Signal Processing Techniques for Seafloor Ground-Range Imaging Using Synthetic Aperture Sonar Systems

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Stefan Leier, M.Sc.
geboren am 17.07.1981 in Lich

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr.-Ing. Marius Pesavento
Tag der Einreichung:	17.06.2014
Tag der mündlichen Prüfung:	18.07.2014

D 17
Darmstadt, 2014

Acknowledgments

I would like to thank all the people who have helped, inspired and supported me during my doctoral study.

First and foremost I would like to sincerely thank Prof. Dr.-Ing. Abdelhak M. Zoubir for giving me the chance to pursue my PhD at his institute and offering me the necessary research flexibility. I am grateful you arouse my interest for signal processing and for your entire support throughout my studies. I wish to express my gratitude to Prof. Dr.-Ing. Marius Pesavento for being my co-supervisor as well as to Prof. Dr. rer. nat. Andy Schürr and Prof. Dr.-Ing. Franko Küppers who acted as my examiners in the PhD committee.

Many thanks go to all the current and former colleagues at the Signal Processing Group. It has been a real pleasure to work with you all. Thanks to Dr.-Ing. Christian Debes, Dr.-Ing. Philipp Heidenreich, Dr.-Ing. Marco Moebus, Dr.-Ing. Michael Muma and Dr.-Ing. Fiky Suratman. I am deeply grateful to Dr.-Ing. Raquel Fandos for spending her free time working with me and for her friendship. Special thanks to Nevine Demitri, Sara Al-Sayed, Mark Ryan Balthasar and Michael Leigsnering for proofreading and all the useful comments. Let me know when it is my turn. I also enjoyed working together with Michael Fauss, Jürgen Hahn, Christian Weiss and Feng Yin. Among all students I supervised, I was extremely pleased to work with Martin Kronig, and I always appreciated the support from Renate Koschella and Hauke Fath.

I want to thank all the people at ATLAS ELEKTRONIK GmbH for their support, hospitality during my visits in Bremen and all the useful discussions throughout the last years. In particular, I have benefited a lot from Ursula Hölscher-Höbing's sonar experience and from all the fruitful discussions and comments I got from Dr. Johannes Groen. Hans, it has been a real pleasure to work with you and to learn from someone with such an expertise and knowledge. Furthermore, I would like to thank the people at the CMRE for a great working experience.

A big thank you goes to the "Nationalmannschaft" for the great distraction.

I wish to deeply express my gratitude to my parents, Kurt and Margarethe, for their support, love and their time listening to me without understanding what I was talking about. Many thanks to Nicole, Gilles and Yoann for your advice, help and joy.

Thank you Olga for being there all along and your love.

Darmstadt, 07.08.2014

Kurzfassung

In dieser Dissertation werden fortgeschrittene Verfahren der Signalverarbeitung für bildgebende Sonarsysteme mit synthetischer Apertur behandelt, um hochauflösende Bodenentfernungsaufnahmen des Meeresbodens zu erstellen. Hierdurch werden unter anderem Anwendungen der Objektdetektion, Hydrographie und Pipelineinspektion ermöglicht. Insbesondere werden die Problemstellung der datengestützten Bewegungsschätzung mittels Echosignalen, bekannt als Mikronavigation, sowie Verfahren für die Kompensation von Phasenfehlern betrachtet. Die entwickelte Signalverarbeitungskette zur Rekonstruktion von hochauflösenden Sonarbildern wird daraufhin für eine Empfindlichkeitsstudie verwendet, die den Einfluss der Bildqualität auf ein Detektions- und Klassifikationssystem zur Objekterkennung analysiert. Darüber hinaus wird die Idee des Compressive Sensing eingeführt, um höhere Abdeckungsraten für Systeme mit synthetischer Apertur durch räumliche Unterabtastung in Fahrtrichtung zu erzielen.

Synthetische Aperturverfahren haben gegenüber physikalischen Bildgebungsmethoden den Vorteil, dass durch eine entfernungsunabhängige Bildauflösung Sonarbilder mit hoher Qualität erstellt werden können. Das Verfahren setzt jedoch eine präzise Kenntnis der Positionen der Sender- und Empfangssensorgruppe des Sonarsystems über mehrere aufeinanderfolgende Lotperioden voraus. Aufgrund translatorischer Messungenauigkeiten der Inertialnavigation der Sonar-Trägerplattform bedarf es datengestützter Methoden, um die genaue Trajektorie auf Basis der Empfangssignale zu schätzen. Im Rahmen dieser Dissertation wird eine Signalverarbeitungskette für echte Sonardaten zur Schätzung der Plattformtrajektorie unter Berücksichtigung der Topographie des Meeresbodens entwickelt. Letztere wird anhand eines Schätzverfahrens bestimmt, welches die kontinuierliche Rollbewegung der Trägerplattform des Sonarsystems mit einbezieht. Die Höhenschätzung wird der bildgebenden Methode zusätzlich zur Verfügung gestellt, um eine Bildverschlechterung aufgrund einer unbekannten Topographie im Fall von nichtlinearen Bewegungsabweichungen zu vermeiden. Diese Vorgehensweise führt zu einer deutlichen Verbesserung der Bildqualität in Gebieten, in denen die Höhe des Meeresbodens stark variiert, was anhand von synthetischen Datenbeispielen gezeigt wird. Ein direkter Vergleich von echten Sonarbildern belegt im Anschluss eine deutliche Verbesserung der Qualität durch die Verwendung der entwickelten Verarbeitungskette.

Des Weiteren werden praktische Methoden innerhalb der Mikronavigation zur erwartungstreuen Schätzung der Plattformtrajektorie vorgeschlagen. Diese umfassen sowohl ein Kompensationsverfahren zur Korrektur von verzerrten Zeitversatzschätzungen als auch ein querab Beamforming-Verfahren. Während das Kompensationsverfahren

verzerrte Zeitversatzschätzungen korrigiert, die durch ein großes Verhältnis von Trägerfrequenz zu Bandbreite entstehen, gleicht das Beamforming-Verfahren die variierenden Zeitunterschiede von Signalen aus, die durch Nahfeldszenarien und Systeme mit breiter Abstrahlcharakteristik auftreten. Um Bildverschlechterungen aufgrund von verbleibenden Phasenfehlern zu korrigieren, wird eine bestehende Autofokusmethode so modifiziert, dass die räumlich-veränderlichen Punktspreizfunktionen eines idealen Bildpunkts, wie sie im Stripmap-Operationsmodus für synthetische Aperturverfahren auftreten, besser berücksichtigt werden. Weiterhin wird ein Kalibrierverfahren zur Schätzung einer optimalen mittleren Schallgeschwindigkeit vorgeschlagen, um eine bestmögliche Fokussierung des Sonarbildes für den Fall einer ortsabhängigen Schallgeschwindigkeit zu erzielen.

Obwohl die datengestützte Bewegungsschätzung und Phasenfehlerkompensation die bestmögliche Bildqualität erreichen sollen, ist die Untersuchung des Einflusses der Bildqualität auf die automatische Objekterkennung äußerst wichtig. Dies gilt insbesondere um die Zuverlässigkeit von Detektions- und Klassifikationssystemen für zukünftige autonome Sonarsysteme beurteilen zu können. Die durchgeführte Empfindlichkeitsstudie zeigt den enormen Leistungsverlust in der Bildsegmentierung, der Merkmalsextraktion, sowie in der Klassifikation für die automatische Objekterkennung. Weiterhin wird ein empirischer Zusammenhang zwischen der Bildqualität und der Leistungsfähigkeit des Detektions- und Klassifikationssystems verdeutlicht. Um verlässliche Eingangsgrößen für die Objekterkennung sicherzustellen, wird eine Strategie zur sequenziellen Beurteilung der Qualität von Sonarbildern anhand der momentanen geometrischen Auflösung der synthetischen Apertur vorgeschlagen und erfolgreich umgesetzt.

