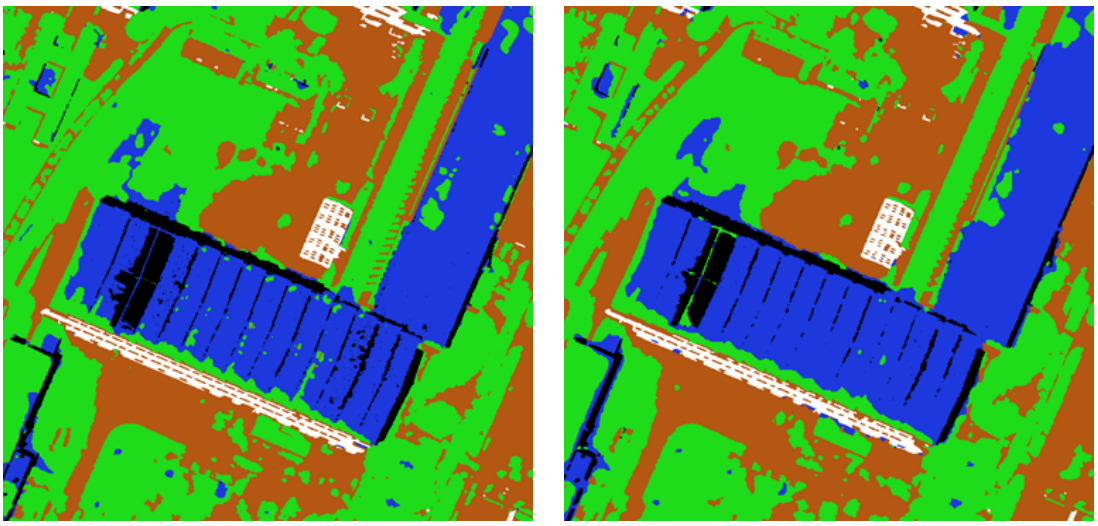
The inputs to the MRF classification are the optical image, the original SAR image and the Frost filtered SAR image. Additionally, the MRF classification uses the SVM result as

initialization. Several parameters have to be specified. For each class, its grey value mean and standard deviation has to be provided because the likelihood term of the MRF assumes the images to follow a Gaussian distribution, although we know this is not an accurate model (see chapter 3). Additionally, a weight parameter beta has to be specified. It has to be reconsidered that beta weights the regularization term in relation to the likelihood term. Increasing beta puts a higher weight on the regularization term. Thus, the influence of the cliques of order two within the 8-connectivity neighborhood of the pixel increases.

(a) (b)

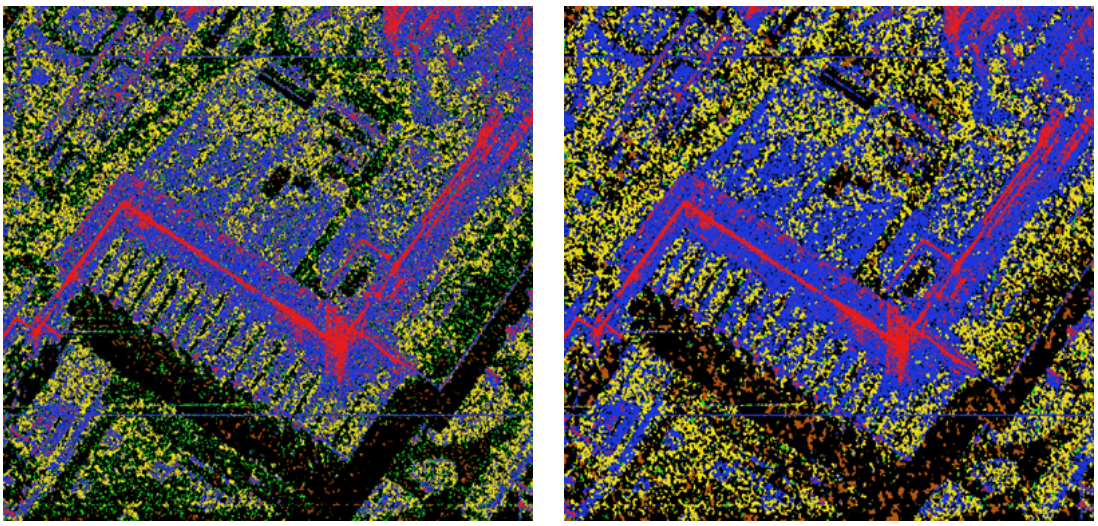
Figure 6.7: (a) The original optical image, (b) The original SAR image

The results of the entire classification, SVM followed up by MRF, do not show very good results. Several reasons account for this lack of classification quality. The first reason to be thought of is the input statistics to the SVM classification. It turns out that the chosen statistics do not describe the image characteristics comprehensively enough for classification purposes. As soon as the input image does not show significant grey value or textures differences between classes, the SVM algorithm will not succeed in distinguishing between the feature vectors. Taking a look at our optical grey value image, we can clearly see that only very small texture differences exist between ground and roof. Additionally, the grey value amplitudes are very similar. The computed image statistics do not succeed in solving this problem. Another criteria for the performance of an SVM algorithm is its mathematical sophistication. In our case, the SVM algorithm is based on a linear model. In consequence of this simple mathematical model, similar feature vectors cannot be separated by a hyperplane, i.e. the classification fails. Therefore, non-linear SVMs should be tested in order to enhance



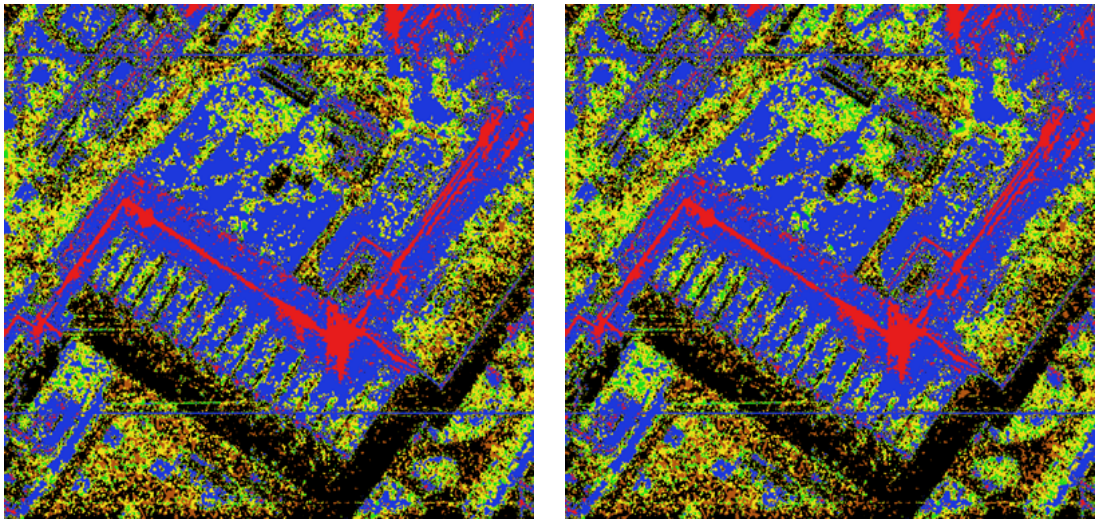
(a) (b)

Figure 6.8: (a) The original optical image after classified with MRF, beta set to 4, initialization with SVM classification result from Fig. 6.4(a), (b) The original optical image after classification with MRF, beta set to 5, initialization with the SVM classification result from Fig. 6.4(b)



(a) (b)

Figure 6.9: (a) The original SAR image classified with MRF, beta set to 3, initialization with SVM result from Fig. 6.5(a), (b) The original SAR image classified with MRF, beta set to 4, initialization with SVM result from Fig. 6.5(b)



(a) (b)

Figure 6.10: (a) The Frost filtered SAR image after classification with MRF, beta set to 4, initialization with SVM result from Fig. 6.6(a), (b) Frost filtered SAR image classified with MRF, beta set to 4, initialization with SVM classification result from Fig. 6.6(b)

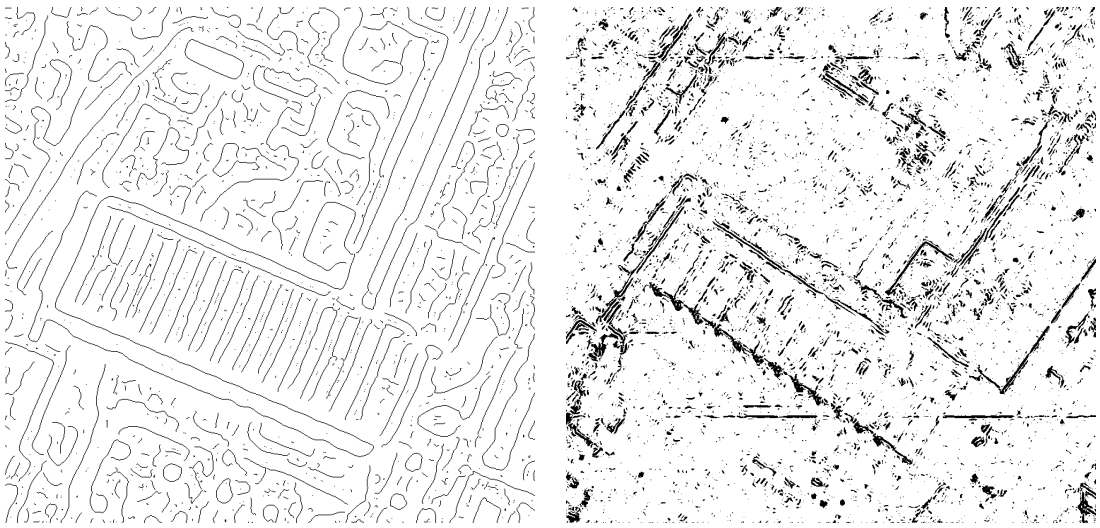
the classification results. Taking into account those drawbacks, the initial classification for the MRF refinement already lacks quality.

The MRF classification used in this project assumes the image grey values to follow a Gaussian distribution. This fact can more or less be justified for the optical image but not for the SAR image. The SAR image classification wit SVM suffers from similar problems as the optical image but the consequences are worse. The statistical parameters which are put into the SVM algorithm do not allow for sufficient distinction between the classes. In particular, the distinction between vegetation and roof has to be considered almost a failure. The results do not significantly improve using the MRF classification for a very particular reason. The MRF classification deployed in this project considers the grey values of the image to follow a Gaussian distribution. While this fact can be justified for the optical image, it lacks justification for the SAR image. As explained in detail in section 3.4.2, the mathematical model of the SAR image is considered multiplicative and the speckle effect exists. The distribution of the SAR amplitude image is not Gaussian. Hence, the distribution of the SAR image used in this project cannot be assumed to be Gaussian. A possible solution to this issue is given in [Tison et al., 2004]. A mathematical model relying on the Fisher distribution is proposed in order to model high resolution scenes of urban areas. An alternative approach

for the classification of optical imagery based on watersheds and self-organizing Kohonen maps [Kohonen, 1990] was proposed in [Poulain, 2007]. The watershed segmentation leads to segments that are statistically and geometrically analysed. For example, invariant geometric moments proposed in [Flusser and Suk, 2006] are one component of the input feature vector to the Kohonen map. Using rotationally invariant geometric moments could improve our classification results as well because we deal with urban areas. In urban areas, geometric information is densely distributed. In our project, no geometric information has been included for the classification. Its introduction would enhance classification results.

6.3 Feature Extraction Results

Lines were extracted from the optical image (Fig. 6.11(a)) and from the SAR image (Fig. 6.11(b)). The lines in the optical image were extracted with the Canny edge detector. Before line extraction was conducted, the optical image was filtered with the anisotropic diffusion filter.



(a) (b)

Figure 6.11: (a) Lines extracted with the Canny algorithm applied to the anisotropic diffusion filtered optical image, (b) Lines extracted with the Tupin algorithm applied to the Frost filtered SAR image

Lines in the SAR image were extracted using the algorithm developed by [Tupin et al., 1998] and a threshold was applied. The input to the algorithm was the Frost filtered SAR image. This line extraction algorithm was adapted to SAR imagery. Hence, good results can also be obtained with the original SAR image as input. However, testing showed that the

line extraction based on the Frost filtered image improves results. The grey values of the line images have been inverted for visualization reasons. Obviously, the line extraction results call for further enhancement. In particular the lines obtained from the SAR image show many small line pieces which leads to problems during the registration process. Improvements should comprise the assembly of many small line pieces to longer line segments. Small line pieces that do not contribute to the construction of longer line segments should be eliminated. Different approaches may be thought of e.g. based on a Hough transformation. Additionally, algorithms capable of providing the desired properties already exists in several road network extraction or road network updating softwares. They should be tested on the images of this project in order to refine line extraction results.

6.4 Fusion Results

After ortho-rectification, preprocessing, classification and line extraction, distance maps are calculated (see Fig. 4.8). Their computation is based on the line images. Each pixel value in the distance image displays its Euclidean distance to the closest pixel of a line object. Distance mapping was found necessary in order to prevent residuals that arise from different absolute line positions. For the fusion step, the optical distance map and the SAR distance map are input to the ITK registration framework. We define one image to be the fixed image and the other one to be the moving image. The moving image will be deformed and mapped onto the fixed image. In our case, we choose the optical distance map as fixed image. Hence, the SAR distance map is mapped onto the optical one. In order to visualize the final registration results, a checkerboard image is introduced (Fig. 6.12). The inputs to the checkerboard image are the fixed image and the resulting deformed SAR image after the registration.

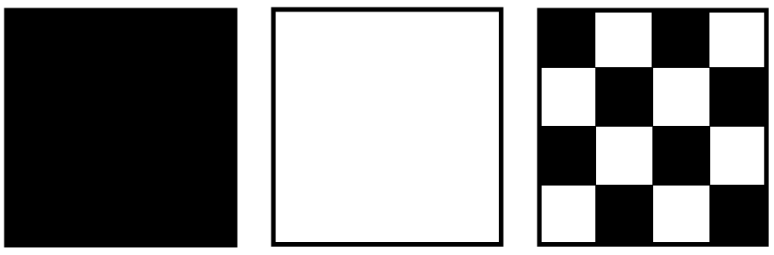


Figure 6.12: (a) Fixed image (optical), (b) Deformed moving image (SAR), (c) Checkerboard of the two images

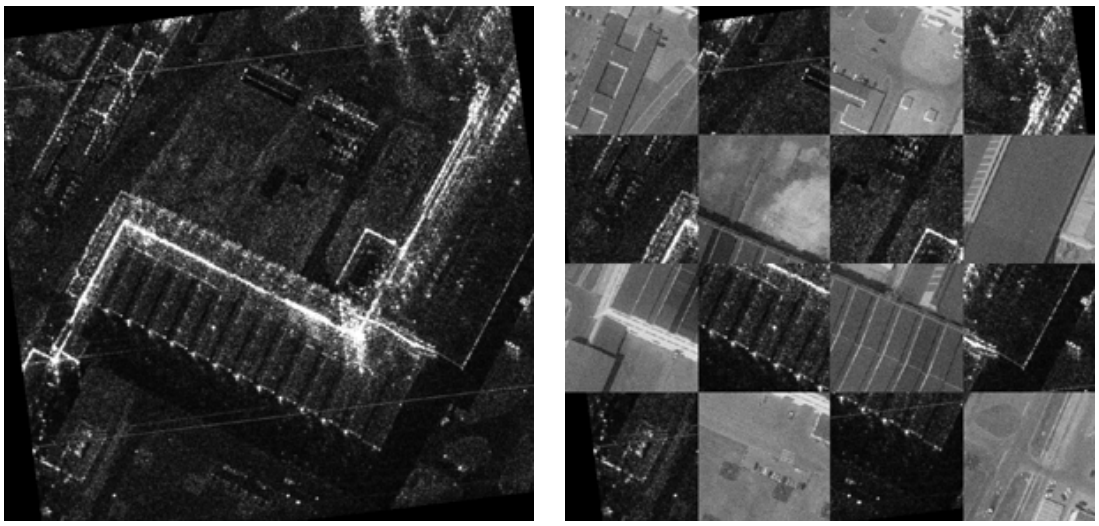
Remember that the registration has a modular architecture. Its four modules have to be chosen carefully in order to achieve meaningful results. A transform and an interpolator are specified for the mapping. In order to measure the similarity of the fixed and the deformed moving image, a metric is specified. An optimizer drives the adjustment of the iterative search for appropriate transform parameters, depending on the metric values. The greater the similarity between fixed and deformed moving image, the better the optimizer converges towards a sufficient solution. A variety of algorithms for any modul exist. The algorithms chosen for our image registration are described in chapter 5. All of them have different properties. By chosing appropriate module combinations, the registration framework can be adapted to the particular input images. In this project, we test four different module combinations in separate test programs (Tab. 6.1).