Eine Erhöhung der Abdeckungsrate für konventionelle Verfahren mit synthetischer Apertur ist ausschließlich durch eine Vergrößerung der physikalischen Aperturlänge möglich. Als Alternative wird die Verwendung von Compressive Sensing vorgeschlagen, welches eine räumliche und zeitliche Unterabtastung bei gleichbleibender Bildqualität ermöglicht. Hierdurch können höhere Plattformgeschwindigkeiten erzielt werden. Das entwickelte Bildgebungsverfahren verhindert das Auftreten von Bildambiguitäten durch die Unterabtastung und zeigt ein enormes Potenzial zur Datenreduktion anhand von Simulationen mit synthetischen Daten auf. Gleichzeitig bestätigt ein experimentelles Laborsystem eine mögliche Verdopplung der Plattformgeschwindigkeit, was einen ersten Schritt in Richtung der Reduktion der Missionsdauer für zukünftige praktische Systeme darstellt. Alle in dieser Arbeit entwickelten Verfahren werden sowohl auf synthetisch generierten Daten als auch auf realen Sonarmessdaten erfolgreich angewandt.

Abstract

In this Ph.D. thesis advanced signal processing techniques are addressed in order to reconstruct high-resolution seafloor imagery using synthetic aperture sonar ground-range imaging. This enables applications such as object detection, hydrography, and pipeline inspection, among others. In particular, the problems of echo-data-driven motion estimation known as micronavigation, and compensation of phase errors are considered. Based on the developed processing chain, a sensitivity study is conducted that points out the impact of distorted seafloor imagery on an automatic detection and classification system for target recognition. Furthermore, the framework of compressive sensing is introduced for synthetic aperture imaging to attain higher coverage rates using along-track undersampling.

Synthetic aperture techniques are advantageous over conventional real aperture imaging techniques as they achieve a range-independent resolution that enables the reconstruction of high-quality sonar imagery. However, this requires an accurate knowledge of the positions of the transmitter and the receiving elements for multiple consecutive transmission and reception times. As inertial navigation systems are imprecise in tracking translational platform motion, echo-data-driven motion estimation is additionally employed to estimate the platform trajectory. For this purpose, a motion estimation processing chain for real sonar measurements is designed in this thesis, which considers height information about the seafloor. To this end, a topography estimation technique, which takes into account a continuous roll movement of the imaging platform, is developed. In order to avoid image defocus in environments with a strongly varying topography in case of nonlinear trajectories, the obtained height estimates are used during image reconstruction. This leads to significant image quality improvements, which is demonstrated on synthetic data. A comparison of real synthetic aperture sonar images then highlights the quality enhancement using the developed processing chain.

Furthermore, practical methods for an unbiased estimation of platform motion are proposed. These involve a compensation technique for correcting the occurrence of biased time delay estimates due to a high ratio of carrier frequency to bandwidth as well as a broadside beamforming method that equalizes the varying time delays due to near-field scenarios and widebeam systems. In order to avoid image defocusing in the presence of residual phase errors, modifications of an existing autofocus technique are proposed to better cope with spatially varying point spread functions in stripmap synthetic aperture imaging. Additionally, a data-driven calibration method

is developed so as to correctly estimate an optimal average sound speed yielding the best focused imagery in situations of a spatially varying sound-speed profile.

Although data-driven motion estimation and phase error compensation techniques aim at achieving high-quality imagery of the seafloor, investigating the influence of distorted sonar imagery on automatic target recognition systems is of utmost importance for future autonomous sonar systems in order to judge their reliability. A sensitivity study is conducted that demonstrates significant performance loss in image segmentation, feature quality as well as in classification performance of a specific automatic detection and classification system. Further, an empirical relation between image degradation and performance loss of the individual stages of the automatic detection and classification system is highlighted. In order to guarantee reliable inputs for automatic target recognition, a strategy is proposed to sequentially assess the image quality of synthetic aperture sonar imagery during reconstruction, which is based on the instantaneous cross-range resolution.

Increasing the coverage rate of a synthetic aperture system in the case of conventional imaging techniques is only feasible by increasing the physical array size. Alternatively, a compressive sensing framework is applied to perform aperture undersampling and, thus, offer a higher platform speed while still maintaining imaging performance. A stripmap imaging technique is developed to avoid the occurrence of azimuth image ambiguities. Synthetic data simulations then demonstrate the huge potential in data reduction, and laboratory experiments using compressive sensing further show an increase in platform speed by a factor of two. This possibly reduces the overall mission time of future synthetic aperture systems in real-life scenarios. All developed methods are applied to synthetic data as well as to real sonar measurements.

Contents

Chapter 1

Introduction

Synthetic aperture sonar (SAS) provides high-resolution imagery of the seafloor which, in contrast to conventional side-looking sonar systems, achieves a range-invariant cross-range resolution [Hayes and Gough (2009), Hansen et al. (2011)], thereby maintaining a constant resolution of the SAS image for the entire seafloor scenery. This is highly desirable for any post-processing application, *e.g.*, object detection [Midelfart et al. (2009), Williams and Groen (2011), Fandos and Zoubir (2011)], mine hunting [Groen et al. (2010), Fandos et al. (2013)], seafloor mapping, and hydrography [McRea et al. (1999)]. Further, phase difference extraction of high-quality SAS image pairs, reconstructed by an interferometric SAS configuration, yields an accurate estimate of the seafloor topography [Sæbø et al. (2007)]. This enables potential applications such as pipeline inspection [Hansen et al. (2010)] or reconnaissance for installations of pipelines and offshore platforms [Bjørnø (2013)]. Ultimately, all aforementioned application tasks are performed autonomously without the need for permanent operator control. In order to maintain resolution during imaging, a synthetic aperture of variable length is formed by coherently processing a varying number of backscattered echo signals. These signals are recorded over consecutive transmission times by a physical hydrophone array that is either mounted onto a moving autonomous underwater vehicle (AUV) or employed as a towed array [Hayes and Gough (2009)]. The challenge of synthetic aperture formation is its high sensitivity (up to a fraction of a wavelength) to unknown platform motion, which arises from an imprecise on-board inertial navigation system (INS) [Hansen et al. (2011)]. As a consequence, image defocusing may occur, which is in conflict with the aim of achieving high-quality SAS imagery. Thus, additional means are required to obtain precise estimates of the platform motion in order to enable the reconstruction of high-quality SAS images.

1.1 Motivation

Although SAS imaging has become a mature seafloor mapping technique that produces images with remarkable quality under friendly environmental conditions, severe image defocusing still occurs in harsh or difficult situations, *e.g.*, for strongly varying seafloor topographies, in shallow water, and under varying sound speed conditions

[Hansen et al. (2011)]. While motion estimation and compensation, also known as micronavigation [Bellettini and Pinto (2002)], is limited in shallow water due to strong multipath effects [Bellettini and Pinto (2009)], seafloor height estimates are necessary for ground-plane micronavigation in environments with a rough topography. Otherwise, the unknown motion is estimated inaccurately. Particularly, in ground-plane imaging using space-time reconstruction techniques, spatial height grids are required in addition to ground-plane motion estimates to attain high-resolution SAS images. Ground-plane or ground-range imaging is typically more suitable for SAS interferometry than slant-range imaging [Sæbø et al. (2007)]. In widebeam systems [Hayes and Gough (2009), Callow et al. (2009)], the accuracy requirement of estimating platform motion becomes even more stringent as more echo signals have to be processed coherently. Moreover, the broadside narrowbeam assumption regarding micronavigation is no longer valid, which demands additional processing steps to avoid biased motion estimates. However, widebeam SAS systems are favored as they attain the same resolution using lower frequencies, which are less attenuated in water [Lurton (2002)]. This makes widebeam SAS systems suitable for long-range imaging applications.

In any sonar imaging application, image reconstruction is usually the first stage of the processing chain after collecting raw sonar measurements. The reconstructed images serve then as an input for an application specific post-processing stage, *e.g.*, an automatic target recognition (ATR) system. Most existing ATR systems that operate on SAS images assume a uniformly excellent quality for the entire image [Williams and Groen (2011)]. However, for the aforementioned reasons, this may not always be fulfilled in practice. Therefore, the question arises, how ATR performance is affected by distorted image quality scenarios. In current mine hunting applications, sonar images are still inspected by an operator [Zerr et al. (2009)] so that real-time intervention is possible. Omitting the operator therefore requires an assessment scheme to guarantee the use of high-quality SAS images in future ATR systems.