Test Program	Transformation	Interpolator	Metric	Optimizer
Program A	Rot. + Transl.	Bilinear	Mean Squares	Reg. Step Grad. Desc.
Program B	Rot. + Transl.	Bilinear	Norm. Cross-Corr.	Reg. Step Grad. Desc.
Program C	Rot. + Transl.	Bilinear	Mutual Inf.	1+1 Evolutionary
Program D	Rot. + Transl.	Bilinear	Norm. Cross-Corr.	1+1 Evolutionary

Table 6.1: Registration test programs

Since the performance very much depends on the transform parameters, the same simple transform is used for all four test programs. It applies a centered rotation, followed by a translation in x and y. The center can either be the grey value gravity center or the geometric center of the image. Since our test images cover the same ground region, the transform is centered on the geometric image center. The interpolation has the least impact on the registration result. Hence, bilinear interpolation is conducted in all registration test programs. Hence, the entire mapping part (transform and interpolation) is not varied. Three different metrics are tested: mean squares, normalized cross-correlation, and mutual information. Optimizers drive the execution of the registration process. We test two optimizers: regular step gradient descent and 1+1 evolutionary. Different parameters have to be specified for each module. Thus, the number of parameters varies for the four test programs. Initial parameters have to be input to them. They have a big impact on the quality of the result, the computational costs and the convergence. We distinguish between the initial transform parameters and the parameters for the optimizer. Although the chosen transformation has only five parameters (the rotation angle, the image center coordinates, two translations), attention has to be paid to differences in the corresponding units. The angle has the unit radian ([rad]) and hence the values are in the range $[-\pi,\ \pi]$. However, the rotation center and the translations are measured in millimeters ([mm]). Hence, different scales have to be

chosen for the parameter optimization. This normalization of the parameter scales is absolutely necessary in order to obtain sufficient results with the ITK registration framework. The 1+1 evolutionary optimizer parameters and the regular step gradient descent optimizer parameters define the capture radius and the maximum number of iterations. Additionally, the regular step gradient descent optimizer calls for an initial step length, a minimum step length and a relaxation factor. The initial step length defines the initial change of the transform parameters. A relaxation factor specifies the rate at which the optimizer's step length in parametric space is reduced. Minimum step length defines the convergence tolerance, i.e. the parameter change for the final transform. As soon as minimum step length is reached, the registration ends. In case, the optimizer does not reach the minimum step length, the maximum iteration number terminates the registration process.

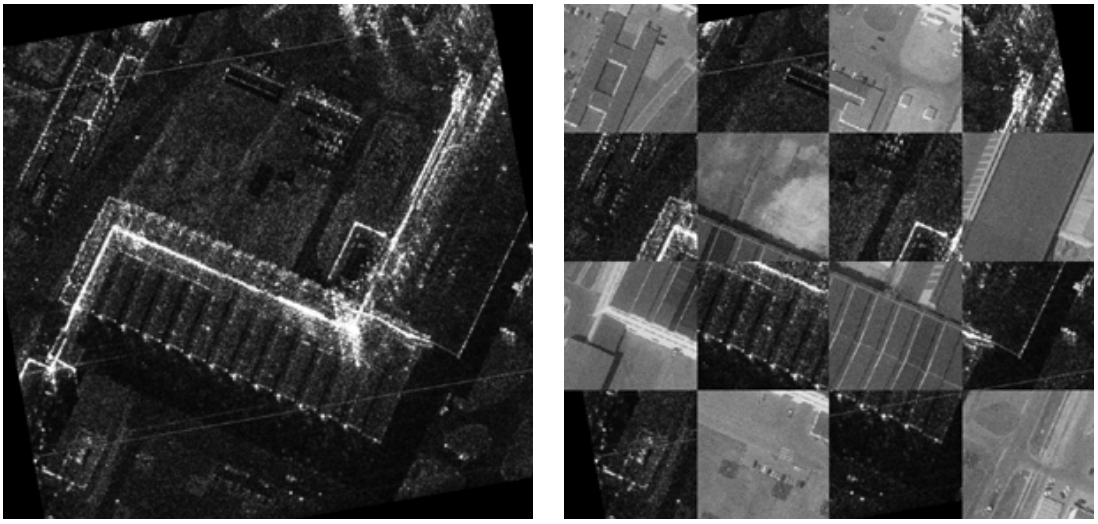


(a) (b)

Figure 6.13: (a) SAR Image after registration with Program A, (b) SAR Image registered onto the optical image with Program A

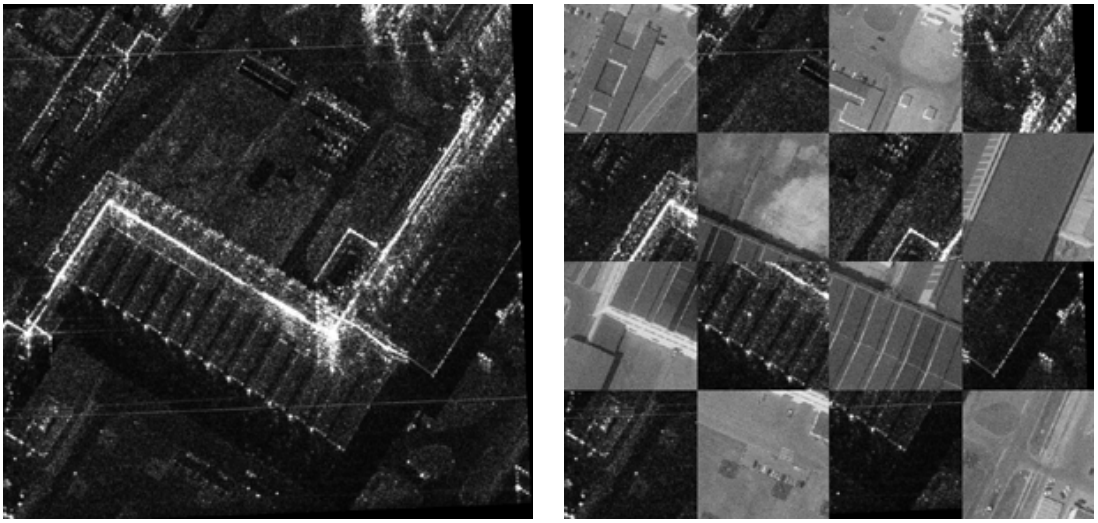
Test Program	Rotation [deg]	Translation in x, y [mm]
Program A	7.15	39.29, -45.83
Program B	10.01	71.02, -60.03
Program C	2.30	31.78, -6.84
Program D	4.33	33.12, -27.59

Table 6.2: Final transformation parameters that are applied to the SAR image



(a) (b)