The advance per ping of current SAS systems is dictated by the spatial sampling theorem [Soumekh (1999)], thereby, restricting their practical coverage rates [Hagen and Hansen (2007)]. A violation of the theorem leads to azimuth image ambiguities resulting in either masked image content such as objects, or destroyed shadow information that is important for ATR. Compressive sensing (CS) [Donoho (2006), Candès et al. (2006)] describes a new sampling paradigm that enables sub-Nyquist sampling. As a consequence, SAS imaging techniques based on the CS framework have the potential to increase the coverage rate, and hence, reduce the mission time while maintaining image quality.

1.2 State-of-the-art

Current SAS systems consist of a single-transmitter and multi-receiver configuration [Hayes and Gough (2009)], which enables the exploitation of the redundant phase center (RPC) principle via the displaced phase center antenna (DPCA) algorithm [Bellettini and Pinto (2002), Cook et al. (2006)]. This micronavigation technique estimates unknown path deviations of the imaging platform from a predefined rectilinear trajectory. Depending on the imaging modality, *i.e.*, slant-range or ground-range, either a range difference known as slant-range sway [Bellettini and Pinto (2002), Bellettini and Pinto (2009)] or a ping-to-ping motion vector is estimated with respect to a global Cartesian coordinate system [Cook et al. (2006)]. Furthermore, the type of imaging category, *i.e.*, wavenumber-frequency or space-time reconstruction, plays a major role in how motion compensation is applied [Callow et al. (2009)]. Some systems also use a combination of both imaging types. While large-scale imaging is handled by wavenumber-frequency imaging given its computational advantages, space-time imaging is employed to construct small-scale images as it provides better resolution capabilities by using a more accurate motion compensation [Hansen et al. (2011)]. However, given the close links of SAS systems to non-civilian applications, a complete and especially detailed processing overview of state-of-the-art SAS systems is usually not available, except for [Cook (2007)], which considers a SAS processing chain in reasonable detail. In [Sæbø (2010), Schmaljohann and Groen (2012)], the influence of seafloor topography on motion estimation is mentioned but no details are provided. The same applies to an additional use of autofocus techniques, which have the potential to further enhance SAS image quality. In [Hansen et al. (2011)], a modified phase gradient autofocus (PGA) algorithm is employed to improve image quality. The same technique has been used to estimate sound speed [Hansen et al. (2007)]. However, neither the mosaic phase gradient (mPGA) proposed in [Bonifant (1999)] nor the stripmap PGA proposed by [Callow et al. (2003a)] are documented to have been employed in practical systems on real SAS images except for a single example in the cited references, to the best of knowledge. Modifications to the mPGA are mentioned in [Zhu et al. (2011)] but lack a thorough description and validation.

State-of-the-art ATR systems operate on SAS images [Midelfart et al. (2009), Groen et al. (2009), Fandos et al. (2013)], as they provide superior resolution in contrast to conventional real aperture sonar (RAS) images. Thus, higher classification performance is attained due to a gain in the richness of detail. However, SAS image quality has to be sufficiently high in order to obtain better segmentation results along with extracting more suitable features and, ultimately, improved classification performance. Distorted SAS images affect object and shadow segmentation, and

thereby, the feature extraction process of the ATR system. For example, geometrical shape information of a shadow changes, which is usually associated with the shape of the object. Thus, the classification performance deteriorates. To the best of knowledge, a study on the degradation of classification performance under residual platform motion has not yet been conducted. In order to avoid loss in ATR performance, state-of-the-art systems are based on adaptive survey planning [Williams (2012)] by estimating a correlation quality measure as a function of range. Typically, this measure arises as a side product during motion compensation [Williams et al. (2012)] or from real aperture interferometry [Synnes et al. (2009)]. For low quality values, the spacing between parallel tracks of the AUV is reduced to guarantee the collection of good quality sonar data over range. Both correlation measures are highly related to SAS image quality, as they are a prerequisite for successful motion estimation. However, a high correlation measure does not imply good SAS image quality as there are potential scenarios, *e.g.*, mismatches in sound speed or difficult seafloor topographies, which may yield distorted SAS imagery [Hansen et al. (2011)].

1.3 Contributions

In the sequel, original contributions to the field of SAS signal processing are summarized. The thesis provides a complete description of a developed processing chain to reconstruct high-resolution ground-range SAS images given raw sonar measurements. While SAS has become a mature imaging technique under mild operational conditions, image quality is still distorted in difficult environmental situations. The following novel techniques are integrated into the SAS processing chain to improve image quality and are addressed in Chapter 4:

- Phase wrap error correction: A high ratio of carrier frequency to bandwidth introduces biased time delay estimates due to phase cycle ambiguities. A scheme based on binary image processing techniques is proposed to account for these biased estimates.
- Near-field and widebeam correction: Motion compensation of widebeam SAS systems is affected by biased time delays due to non-broadside targets located in the near-field. The effect is theoretically studied for RPC arrays and a broadside beamforming scheme is proposed to mitigate it.
- Height estimation for motion compensation: Neglecting the seafloor topography may lead to inaccurate ping-to-ping motion estimates. The topography influence

on image quality is studied and a processing scheme is developed to estimate the seafloor height variation for motion compensation. Further, the height estimates can be incorporated into the image reconstruction method in an ensuing processing iteration.

- Stripmap autofocus extension: In contrast to spotlight autofocus algorithms, original stripmap autofocus algorithms do not operate solely on image data but also require raw sonar measurements to iteratively estimate the unknown platform path. As this is computationally expensive, spotlight autofocus algorithms are employed on stripmap SAS images assuming a locally invariant point spread function (PSF). Modifications to an existing algorithm (mPGA) are introduced to overcome its drawbacks and make it more applicable for stripmap SAS images. A validation is demonstrated on synthetic and real sonar data.

- Sound speed estimation: A mismatch in the assumed sound speed leads to image defocusing during reconstruction. Since water is an inhomogeneous medium, sound speed may vary along the acoustical pathway. However, sound speed can only be easily measured in the vicinity of the imaging platform. Thus, a technique for estimating an optimal average sound speed is developed based on an iterative image quality assessment scheme.

Based on the proposed SAS processing chain, the influence of image quality on an ATR system is studied for the first time and a novel strategy for sequentially assessing image quality is developed in Chapter 5:

- Image quality impact on automatic target recognition: For man-made object detection, ATR systems typically assume SAS images to be of a homogeneously good quality. An empirical analysis is carried out to study the influence of residual motion errors on the performance of individual ATR processing steps such as segmentation, feature extraction, and classification performance. An empirical relation to image degradation is demonstrated.

- Sequential focus assessment: An evaluation strategy is proposed to sequentially assess the focusing capability of the synthetic aperture imaging technique, and therefore, to judge the quality of the sonar images. The developed scheme can be used to determine the number of single ping images for which coherent processing leads to an improvement in resolution and overall image quality.

The framework of CS is considered as a new methodology for future SAS imaging systems to overcome the spatial sampling limitations of conventional synthetic aperture systems:

- A reconstruction scheme based on CS is proposed as an imaging technique for synthetic apertures that allows for aperture undersampling. Simultaneously, the proposed technique suppresses occurring azimuth image ambiguities, which can mistakenly be interpreted as real targets. An experimental ultrasound synthetic aperture system has been set up to record real data measurements and to verify the imaging results obtained by synthetic data experiments.

1.4 Publications

The following publications have been produced during the period of doctoral candidacy.

Internationally refereed journal articles

- S. Leier and A. M. Zoubir, "Aperture undersampling using compressive sensing for synthetic aperture stripmap imaging," *EURASIP J. Advances Signal Process.*, Mar. 2014, under review.
- S. Leier, R. Fandos and A. M. Zoubir, "Motion error influence on segmentation and classification performance in SAS based automatic mine countermeasures," *IEEE J. Ocean. Eng.*, Mar. 2014, early access.

Internationally refereed conference papers

- S. Leier, M. Kronig and A. M. Zoubir, "A modified version of the mosaic phase gradient autofocus," in *Proc. Eur. Signal Process. Conf.*, Marrakech, Morocco, Sep. 2013.
- S. Leier, A. M. Zoubir and J. Groen, "Sequential focus evaluation of synthetic aperture sonar images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Vancouver, Canada, May 2013.
- C. Debes, S. Leier, F. Nikolay and A. M. Zoubir, "Compressive sensing for synthetic aperture imaging using a sparse basis transform," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Munich, Germany, Jul. 2012.
- S. Leier and A. M. Zoubir, "Phase wrap error correction for micronavigation in synthetic aperture systems," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Munich, Germany, Jul. 2012

- S. Leier and A. M. Zoubir, "Time delay estimation for motion compensation and bathymetry of SAS systems," in *Proc. Eur. Signal Process. Conf.*, Bucharest, Romania, Aug. 2012.