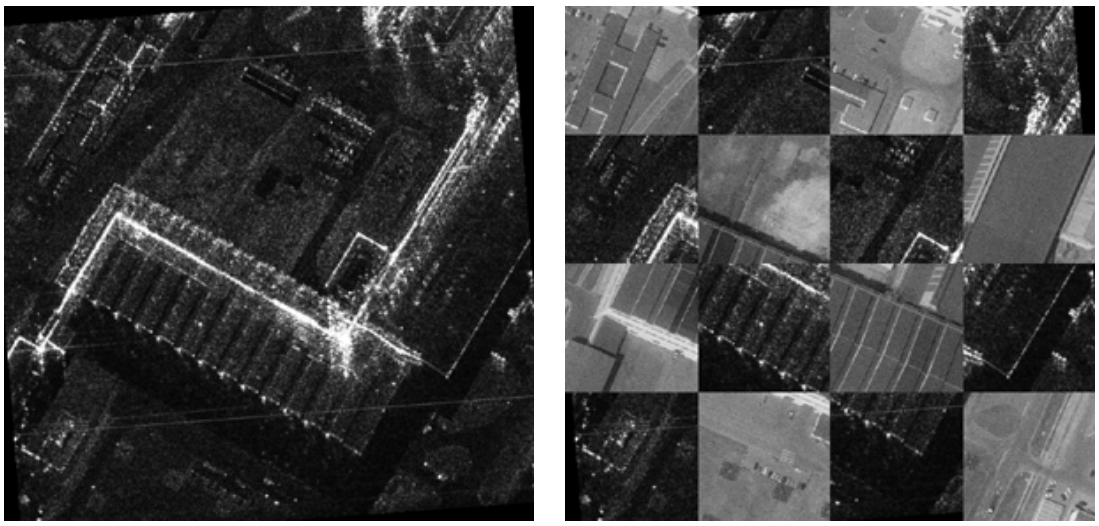
Figure 6.14: (a) SAR Image after registration with Program B, (b) SAR Image registered onto the optical image with Program B



(a) (b)

Figure 6.15: (a) SAR Image after registration with Program C, (b) SAR Image registered onto the optical image with Program C

A performance evaluation of the four test programs reveals the pros and cons. Test program A gives good results (Fig. 6.13). It is relatively fast to compute, too. Image metric (mean squares) and optimizer (regular step gradient descent) work well on the distance images. The



(a) (b)

Figure 6.16: (a) SAR Image after registration with Program D, (b) SAR Image registered onto the optical image with Program D

outcome is also fairly accurate. Program B provides the best results in terms of accuracy (Fig. 6.14) and computation speed. The metric based on normalized cross-correlation in combination with the regular step gradient descent optimizer is the best combination for our case. Program C is the least accurate, probably due to its metric based on mutual information. Actually, mutual information is usually applied directly to multi-modality imagery with different radiometries. In our case, there is no need for using a mutual information metric. Due to our main idea of feature based registration, the radiometric differences disappear. Hence, we measure the similarity of two images with the same radiometry: the distance images. Program D (Fig. 6.16) provides better results than C but the inital parameters are difficult to adjust. Small changes in the initial parameters lead to big changes in the outcome. Metric (normalized cross-correlation) and optimizer (1+1 Evolutionary) do not work well together. Considering the results of the test programs, it becomes obvious that the 1+1 evolutionary optimizer and the mutual information metric are not appropriate for registering the distance images. Combinations of the normalized cross-correlation metric and the mean squares metric with the regular step gradient descent optimizer give good results. Since no radiometric difference appear, these two metrics are sufficient for registering the distance images.

In fact, the transformation implemented in the ITK registration framework is too simple for a comprehensive modelling of the SAR residuals. Other rigid transformations would

provide better results e.g. an affine transformation or a rational polynomial transform. It is also possible to use the initial collinearity equation or the radar equation respectively as transformations. Both collinearity and radar equation were tested but showed to be very complexe in terms of parameter estimation. Since they show at least eleven parameters, the optimizers provided in ITK showed convergence difficulties. It was found that the ITK optimizers work fine with transforms containing not more than seven parameters. Transforms that show more than seven parameters, like the collinearity and the radar equations, have to be split up. The optimization may then be conducted in several steps. Usually, the first step to be optimized is the rotation, followed by a translation. Thereafter, scales can be optimized and so on. This step-wise optimization could be carried out iteratively thus refining the entire optimization. Additionally, tests should be conducted with non-rigid transformations that align the images locally. For example, an approach based on the finite elements method (FEM) provided in OTB may be thought of.

7 Conclusions and Future Perspectives

We began the description of the optical/SAR image fusion with the theoretical background. Thereafter, the test data was introduced and the entire registration strategy was outlined. In the following chapters and sections, a detailed understanding of each image registration component was provided step by step. Results were displayed and discussed component-wise in the previous chapter. This chapter gives a summary of the project (section 7.1), draws some final conclusions and finishes with an outlook (section 7.2).

7.1 Summary

The goal of this project was the development of a fusion strategy for high resolution optical and SAR images (Fig. 7.1). Due to the different sensors and the high resolution of the images, classical pixel based approaches fail. Hence, our main idea was image fusion based on extracted features. Since we deal with urban scenes, lines were extracted. In order to prepare for line extraction, both images were ortho-rectified. Distortions due to the different sensor geometries and the terrain were reduced with rather general geometric models. For the optical case, the collinearity equations were used. The SAR image was rectified with the Toutin equations. In the next step, the optical image was filtered with an anisotropic diffusion filter and the SAR image was Frost filtered for despeckling reasons. Following up was a classification in order to distinguish between rectified ground and unrectified buildings. Initially, a Support Vector Machines classification was carried out. Thereafter, preliminary classification results were refined with a Markov Random Field classification. Finally, lines could be extracted. We applied the Canny filter to the optical image and the Tupin line detector to the SAR image. Then, distance maps were calculated. This step was found necessary in order to avoid errors introduced by different absolut positions of the extracted lines. The distance maps were input to the ITK registration framework and registered iteratively. Four programs, combining various mappings, metrics and optimizers, were tested. A rather simple mapping algorithm was implemented. It consists of a centered rotation plus translation and a bilinear interpolation. Best performances could be achieved with a normalized cross-correlation metric and a regular step gradient descent optimizer. Finally, we obtained a deformed SAR image that was successfully registered onto the optical image.

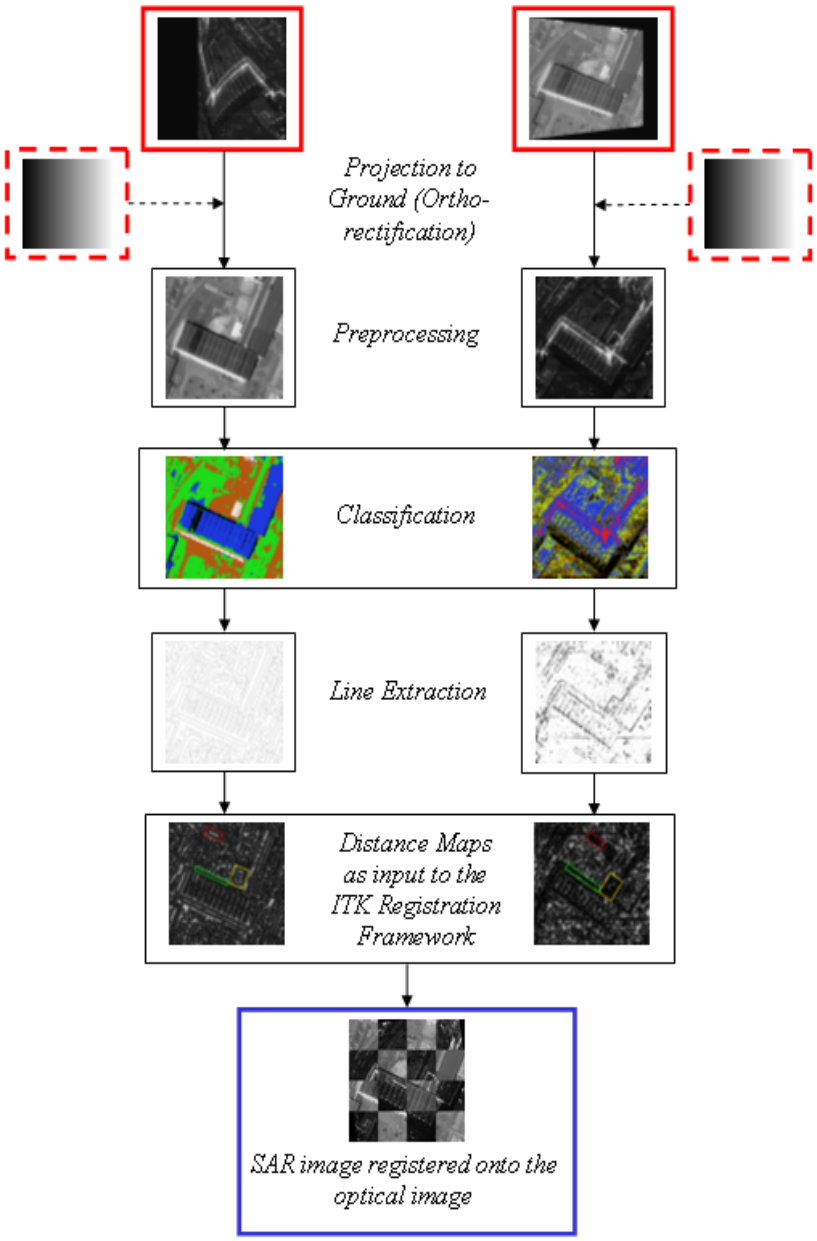


Figure 7.1: Review of the entire image optical/SAR image fusion

7.2 Conclusions and Outlook

In conclusion, the overall registration strategy proved to be successfull. Its integration into the ORFEO Toolbox will provide additional possibilities. It will give new ideas and serve as a basis for further improvements. None the less, further refinement of all components would significantly improve results.

- The ortho-rectification of the SAR image could possibly be enhanced with an approach based on Doppler and range equations [Raggam et al., 1993; Gelautz et al., 1998; Sörgel, 2003]. Since this alternative approach contains the real physical parameters of the SAR sensor, a refinement of the satellite's parameters is also possible [CNES and ENST, 2005].

- The detection of occluded and layover regions is imperative and should be further tested in the first place. The implemeted algorithms could only be tested with the simulated DEMs, yet. Disturbing effects occur due to the high difference between neigboring height values. They should be tested with real DEMs in order to evaluate the performance of the programmed algorithm.

- For this project, the SVM classification had to be tested but does not show sufficient outcomes. Alternative classification approaches should be tested, too, in particular for the optical image. Watershed segmentation followed up by a classification with self-organizing Kohonen feature maps shows promising results. For the SVM classification, the incorporation of further descriptive features, e.g. geometric moments, would probably also improve results. The inclusion of semantic information could be thought of, too. For example, the distinction between ground and roof in the SAR image could be enhanced by declaring all regions, framed by L-shaped bright lines and occludded areas, as roofs.

- The Markov Random Field classification of the SAR image assumes a Gaussian distribution. As already previously outlined, an amplitude SAR image is Rayleigh distributed. For urban scenes, approaches, based on the Fisher distribution, have been successfully tested. In order to improve MRF classification results of the SAR image, an appropriate distribution has to be chosen and integrated into the source code.

- Feature extraction already works well and thus only calls for minor changes. The number of small line pieces should be significantly reduced. Lots of small line pieces lead to disturbing effects in the distance maps. Longer line segments could be derived from the assembly of multiple line pieces. An approach originally developed at the

CNES for the extraction of road networks could possibly solve this issue. It still has to be adapted to the SAR image though.

- So far, the threshold of the Tupin line detector is chosen manually. In order to not only adapt the algorithm to our test image but to all kinds of SAR images, an automatic threshold determination is necessary.

- Another part calling for improved performance is the mapping component of the ITK registration framework. In order to really take care of the residuals still present in the SAR image, more sophisticated transforms are absolutely necessary. At least, anisotropic scaling has to be integrated as soon as possible. Furthermore, non-rigid transformations should be tested. An algorithm based on the Finite Elements Method (FEM) already exists in OTB. It will be applied to the test images leading to disparity maps.

- After refinement of the image fusion algorithm, the computation of Digital Surface Models (DSM) seems possible. Reconsider that all elevated objects in the optical and in the SAR images stay distorted. Due to the different sensor geometries, they show different perspectives of the same objects. Thus, three-dimensional construction of buildings, based on the unrectified regions in optical and SAR image, could be integrated.

- Finally, the entire fusion strategy should be tested on further test images and corresponding DEMs! Throughout this entire project, all algorithms developed and programmed have been evaluated with only one optical and one SAR image. All DEMs have been simulated.

In conclusion, the entire fusion, classification and DSM computation should be regarded as one step. Since improved classification results will lead to an improved fusion and vice versa, an iterative procedure could be thought of. Results of a first fusion/classification/DSM should be input to a second one and so on.

Bibliography

Agenzia Spaziale Italiana (ASI). Cosmo - skymed, accessed 9th July at 10:30 2007. `http://www.asi.it/cosmoskymed/About_COSMO-SkyMed.htm`.

David Assemat, Jean-Claude Durand, and Joelle Guinle, editors. *Space Technology Course - Spacecraft Techniques and Technology - Volume 2: Payloads*. Centre National d'Études Spatiales (CNES) Space Technology Course. Cépaduès - Edition, Toulouse, 2005.

J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Stat. Soc.*, pages 259 – 302, 1986.

Francois Bignalet-Cazalet. Recalage automatique d'images sans modèle de prise de vue. Master's thesis, École Nationale Supérieure de l'Aéronautique et de l'Espace (SUPAERO), 2004. Projet de fin d'étude accomplished at the Centre National d'Études Spatiales (CNES), Toulouse.

John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679 – 698, November 1986.

CNES. International Charter "Space and Major Disasters", accessed 26th July at 11:30 2007a. `http://www.disasterscharter.org`.

CNES. PLEIADES - Dual optical system for metric resolution observations - ORFEO Accompaniment Program, accessed 25th July at 15:20 2007b. `http://smsc.cnes.fr/PLEIADES/Fr/A_prog_accomp.htm`.

CNES. Pléiades - Dual optical system for metric resolution observations, accessed 9th July at 09:00 2007c. `http://smsc.cnes.fr/PLEIADES/index.htm`.

CNES and ENST. Étude CNES DCT/SI/AP 04-259 - Recalage d'images optique et radar par extraction de primitives. Technical report, École Nationale Supérieure des Télécommunications Paris - Départment Traitement de Signal et des Images, June 2005.

C. Cortes and V. Vapnik. Support Vector Networks. *Machine Learning*, 20(3):1 – 25, 1995.

P.E. Danielsson. Euclidean distance mapping. *Computer Graphics and Image Processing*, 14: 227 – 248, 1980.

Jan Flusser and Tomás Suk. Rotation Moment Invariants for Recognition of Symmetric Objects. *IEEE Transactions on Image Processing*, 15(12):3784 – 3790, 2006.

Clive Fraser. Departures from Collinearity and Camera Calibration. Lecture notes of the lecture Photogrammetry 451-447, part of the Geomatics course at the University of Melbourne, `http://www.sli.unimelb.edu.au/cfraser`, downloaded 8th August at 12:00 2007a.

Clive Fraser. High-Resolution Satellite Imagery for Mapping. Lecture notes of the lecture Photogrammetry 451-447, part of the Geomatics course at the University of Melbourne, `http://www.sli.unimelb.edu.au/cfraser`, downloaded 8th August at 12:00 2007b.

V.S. Frost, J. Abbott Stiles, K.S. Shanmugan, and J.C. Holtzman. A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(2):157 – 166, March 1982.

Frédéric Galland, Florence Tupin, Jean-Marie Nicolas, and Michel Roux. Registering of Synthetic Aperture Radar and optical data. In *IEEE International Geoscience and Remote Sensing Symposium, 2005*, volume 5, pages 3513 – 3516. IGARSS '05, July 2005.