International conference papers

- M. R. Balthasar, S. Al-Sayed, S. Leier and A. M. Zoubir, "Optimal area coverage in autonomous sensor networks," in *Proc. 2nd Int. Conf. Underwater Acoust.*, Rhodes, Greece (invited paper), Jun. 2014.
- S. Leier, J. Groen, A. M. Zoubir, U. Hölscher and I. Campbell, "The influence of sound speed on synthetic aperture sonar imagery," in *Proc. 1st Int. Conf. Underwater Acoust.*, Korfu, Greece (invited paper), Jun. 2013.
- S. Leier and A. M. Zoubir, "Quality assessment of synthetic aperture sonar images based on a single ping reference," in *Proc. IEEE OCEANS Europe Conf.*, Santander, Spain, Jun. 2011.

Filed patent application

- S. Leier, A. M. Zoubir, and J. Groen, (Technische Universität Darmstadt, ATLAS ELEKTRONIK GmbH, Applicants) "Verfahren zum Bestimmen einer optimalen Unterwasserschallgeschwindigkeit sowie Vorrichtung zum Durchführen des Verfahrens," Patent Application DE 10 2013/106 359.4.

1.5 Thesis overview

The thesis outline is as follows: Chapter 2 describes the fundamentals of synthetic aperture processing. It introduces a signal model of a mono-static synthetic aperture system and addresses resolution aspects as well as sampling constraints.

Chapter 3 describes two real SAS systems from ATLAS ELEKTRONIK GmbH, Bremen, Germany, and extends the established signal model to a single-transmitter and multi-receiver system. It further addresses the concepts of space-time imaging and provides a description of a proposed ground-grid construction scheme to facilitate coherent RAS image summation. Synthetic aperture shading is introduced as a

technique for controlling the sidelobe level (SLL) of a PSF, and broadside bathymetry estimation is discussed to obtain seafloor topography information.

In Chapter 4, a data-driven motion compensation method based on the received echo signals is developed after introducing a motion error model for stripmap SAS imaging. The proposed processing chain for motion compensation mainly consists of surge estimation, time delay estimation, and a model-fitting approach for estimating ground-range sway as well as heave motion. In this context, a phase wrap compensation method is proposed to compensate for biased time delays and successfully applied to experimental sonar data. A technique for correcting near-field and widebeam effects is developed based on broadside beamforming after deriving the relation between time delays and the corresponding displacement between RPC arrays. Based on the introduced nonlinear least squares approach used for ping-to-ping motion estimation, real SAS image examples are shown, highlighting the focusing enhancement using data-driven motion compensation. Afterwards, an existing autofocus technique is modified to overcome two original shortcomings, followed by a proposed scheme for sound speed estimation based on quality evaluation of SAS images.

Chapter 5 first briefly introduces the fundamental steps of an ATR system, namely, segmentation, feature extraction, and classification. After the description of the used test database, the impact of residual motion on the individual ATR processing blocks is empirically studied and discussed using different motion error amplitudes. A relation to image quality is demonstrated. Thereupon, a sequential focusing strategy is proposed to assess SAS image quality by probing the instantaneous cross-range resolution during image reconstruction.

In Chapter 6, a synthetic aperture imaging method based on a CS framework is developed allowing for aperture undersampling while suppressing azimuth image ambiguities that occur for conventional imaging. The proposed technique is verified based on synthetic data as well as on ultrasound measurements collected by a real laboratory SAS rail system. The latter has been set up during the period of doctoral candidacy.

Conclusions are drawn in Chapter 7 and recommendations for future work are presented.

Chapter 2
Fundamentals

The principle of active synthetic aperture techniques is to synthesize a large aperture by moving a small physical aperture, which constantly transmits pulses towards a scene of interest. While a single transceiver is typically employed in synthetic aperture radar (SAR) [Carrara et al. (1995), Jakowatz et al. (1996), Soumekh (1999)], the use of a single-transmitter and multi-receiver array is more common in synthetic aperture sonar (SAS) imaging [Hayes and Gough (2009), Hansen et al. (2011)] due to the relatively slow sound speed in water in contrast to the speed of light in air leading to a slow platform speed [Cutrona (1975)]. The physical aperture is mounted onto an imaging platform, *e.g.*, an airplane or satellite in SAR or an autonomous underwater vehicle (AUV) in SAS imaging.

After recording the backscattered echo signals along the entire synthetic aperture, a SAR or SAS image can be reconstructed in terms of a reflectivity map of the illuminated target area. As these images provide a resolution that is theoretically constant for the entire scene due to a dynamic adjustment of the synthetic aperture length, synthetic aperture imaging is usually advantageous over real aperture imaging. In general, different synthetic aperture imaging modes are available, namely, spotlight and stripmap [Carrara et al. (1995), Soumekh (1999)]. For the former, the antenna is either mechanically or electronically steerable via phased arrays in along-track direction to illuminate a single stationary spot of the target area during the entire data acquisition. Contrarily, in stripmap operation, the antenna is steadily pointing in broadside direction illuminating a strip of the target area that changes during acquisition. For the remainder of the thesis, stripmap mode imaging is considered as it is most common for SAS [Hayes and Gough (2009), Hansen et al. (2011)].

In this chapter, an overview of the key ideas for synthetic aperture signal processing is presented and important parameters such as the resolution and sampling constraints are introduced. Section 2.1 states the signal model of a synthetic aperture transceiver system taking into account demodulation aspects and pulse compression. The latter is typically used as a technique in active radar and sonar systems to improve range resolution. In Section 2.2 the basic principle of synthetic aperture imaging is addressed, namely, its dynamic length adjustment so as to achieve a constant cross-range resolution. Section 2.3 then covers spatial sampling requirements and points out its relation to the area coverage rate and its impact on SAS imaging.

2.1 Synthetic aperture signal model

In its basic configuration, a synthetic aperture is formed by a moving transceiver that repeatedly sends a pulsed signal $s^{\mathrm{BP}}(t)$ at time instants t_p from its current position $\boldsymbol{a}_p = [0, p\,\Delta^{\mathrm{A}}, h^{\mathrm{og}}]^T$, where p is the ping index of transmission time t_p, Δ^{A} denotes the advance per ping (APP) and h^{og} is the altitude of the imaging system. Simultaneously, the reflected echo signals $e_p^{\mathrm{BP}}(t)$ are recorded, where the superscript $(\cdot)^{\mathrm{BP}}$ indicates bandpass domain signals. Typically, the continuous process of moving along the trajectory is neglected in synthetic aperture signal models by assuming a so-called stop-and-hop situation [Richards (2005)]. Here, the sensor element remains at its current position $\boldsymbol{a}_p$ after signal transmission and waits for the reception of echo signals. The sensor then hops with a step size Δ^{A} to its next position $\boldsymbol{a}_{p+1}$. Whether this assumption is applicable mainly depends on the relation between platform speed v with respect to the speed of wave propagation c and the maximum imaging range $R_{\max}$. While in SAR applications, especially in air-borne systems, the condition is typically fulfilled, it may or may not be valid in SAS imaging. However, given the real sonar system parameters as provided in Section 3.1, the stop-and-hop model can be assumed for the remainder of this thesis without any consequences on SAS image quality.

A typical geometry of a synthetic aperture imaging system operating in stripmap mode is depicted in Figure 2.1, where the traveling direction of the imaging platform and the direction of wave propagation are called cross-range y (along-track) and range (or slant-range), respectively. The latter has to be distinguished from the ground-range dimension x. In general, the imaging swath of the target area is unbounded, which is illustrated by gray dashed lines. However, the area, for which the synthetic aperture image is reconstructed, is bounded in ground-range and cross-range direction by $x \in [X_c - X_0, X_c + X_0]$ and $y \in [Y_c - Y_0, Y_c + Y_0]$, respectively. Here, X_c and Y_c denote the center of the target area in ground-range and cross-range direction, respectively, and X_0 and Y_0 represent half of the swath width. While the boundaries in the ground-range direction are mainly determined by the signal-to-noise ratio (SNR) and the geometry of the imaging system, the boundaries in cross-range are dictated by the physical aperture position along the vehicle trajectory and its horizontal beamwidth. Thus, the boundaries specify the area for which a constant resolution is attained by synthetic aperture imaging as highlighted by the red dashed area in Figure 2.1.