M. Gelautz, H. Frick, J. Raggam, J. Burgstaller, and F. Leberl. SAR Image Simulation and Analysis of Alpine Terrain. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53: 17 – 38, 1998.

J. W. Goodman. *Statistical Optics*. John Wiley and Sons, Inc, USA, 1st edition edition, 1985.

Christian Heipke. Photogrammetrie und Fernerkundung I - III. Lecture notes of the diploma course Geodesy and Geoinformatics at the Leibniz University Hannover, July 2007.

Luis Ibánez, Will Schroeder, Lydia Ng, and Josh Cates. *The ITK Software Guide*, 2 edition, November 2005. Updated for ITK version 2.4.

Jordi Inglada and Frédéric Adragna. Automatic Multi-Sensor Image Registration by Edge Matching using Genetic Algorithms. In *IEEE International Geoscience and Remote Sensing Symposium, 2001*, volume 5, pages 2313 – 2315. IGARSS '01, 9-13 July 2001.

Jordi Inglada, Thomas Feuvrier, Patrick Imbo, Caroline Ruffel, and Romain Garrigues. *The ORFEO Tool Box Software Guide*, 2 edition, October 2006. Updated for OTB version 1.0.2.

Jordi Inglada and Alain Giros. On the Possibility of Automatic Multisensor Image Registration. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2104 – 2120, October 2004.

Jordi Inglada and Hélène Vadon. Fine registration of SPOT5 and Envisat/ASAR images and ortho-image production: a fully automatic approach. In *IEEE International Geoscience and Remote Sensing Symposium, 2005*, volume 5, pages 3510 – 3512. IGARSS '05, July 2005.

Karsten Jacobsen. Geometry and Information Contents of large size Digital Frame Cameras. In *ISPRS Hannover Workshop 2007: High-Resolution Earth Imaging for Geospatial Information*, May 29 - June 1 2007.

Michel Kasser and Yves Egels, editors. *Photogrammétrie numérique*. Collection ENSG/IGN, Ecole Nationale des Sciences Géographiques. Hermès Science Publications, Paris, 2001.

Kitware. Nlm Insight Segmentation and Registration Toolkit (ITK), accessed 25th July at 15:10 2007. `http://www.itk.org`.

T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78:1464 – 1480, September 1990.

Jong-Sen Lee. Speckle Analysis and Smoothing of Synthetic Aperture Radar Images. *Computer Graphics and Image Processing*, 17:24 – 32, 1981.

A. Lopès, E. Nezry, R. Touzi, and H. Laur. Structure Detection and Statistical Adaptive Speckle Filtering in SAR Images. *International Journal of Remote Sensing*, 14(9):1735 – 1758, 1993.

E. Meier, U. Frei, and D. Nüesch. *Precise Terrain Corrected Geocoded Images*. In: G. Schreier: *SAR Geocoding: Data and Systems*. Wichmann, Karlsruhe, Germany, 1993.

D. F. Morrison. *Multivariate Statistical Methods*. McGraw-Hill Book Company, New York, USA, 2nd edition edition, 1976.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629 – 639, July 1990.

Vincent Poulain. Détection de changements entre images et bases de données vecteur. Master's thesis, École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT), 2007. Projet de fin d'étude accomplished at the Centre National d'Études Spatiales (CNES), Toulouse.

J. Raggam, D. Strobl, and W. Hummelbrunner. *Product Quality Enhancement and Quality Evaluation.* In: G. Schreier: *SAR Geocoding: Data and Systems.* Wichmann, Karlsruhe, Germany, 1993.

Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond.* The MIT Press, Cambridge, Massachusetts, 2002.

John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis.* Cambridge University Press, Cambridge, United Kingdom, 2004.

Marc Sigelle and Florence Tupin. Champs de Markov en traitement d'image. Technical Report C3M, École Nationale Supérieure des Télécommunications Paris - Départment Traitement de Signal et des Images, July 1999.

Uwe Sörgel. *Iterative Verfahren zur Detektion und Rekonstruktion von Gebäuden in SAR- und InSAR-Daten.* PhD thesis, Fachbereich Elektrotechnik und Informationstechnik der Universität Hannover, March 2003.

Uwe Sörgel. Radarfernerkundung: Elektromagnetische Wellen und Materie. Lecture notes of the diploma course Geodesy and Geoinformatics at the Leibniz University Hannover, April 2006.

Martin Styner, Christian Brechbühler, Gábor Székely, and Guido Gerig. Parametric Estimate of Intensity Inhomogenities Applied to MRI. *IEEE Transactions on Medical Imaging*, 19 (3):153 – 165, 2000.

Céline Tison, Jean-Marie Nicolas, Florence Tupin, and Henri Maître. A new statistical model of urban areas in high resolution SAR images for Markovian segmentation. *IEEE Transactions on Geoscience and Remote Sensing (IGARSS'03 special issue)*, 42(10):2046 – 2057, October 2004.

Jean-Yves Tourneret. Classification et Reconnaissance des Formes. Lecture notes of the course Mastère Recherche Signal Image Acoustique Optimisation at the École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT), Toulouse, 2003.

Thierry Toutin, Yves Carbonneau, and Louiselle St-Laurent. An Integrated Method to Rectify Airborne Radar Imagery using DEM. *Photogrammetric Engineering and Remote Sensing*, 58(4):417 – 422, April 1992.

R. Touzi, A. Lopès, and P. Bousquet. A statistical and geometrical edge detector for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 26:764 – 773, November 1988.

Florence Tupin, Henri Maître, Jean-Marie Mangin, Jean-Marie Nicolas, and Eugène Pechersky. Detection of Linear Features in SAR Images: Application to Road Network Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 36(2):434 – 453, March 1998.

P. Viola and W.M. Wells. Alignment by maximization of mutual information. In *Fifth International Conference on Computer Vision*, pages 16 – 23, 20 - 23 June 1995.

F.Y. Wu. The Potts Model. *Reviews of modern physics*, 54(1):235 – 267, 1982.

List of Figures and Tables

Figures: All non-referenced figures have been drawn by the author himself.

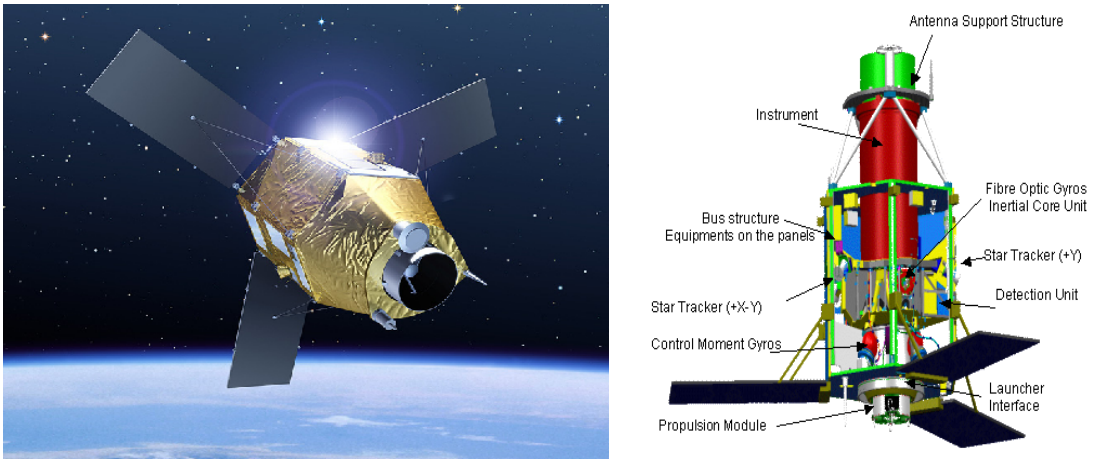
List of Figures

List of Tables

Annex A

The ORFEO Program

The Optical and Radar Federated Earth Observation program (ORFEO) is a French-Italian high resolution earth observation satellite programme, incorporating the optical system Pléiades (France) and the SAR system COSMO-SkyMed (Italy). It was designed to satisfy both civilian and military needs. Key fields of applications are: Defence, risk management and assessment, humanitarian aid, cartography, urban and rural planning, geology, geophysics, hydrology, agriculture, forestry, sea and coastline monitoring.