In general, the target area $f(x, y)$ can be modeled as a set of D stationary point targets with reflectivity σ_d and location $\boldsymbol{q}_d = [x_d, y_d, 0]^T,\ d = 1, \ldots, D$, considering a flat bottom scenario. To simplify matters, the reflectivity σ_d is assumed to be independent of frequency and incidence angle of the impinging wave. Moreover, any path losses are

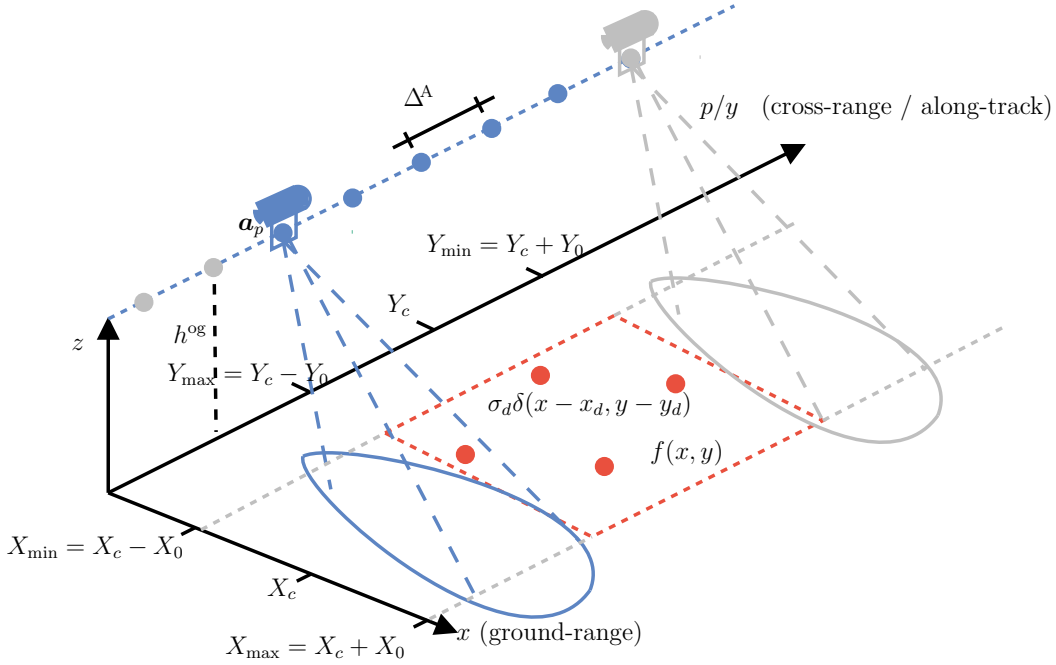


Figure 2.1: Stripmap mode geometry of a synthetic aperture system consisting of a transceiver at location $\boldsymbol{a}_p$ that records the echo signals scattering back from the target scene $f(x, y)$ during ping p. The scene consists of point targets with reflectivity σ_d. In general, the target scene is unbounded, but the synthetic aperture image is bounded in range direction by $[X_c - X_0, X_c + X_0]$ and in cross-range direction by $[Y_c - Y_0, Y_c + Y_0]$.

incorporated into σ_d. The ideal reflectivity function [Soumekh (1999)] of the target area can then be expressed as

$$f(x, y) = \sum_{d=1}^{D} \sigma_d \, \delta(x - x_d, y - y_d), \tag{2.1}$$

where $\delta(x, y)$ describes the two-dimensional Dirac function of ground-range direction x and cross-range direction y. The distance $r_{d,p}$ between a target at location $\boldsymbol{q}_d$ and the imaging platform at position $\boldsymbol{a}_p$ during ping p is given by

$$r_{d,p} = \left\| \boldsymbol{q}_d - \boldsymbol{a}_p \right\|_2 , \tag{2.2}$$

yielding a round-trip delay of $\tau_{d,p} = 2r_{d,p}/c$. In (2.2), $\|\cdot\|_2$ denotes the Euclidean norm. The echo signals for a total number of M_p pings of a mono-static synthetic aperture system under the stop-and-hop assumption [Richards (2005)] are then modeled by the superposition of individually weighted and delayed versions of the transmitted pulse $s^{\mathrm{BP}}(t)$ as follows

$$e_p^{\mathrm{BP}}(t) = \sum_{d=1}^{D} \sigma_d \, b_{\mathrm{phy}}(\theta_{d,p}) \, s^{\mathrm{BP}}\left(t - \tau_{d,p}\right) + v_p^{\mathrm{BP}}(t) \tag{2.3}$$

$$\text{with} \quad p = 0, \ldots, M_p - 1 \quad \text{and} \quad t \in [0, 2R_{\max}/c + T_{\Pi}].$$

Here, $v_p^{\mathrm{BP}}(t)$ describes an additive white noise process, T_{Π} is the pulse length of the transmitted signal $s^{\mathrm{BP}}(t)$, and c denotes the speed of wave propagation in the medium, *e.g.*, speed of light or sound speed. Moreover, the function $b_{\mathrm{phy}}(\theta)$ describes an indicator function resembling an ideal beampattern of the transceiver that is flat over the supported beamwidth and independent of frequency. Given the azimuth aspect angle $\theta_{d,p}$ between target location $\boldsymbol{q}_d$ and transceiver position $\boldsymbol{a}_p$ as

$$\theta_{d,p} = \arctan\left(\frac{y_d - p\,\Delta^{\mathrm{A}}}{x_d}\right), \tag{2.4}$$

the ideal beampattern function $b_{\mathrm{phy}}(\theta_{d,p})$ determines whether target d is observed by the transceiver at ping p. This is written as follows

$$b_{\mathrm{phy}}(\theta) = \begin{cases} 1 & |\theta| \leq \theta_{\mathrm{BW}}/2 \\ 0 & \text{elsewhere,} \end{cases} \tag{2.5}$$

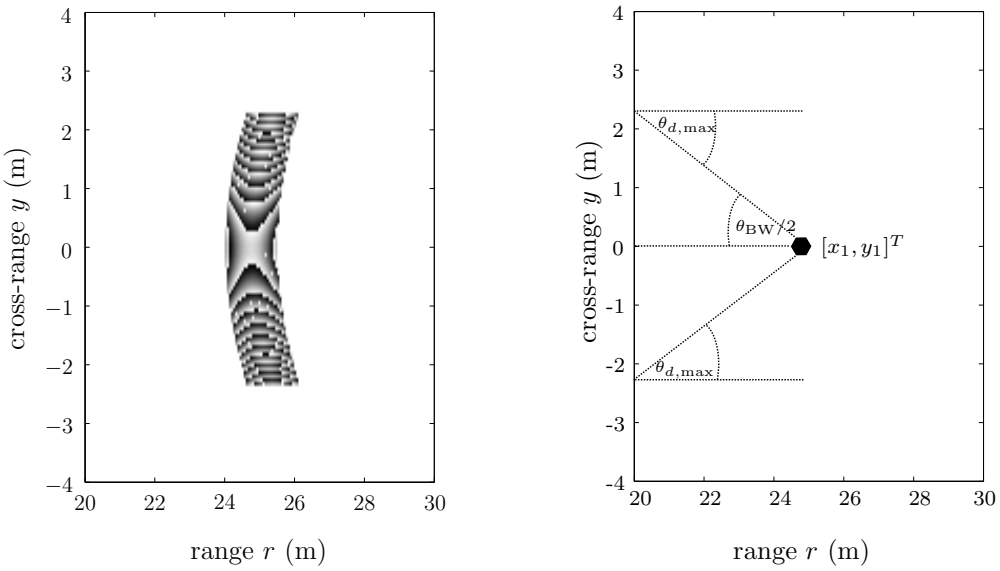
where θ_{BW} is the angular beamwidth of the physical sensor. As mentioned before, the echo signals in (2.3) represent a bandpass model in the continuous-time domain. However, sophisticated signal processing methods are usually applied in the discrete-time domain. Thus, an adequate model is required, which is obtained by demodulation and analog-to-digital conversion (ADC) of the received echo signals as

$$e_p^{\mathrm{LP}}(n) = e_p^{\mathrm{BP}}(t)\, e^{-j\Omega_c t}\Big|_{t=nT_s}, \tag{2.6}$$

where Ω_c denotes the continuous-time angular carrier frequency, T_s is the sampling interval of the ADC, and the superscript $(\cdot)^{\mathrm{LP}}$ indicates lowpass domain signals. The baseband model, also known as phase history data [Jakowatz et al. (1996)], in the discrete-time domain is then obtained by substituting (2.3) into (2.6), which leads to

$$\begin{aligned} e_p^{\mathrm{LP}}(n) &= \left(\sum_{d=1}^{D} \sigma_d\, b_{\mathrm{phy}}(\theta_{d,p})\, s^{\mathrm{BP}}(nT_s - \tau_{d,p}) e^{j\Omega_c(nT_s-\tau_{d,p})} + v_p^{\mathrm{BP}}(nT_s)\right) e^{-j\Omega_c nT_s} \\ &=: \sum_{d=1}^{D} \sigma_d\, b_{\mathrm{phy}}(\theta_{d,p})\, s^{\mathrm{LP}}(n-\eta_{d,p})\, e^{-j\omega_c\eta_{d,p}} + v_p^{\mathrm{LP}}(n), \\ &\text{with}\;\; n = 0,\ldots,M_n-1 \;\;\text{and}\;\; p = 0,\ldots,M_p-1. \end{aligned} \tag{2.7}$$

Here, $\eta_{d,p} = \tau_{d,p}/T_s$ denotes the scaled time delay, and $\omega_c = \Omega_c T_s$ is the angular center frequency. Moreover, $s^{\mathrm{LP}}(n)$ and $v_p^{\mathrm{LP}}(n)$ denote the equivalent lowpass representation of the transmitted signal and additive noise process in discrete-time, respectively. Both superscripts are dropped for notational convenience later on. Note that η is in general a non-integer variable. The phase history data in (2.7) highlights that the round-trip delay is contained in the envelope function of the transmitted baseband signal as well as in a carrier phase term. The latter is exploited in Section 4.2.3 to estimate time delay differences with subsample precision.



(a) Phase response

(b) Single point target geometry

Figure 2.2: Phase of the received echo signals (a) that are recorded along the synthetic aperture resulting from a point target in the scene of interest as shown in (b). The phase values of a linear frequency modulated pulse used as transmitted signal in (a) are depicted between $-\pi$ (light-gray color coding) and π (black color coding). The scenario in (b) illustrates that the maximum target aspect angle $\theta_{d,\max}$ equals half the angular beamwidth $\theta_{\mathrm{BW}}/2$.

A noise-free example of phase history data of a single point target is illustrated in Figure 2.2a using the relations $y = p\,\Delta^{\mathrm{A}}$ with $p = 0, \ldots, M_p - 1$ in cross-range direction and $r = nT_s\,c/2$ with $n = 0, \ldots, M_n - 1$ in range direction. Here, M_p and M_n denote the number of slow-time positions (number of total pings) and fast-time samples, respectively. The terms slow-time and fast-time originate from the slow speed of the vehicle motion relative to the fast speed of the propagating wave. Note that the phase history data is given in range (or slant-range) direction r, rather than in ground-range direction x. Figure 2.2b shows the corresponding scene of interest containing a single point target. While the angular beamwidth θ_{BW} of the physical aperture limits the target signature in cross-range direction, the width of the target signature in range direction is determined by the pulse length T_{Π} of the transmitted pulse $s^{\mathrm{BP}}(t)$. In Figure 2.2a, the phase response shows a quadratic behavior due to the linear frequency modulated (LFM) pulse form, which has been used to generate the echo signals. In the following section, more details are provided about the characteristics of the transmitted pulse and the required pre-processing steps to improve range resolution.

2.1.1 Pulse compression via matched-filtering

Active radar and sonar imaging systems often use an LFM pulse for the transmitted signal $s^{\mathrm{BP}}(t)$ due to its properties with respect to range resolution as well as its sufficient Doppler shift tolerance during pulse compression [Richards (2005), Skolnik (2001)]. Especially for common SAS imaging applications, the relatively slow platform motion hardly causes any significant Doppler effects. Here, pulse compression denotes a technique that aims at narrowing the shape of a signal in the time-domain in order to improve resolution, *i.e.*, to increase the capability of discriminating targets. Classically, there are two methods for pulse compression, namely, deramping and matched-filtering [Soumekh (1999)]. Deramping is usually used in SAR applications [Jakowatz et al. (1996)] to compress the received pulse as it is computationally more attractive. However, the method requires the pulse length to cover the entire range of interest. In case of sonar applications, the equivalent pulse length in meter (scaled by the sound speed) is rather small compared to the maximum range of interest, which is a consequence of the relatively slow sound speed in water. Additionally, longer pulse lengths to cover the range of interest are usually not employed due to (i) avoiding large blind zones, and (ii) duty cycle considerations. Thus, the deramping method is inapplicable for SAS. Hence, the focus lies on matched-filtering, which uses the conjugated time-reversed transmitted signal as an impulse response, and maximizes the SNR in presence of additive white noise. This filtering approach is identical to determining the auto-correlation function of the transmitted signal.

First of all, general resolution aspects are addressed. Therefore, consider the transceiver in Figure 2.3 that sends a pulse signal $s^{\mathrm{BP}}(t)$ in the direction of two closely spaced point targets, which are separated by δ_r. According to (2.3), the received echo signal is a superposition of both target responses, which have a time delay difference of $\Delta\tau_{12} = 2(r_2 - r_1)/c = 2\delta_r/c$. Consequently, they can be clearly resolved if the condition $\delta_r > (cT_{\Pi})/2$ holds. Taking into account the relationship between pulse length and bandwidth, *i.e.*, $T_{\Pi} = 1/f_B$ that results from the Fourier transform of a rectangular pulse [Oppenheim et al. (1989)], the range (slant-range) resolution [Jakowatz et al. (1996), Soumekh (1999)] can be expressed as

$$\delta_r = \frac{c}{2f_B}. \tag{2.8}$$

Consequently, short pulses with a high bandwidth are desirable to attain a high range resolution for monochromatic pulses. However, taking path propagation losses into account, a minimum power has to be transmitted in order to receive a signal strength that is detectable [Skolnik (2001)]. This trade-off leads to the use of wideband frequency modulated signals, *e.g.*, LFM signals that can achieve a time-bandwidth

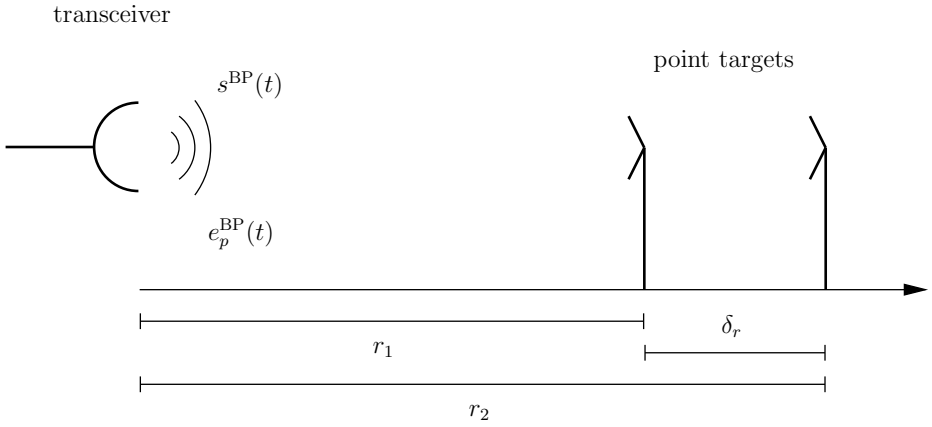


Figure 2.3: Scenario of resolving two point targets in range direction.