(a) (b)

Figure 7.2: (a) Simulation of a Pléiades satellite on its orbit (© CNES), (b) Structure of the Pléiades satellites (© CNES)

With its two agile satellites the optical system Pléiades is capable of providing (parameters taken from the CNES webpage [CNES, 2007c]):

- daily access to every point on earth (excluding inner polar regions),
- a resolution of 0.7 m in vertical viewing panchromatic mode,

- four spectral bands (blue, green, red and near infrared, see Fig. 7.2) with a resolution of 2.8 m in vertical viewing,
- a field of view of 20 km,
- an acquisition of a 120 km by 120 km mosaic in the same orbit,
- nearly instantaneous stereoscopic image couples (or even triplet) of 20 km by 300 km,
- cloud free images covering a total area of 2,500,000 km^2 per year,
- a very accurate positioning of the images (<1 m/1000 km with ground control points) facilitating data exploitation within Geographical Information Systems (GIS) (see positioning devices in Fig. 7.2(b)).

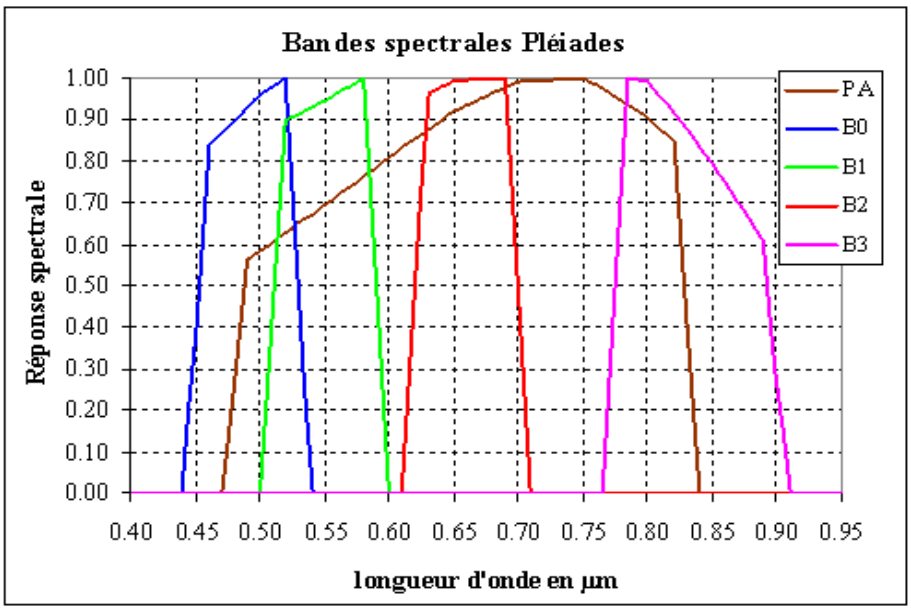


Figure 7.3: Spectral bands of the Pléiades satellites (© CNES)

With its four flexible SAR satellites (see Fig.7.2 for satellite layout) the COSMO-SkyMed system is capable of providing (parameters taken from the ASI webpage [, ASI]):

- all weather and night/daylight observations with X-Band SAR,
- multi-polarimetric and multi-temporal imaging modes,

- images in ScanSAR mode (swathwidth 100 by 100 km^2, geometric resolution 30 - 100 m),
- images in Stripmap mode (swathwidth 30 by 30 km^2, geometric resolution 3 - 15 m),
- images in Spotlight mode (swathwidth 10 by 10 km^2, geometric resolution 1 m),
- revisit time of any point on earth (excluding the inner polar regions) of less than 12 hours with all four satellites operational,
- interferometric image couples with a temporal separation of four days (in normal mode) or 20 seconds (in tandem mode),
- image delivery for very urgent applications within 18 hours,
- overall system lifetime of 15 years with an individual satellite lifetime of five years.

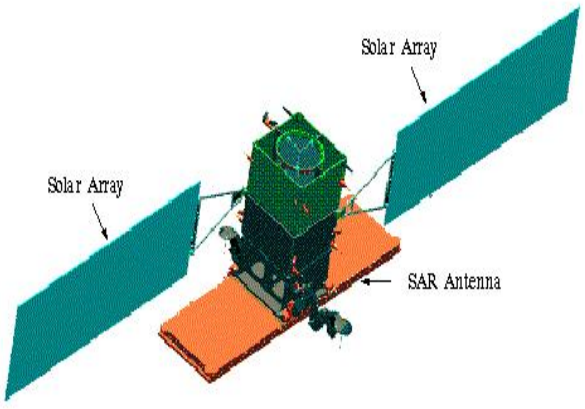


Figure 7.4: Layout of the COSMO-SkyMed satellites (© ASI)

In order to prepare for the exploitation of high resolution images captured by the ORFEO sensors, the ORFEO Accompaniement Programme (refer to the webpage [CNES, 2007b] for latest updates) was set up. Initialized and led by CNES, it was started in mid 2003 and will last until 2009 (Fig.7.2). It consists of a methodological part and a thematic part. The thematic part covers a large range of applications (see first paragraph of this section) which are developed in close cooperation with the end users. Several thematic working groups exist in order to specify and validate value added products and services necessary for a successful operational period. The methodological part's objective is the definition and development

of tools for the achievement of tasks specified by the thematic working groups. The project presented in this report was part of this methodological part and adds functionalities to the ORFEO Toolbox (OTB). OTB is an open source software library for remote sensing imagery processing. It contains a set of algorithms for the operational exploitation of the future submetric radar and optical images (e.g. three dimensional aspects, change detection, texture analysis, pattern matching and optical radar complementarities). It is mainly build around the National Library of Medecine Insight Segmentation and Registration Toolkit (ITK) [Kitware, 2007], an open source software library facilitating the analysis of medical images. The approach presented in this report relies on already developed algorithms of OTB and adds new tools. The proposed image registration algorithm will be integrated into OTB after testing.

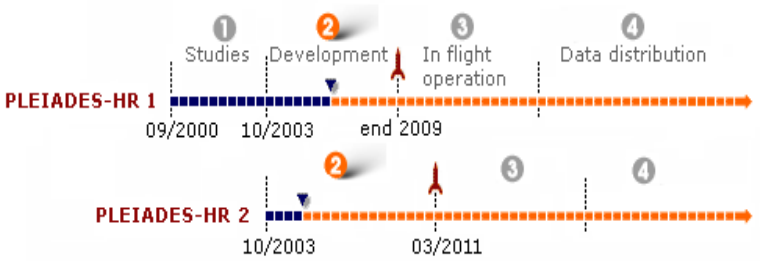


Figure 7.5: Timeline of the Pléiades programme (© CNES)

Annex B

Input statistics and texture parameters for the SVM classification

The following image properties were determined. An image stack was constructed out of these seven image layers.

- Mean:

$$g'_{i,j} = \frac{\sum_{i=i-m}^{i+m} \sum_{j=j-m}^{j+m} g_{i,j}}{n} \tag{7.1}$$

  with the number of rows i, the number of columns j, the filter kernel radius m, the amount of pixels inside the filter kernel n, the current greyvalue of the input image $g_{i,j}$ and the new grey value for the output image $g'_{i,j}$.

- Median:

$$g'_{i,j} = g_{\frac{n+1}{2}} \qquad if\ n\ is\ odd \tag{7.2a}$$

$$g'_{i,j} = \frac{1}{2}\left(g_{\frac{n}{2}} + g_{\frac{n}{2}+1}\right) \qquad if\ n\ is\ even \tag{7.2b}$$

- Entropy:

$$g'_{i,j} = \sum_{i=0}^{2^k-1} P_i log_2 \frac{1}{P_i} \tag{7.3}$$

  with the number of bits per pixel k i.e. each pixel may have a value between 0 and $2^k - 1$. The probability that the pixel of interest has the value i is described with P_i.