product that is much larger than one, *i.e.*, $T_\Pi f_B \gg 1$. An LFM signal with a linear upsweep can be written in the continuous-time domain as follows

$$s^{\mathrm{BP}}(t) = \exp\bigl(j\pi\alpha_c t^2\bigr)\exp\bigl(j\Omega_c t\bigr)\,\Pi(t/T_\Pi) \tag{2.9}$$

$$\text{with } \Pi(t/T_\Pi) = \begin{cases} 1 & -T_\Pi/2 \le t \le T_\Pi/2 \\ 0 & \text{elsewhere,} \end{cases} \tag{2.10}$$

where $\alpha_c = f_B/T_\Pi$ denotes the chirp rate and Ω_c the angular carrier frequency. After deriving an expression for the instantaneous frequency $f(t)$ as the phase derivative [Cohen (1995)] of (2.9) such that

$$f(t) = \frac{1}{2\pi}\frac{\mathrm{d}}{\mathrm{d}t}\bigl(\pi\alpha_c t^2 + \Omega_c t\bigr) = f_c + \alpha_c t, \tag{2.11}$$

it becomes apparent that the bandwidth f_B of the LFM pulse signal can be found by

$$f_B = f(t)\big|_{t=T_\Pi/2} - f(t)\big|_{t=-T_\Pi/2}, \tag{2.12}$$

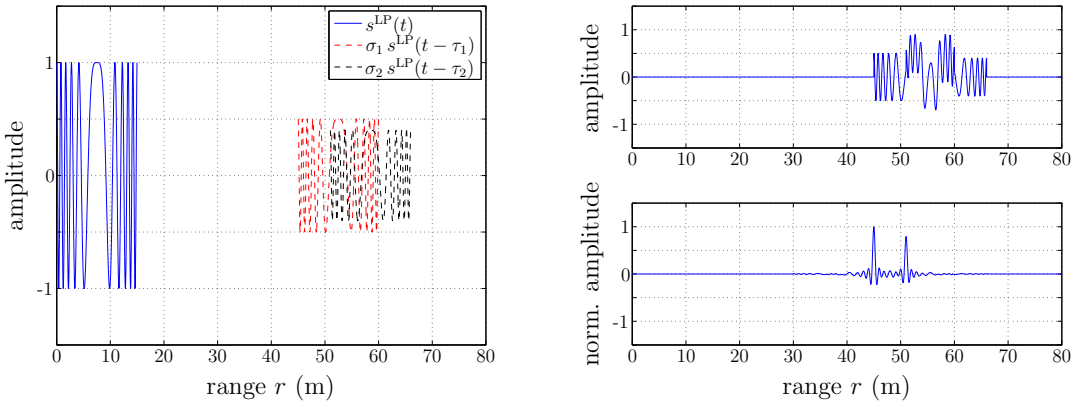
which is independent of a particular pulse length. Hence, an LFM pulse decouples range resolution from aspects relating to pulse energy [Richards (2005)]. However, in order to achieve a gain in resolution, the pulse compression technique in terms of a matched-filter operation has to be performed. Consequently, the pulse-compressed echo signal model in case of an LFM pulse [Hein (2003)] is given by

$$\begin{aligned} e_p(n) &= e_p^{\mathrm{LP}}(t) *_t s^{\mathrm{LP}}(-t)^*\Big|_{t=nT_s} \\ &= \left(\sum_{d=1}^{D} \sigma_d\, b_{\mathrm{phy}}(\theta_{d,p})\, e^{-j\Omega_c\tau_{d,p}}\, s_p^{\mathrm{LP}}(t-\tau_{d,p})\right) *_t s^{\mathrm{LP}}(-t)^*\Big|_{t=nT_s} \\ &= \sum_{d=1}^{D} \sigma_d\, b_{\mathrm{phy}}(\theta_{d,p})\,\Lambda\left(\frac{n-\eta_{d,p}}{M_\Pi}\right)\,\mathrm{si}\left(\frac{\pi f_B(n-\eta_{d,p})}{f_s}\,\Lambda\left(\frac{n-\eta_{d,p}}{M_\Pi}\right)\right)\, e^{-j\omega_c\eta_{d,p}} \end{aligned} \tag{2.13}$$

where $*_t$ and $(\cdot)^*$ denote convolution in time and complex conjugation, respectively, and $\mathrm{si}(x) = \sin(x)/x$. Further, $\Lambda(n/M_\Pi)$ is the triangular function of width $2M_\Pi = 2T_\Pi/T_s$. In (2.13), the additive noise term is omitted. Note that the matched-filter operation is identical to the cross-correlation of $e_p^{\mathrm{LP}}(t)$ and $s^{\mathrm{LP}}(t)$, where $s^{\mathrm{LP}}(t)$ denotes the equivalent lowpass representation of the transmitted signal. Expressing $e_p^{\mathrm{LP}}(t)$ in terms of time shifted replicas of $s^{\mathrm{LP}}(t)$, the matched-filter operation yields the auto-correlation function $\mathrm{r}_{ss}(n)$ of the transmitted pulse. Thus, a more general model of the discrete-time pulse-compressed echo signals reads

$$e_p(n) = \sum_{d=1}^{D} \sigma_d \, b_{\mathrm{phy}}(\theta_{d,p}) \, \mathrm{r}_{ss}(n - \eta_{d,p}) \, e^{-j\omega_c \eta_{d,p}}. \tag{2.14}$$

A comparison of the attained range resolution using LFM echo signals before and after range compression is illustrated in Figure 2.4a. It shows the transmitted signal (blue) and two individual echo signals corresponding to the two closely spaced targets as depicted in Figure 2.3. Figure 2.4b (top) shows the superposition of both echo signals according to (2.7) as well as the pulse-compressed echo signal (bottom) of (2.13). Clearly, the pulse-compressed echo signals enhance the range resolution significantly, and the two targets can be resolved.



(a) Transmitted signal (blue) and two received echoes (red and black)

(b) Superposition of echoes (top) and pulse-compressed echoes (bottom)

Figure 2.4: Illustration of the transmitted LFM signal and individually received echo signals (a) and the gain in resolution that is achievable using pulse compression (b). The abscissa is converted from fast-time t to the range dimension r in all plots.

2.2 Synthetic aperture principle

As outlined in Section 2.1, a synthetic aperture is constructed through repeatedly transmitting pulses and receiving the corresponding echo signals of the scene of

interest at equally spaced positions while moving along a predefined rectilinear trajectory. Thereby, for each spatial position, the target area is illuminated by the footprint of the physical aperture beam. Its beamwidth extent in cross-range direction [Van Trees (2002), Richards (2005)] depends on the angular beamwidth θ_{BW} of the physical aperture and is proportional to range r – see Figure 2.5. Moreover, the beamwidth extent in cross-range direction approximates the synthetic aperture length L_{syn}. Hence, the latter becomes a function of range, which dynamically adjusts itself according to [Jakowatz et al. (1996)]

$$L_{\mathrm{syn}}(r) = 2r \tan(\theta_{\mathrm{BW}}/2) \approx r\theta_{\mathrm{BW}}. \tag{2.15}$$

Consequently, this leads to a prolonged length of the synthetic aperture for larger range values as illustrated in Figure 2.5. Since the second target (right) is further away than the first target (left), *i.e.*, $r_2 > r_1$, it follows that $L_{\mathrm{syn}}(r_2) > L_{\mathrm{syn}}(r_1)$.

In contrast to range resolution, which is related to the bandwidth of the transmitted pulse as discussed in Section 2.1.1, the cross-range resolution is mainly influenced by the wavelength of the transmitted signal and the aperture size. Hence, by collecting more information about targets located further away, and therefore, enlarging the synthetic aperture size, a more confined beam can be formed. This ideally maintains the cross-range resolution over range, which leads to high-resolution images in the case of synthetic aperture imaging [Carrara et al. (1995), Jakowatz et al. (1996), Richards (2005)]. More details on cross-range resolution aspects are discussed in the following.