- Energy:

$$g'_{i,j} = \sum_{i=0}^{2^k-1} g_i^2 \tag{7.4}$$

- Standard deviation:

$$g'_{i,j} = \sqrt{\frac{\sum_{i=i-m}^{i+m} \sum_{j=j-m}^{j+m} (g_{i,j} - \mu)^2}{n-1}} \tag{7.5}$$

  with the greyvalue mean of the pixels inside the filter kernel μ.

- Skewness:

$$g'_{i,j} = \frac{\sum_{i=i-m}^{i+m} \sum_{j=j-m}^{j+m} (g_{i,j} - \mu)^3}{n} \tag{7.6}$$

- Kurtosis:

$$g'_{i,j} = \frac{\sum_{i=i-m}^{i+m} \sum_{j=j-m}^{j+m} (g_{i,j} - \mu)^4}{n} \tag{7.7}$$

All image layers were linearly rescaled between -1 and +1 since it is the required greyvalue input range for the SVM classification.

Annex C

Supplementary Equations

Rotation matrix of the collinearity equations:

$r_{11} = cos\varphi \cdot cos\kappa$	$r_{12} = -cos\varphi \cdot sin\kappa$	$r_{13} = sin\varphi$
$r_{21} = sin\omega \cdot sin\varphi \cdot cos\kappa + cos\omega \cdot sin\kappa$	$r_{22} = -sin\omega \cdot sin\varphi \cdot sin\kappa + cos\omega \cdot cos\kappa$	$r_{23} = -sin\omega \cdot cos\varphi$
$r_{31} = -cos\omega \cdot sin\varphi \cdot cos\kappa + sin\omega \cdot sin\kappa\cdot$	$r_{32} = cos\omega \cdot sin\varphi \cdot sin\kappa + sin\omega \cdot cos\kappa$	$r_{33} = cos\omega \cdot cos\varphi$

Table 7.1: Elements of the rotation matrix

Corrections to image coordinates due to radial distortion (most significant optical perturbation):

$$\delta r = K_1 \cdot r^3 + K_2 \cdot r^5 + K_3 \cdot r^5, \quad r = \sqrt{(x - x_p)^2 + (y - y_p)^2}$$
$$dx_{dist-r} = \frac{x}{r} \cdot \delta r, \qquad dy_{dist-r} = \frac{y}{r} \cdot \delta r$$

Table 7.2: Radial lens dirstortion with K_i: radial distortion coefficients, r: radial distance, x, y: image coordinates of the point of interest, x_p, y_p: projection of the principal point onto the focal plane (does not necessarily coincide with the center of the CCD array), dx_{dist-r}, dy_{dist-r}: corrections to the image coordinates [Fraser, 2007a]

Corrections to image coordinates due to decentering distortion (a small perturbation):

$$dx_{dist-d} = P_1 \cdot \left(3 \cdot x^2 + y^2\right) + 2 \cdot P_2 \cdot xy, \quad dy_{dist-d} = 2 \cdot P_1 \cdot xy + P_2 \cdot \left(x^2 + 3 \cdot y^2\right)$$

Table 7.3: Decentering dirstortion with P_i: decentering distortion coefficients, x, y: image coordinates of the point of interest, dx_{dist-d}, dy_{dist-d}: corrections to the image coordinates [Fraser, 2007a]

Annex D

Programmed source code

The image fusion strategy proposed in this project was programmed in C++, based on OTB. In the following, the most important source code files are listed. Three different types of source codes have to be distinguished: the files implememting the algorithms (.txx), the corresponding header files (.h) and the test programs (.cxx).

Computation of the SAR amplitude image

- `ImageVectorConvertRescale.cxx`

Geometric Transformations for the Ortho-rectification

Inverse collinearity transformation with three rotation angles:

- `itkCollinearity3DTransform_Inverse.txx`
- `itkCollinearity3DTransform_Inverse.h`
- `itkResampleImageDTMFilter.txx`
- `itkResampleImageDTMFilter.h`
- `transform_collinearity3D_7param_inverse.cxx`

Direct collinearity transformation with rotation matrix elements:

- `itkRatioPolynomial3DTransform_15param.txx`
- `itkRatioPolynomial3DTransform_15param.h`
- `itkResampleImageDTMFilter.txx`
- `itkResampleImageDTMFilter.h`

- transform_ratiopoly_3D_15param.cxx

Inverse collinearity transformation with rotation matrix elements:

- itkRatioPolynomial3DTransform_15param_Inverse.txx
- itkRatioPolynomial3DTransform_15param_Inverse.h
- itkResampleImageDTMFilter.txx
- itkResampleImageDTMFilter.h
- transform_ratiopoly_3D_15param_inverse.cxx

Iterative ortho-rectification for the optical image:

- itkCollinearity3DTransformIterative.txx
- itkCollinearity3DTransformIterative.h
- itkResampleImageDTMIterative.txx
- itkResampleImageDTMIterative.h
- transform_collinearity3D_iterative.cxx

Direct Toutin SAR transformation:

- itkSAR3DTransform_11param.txx
- itkSAR3DTransform_11param.h
- itkResampleImageDTMFilter.txx
- itkResampleImageDTMFilter.h
- transform_SAR_3D_11param.cxx

Inverse Toutin SAR transformation with centered image coordinates:

- itkSAR3DTransform_11param_Inverse_Centered.txx
- itkSAR3DTransform_11param_Inverse_Centered.h
- itkResampleImageDTMFilter.txx
- itkResampleImageDTMFilter.h

- transform_SAR_3D_11param_inverse_Centered.cxx

Iterative ortho-rectification for the SAR image:

- itkSAR3DTransformIterative.txx
- itkSAR3DTransformIterative.h
- itkResampleSARDTMIterative.txx
- itkResampleSARDTMIterative.h
- transform_SAR3D_iterative.cxx

Detection of regions affected by occlusion or layover in the SAR image:

- itkLayoverShadowDetection.txx
- itkLayoverShadowDetection.h
- itkLayoverShadowDetection.cxx

Classification

Classification with SVM:

- metricsSVM.cxx
- Vectorization.cxx
- SVMImageEstimatorClassificationMultiExample_5Classes.cxx
- SVMImageEstimatorClassificationMultiExample_6Classes.cxx

Classification with MRF:

- MarkovFrameworkExampleJan_OPTOrig.cxx
- MarkovFrameworkExampleJan_SAROrig.cxx
- MarkovFrameworkExampleJan_SARFrost.cxx

Feature Extraction

Point detection with the Lopes point detector:

- `LopesImageFilter.txx`
- `LopesImageFilter.h`
- `LopesImageFilter.cxx`

Line detection in the optical image:

- `CannyEdgeDetectionImageFilter.cxx`

Line extraction in the SAR image with the Tupin algorithm:

- `AssymmetricFusionOfLineDetectorExample.cxx`

Distance Mapping with the Danielsson approach:

1. `DanielssonDistanceMapImageFilter.cxx`

Registration

Registrations used in this project:

- `reg_MeanSquares_RotTransl_GradientDescent.cxx` (Program A)
- `reg_NormCorr_RotTransl_GradientDescent.cxx` (Program B)
- `reg_MutInf_RotTransl_OnePlusOne.cxx` (Program C)
- `reg_NormCorr_RotTransl_OnePlusOne.cxx` (Program D)

Tests with the collinearity/Toutin equations and an image pyramid as input:

- `reg_test_MeanSquares_Toutin_GradientDescent.cxx`
- `reg_test_MeanSquares_Toutin10param_GradientDescent.cxx`
- `reg_test_MeanSquares_Collinearity_GradientDescent.cxx`
- `reg_test_NormalizedCorrelation_Collinearity_GradientDescent.cxx`
- `reg_test_MattesMutual_Collinearity_OnePlusOne.cxx`