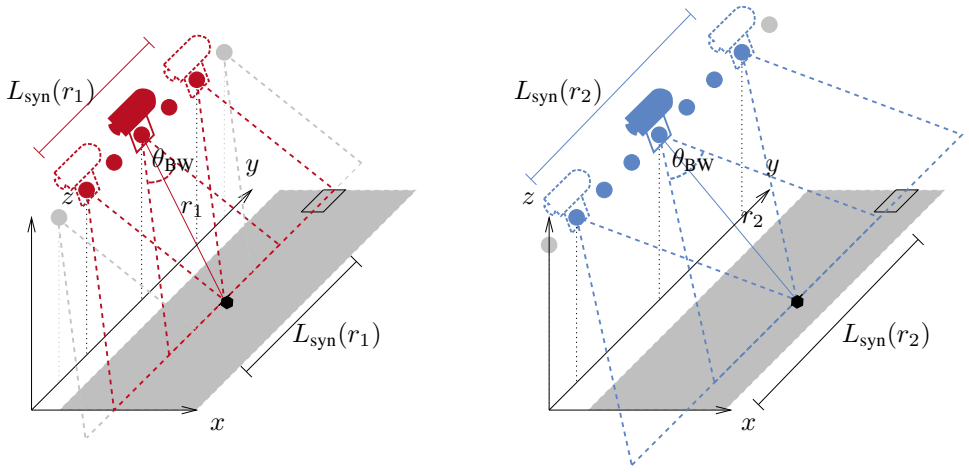


Figure 2.5: Principle of dynamic adjustment of a synthetic aperture length $L_{\mathrm{syn}}(r)$ proportional to range r that is exemplified for two point targets located in range r_1 (left) and r_2 (right) at the point of closest approach.

2.2.1 Cross-range resolution

In order to derive an expression for the cross-range resolution of a synthetic aperture system, the angular beamwidth of a planar antenna and a uniform linear array (ULA) are investigated first. In case of a planar antenna, the aperture is uniformly illuminated along its physical dimension. For plane wave propagation, the electric field intensity $E_A(\theta)$ as a function of the incidence angle θ is given by the inverse Fourier transform of the current distribution $A(y)$ across the aperture dimension y [Skolnik (2001), Richards (2005)] as follows

$$E_A(\theta) = \int_{-D_{\text{phy}}/2}^{D_{\text{phy}}/2} A(y)\, e^{j\frac{2\pi}{\lambda}\sin(\theta)y} dy, \tag{2.16}$$

where λ denotes the wavelength of the plane wave and D_{phy} is the physical size of the planar antenna. Assuming a constant illumination, *i.e.*, $A(y) = A_0 = 1, \forall\, |y| \leq D_{\text{phy}}/2$ and normalizing such that $\bar{E}_A(0) = 1$, the normalized amplitude pattern of a planar antenna can be expressed as

$$\bar{E}_A(\theta) = \frac{\sin\big(\pi(D_{\text{phy}}/\lambda)\sin(\theta)\big)}{\pi(D_{\text{phy}}/\lambda)\sin(\theta)}. \tag{2.17}$$

Given the normalized amplitude pattern $\bar{E}_A(\theta)$, various definitions of the angular beamwidth, such as the Rayleigh angular beamwidth $\theta_R = 2\arcsin(\lambda/D_{\text{phy}})$ or the half-power angular beamwidth $\theta_{3\text{dB}} \approx \arcsin(0.89\lambda/D_{\text{phy}})$, exist [Richards (2005), Van Trees (2002)]. However, for ease of notation and unless stated otherwise, the 4 dB angular beamwidth θ_{BW} and a small angle approximation [Richards (2005)] are used for the remainder of the thesis such that

$$\theta_{\text{BW}} = \arcsin(\lambda/D_{\text{phy}}) \approx \lambda/D_{\text{phy}}. \tag{2.18}$$

Similarly, an angular beamwidth expression of a ULA can be determined as illustrated in Figure 2.6. It shows a ULA consisting of $N_{\text{rx}} = 2\tilde{N}_{\text{rx}} + 1$ receiving elements individually separated by a physical interelement spacing Δ^{u} in array dimension y, where $\tilde{N}_{\text{rx}}$ is assumed even. Considering that the transmitter is co-located with the receiver at the center of the array, the round-trip delay between a target at point (r, θ) and the u^{th} receiver can be written as

$$\tau(u) = \frac{r + r(u)}{c}, \text{ with } u = 1, \ldots, N_{\text{rx}}, \tag{2.19}$$

where $r(u)$ denotes the target distance of receiver u. Provided that the target distance r is much longer than the physical array dimension $L_{\text{phy}} = \Delta^{\text{u}} N_{\text{rx}}$, *i.e.*, $r \gg L_{\text{phy}}$, the

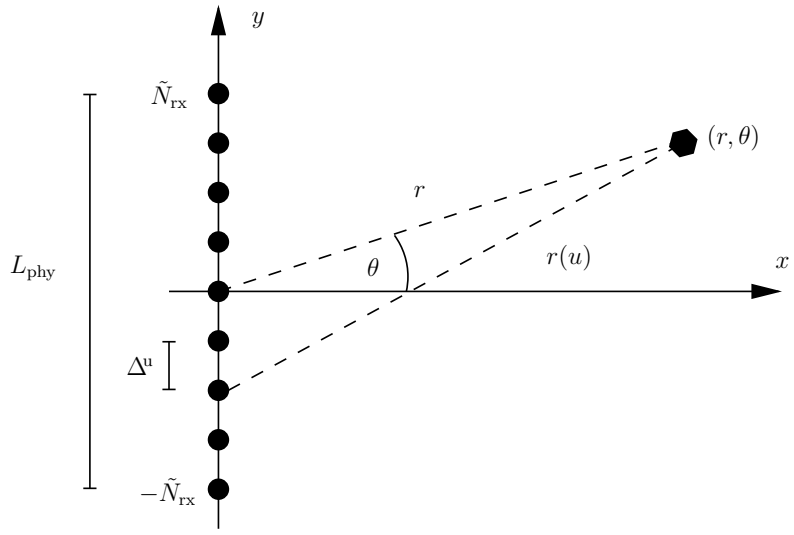


Figure 2.6: Geometry of a uniform linear array consisting of $N_{\mathrm{rx}} = 2\tilde{N}_{\mathrm{rx}} + 1$ receiving elements spaced by Δ^{u} with a total array length of L_{phy} and a target at location (r, θ) – adapted from [Richards (2005)].

target distance $r(u)$ of receiver u can be well approximated by a first-order Maclaurin series expansion as [Richards (2005)]

$$r(u; \theta) \approx r - u\Delta^{\mathrm{u}} \sin(\theta). \tag{2.20}$$

Thus, an ideal received echo signal $e^{\mathrm{BP}}(u, t)$ along the ULA is given by

$$e^{\mathrm{BP}}(u, t) = s^{\mathrm{BP}}(t - \tau(u; \theta)), \tag{2.21}$$

neglecting scaling factors due to propagation losses, beampattern weighting and target reflectivity. For simplicity and without loss of generality, $s^{\mathrm{BP}}(t) = e^{j2\pi f_c t}$ is assumed as a monochromatic transmitted signal with frequency f_c. In order to find the angular beamwidth θ_{phy} of the physical array consisting of ideal isotropic elements, the echo signals $e^{\mathrm{BP}}(u, t)$ are beamformed in broadside direction $\theta_0 = 0$. Again, under far-field and narrowband conditions [Soumekh (1999)], the beamformed broadside signal can be expressed as the sum over all received echo signals as follows [Richards (2005)]

$$\frac{1}{N_{\mathrm{rx}}} \sum_{u=-\tilde{N}_{\mathrm{rx}}/2}^{\tilde{N}_{\mathrm{rx}}/2} e^{\mathrm{BP}}(u, t) = e^{j2\pi f_c t}\, e^{-j\frac{2\pi}{\lambda_c} r} \underbrace{\frac{1}{N_{\mathrm{rx}}} \sum_{u=-\tilde{N}_{\mathrm{rx}}/2}^{\tilde{N}_{\mathrm{rx}}/2} e^{-j\frac{2\pi}{\lambda_c} r(u;\theta)}}_{= \mathrm{AF}_{\mathrm{phy}}(\theta)}. \tag{2.22}$$

Substituting the target distance $r(u; \theta)$ of (2.20) into the expression for $\mathrm{AF}_{\mathrm{phy}}(\theta)$ in (2.22) and evaluating the sum of a finite geometric series [Finckenstein et al. (2006)]

yields the physical array factor as

$$\mathrm{AF}_{\mathrm{phy}}(\theta) = \frac{1}{N_{\mathrm{rx}}} \sum_{u=-\tilde{N}_{\mathrm{rx}}/2}^{\tilde{N}_{\mathrm{rx}}/2} e^{-j\frac{2\pi}{\lambda_c} r(u;\theta)} = \frac{1}{N_{\mathrm{rx}}} \frac{\sin\left(\frac{\pi L_{\mathrm{phy}}}{\lambda_c}\sin(\theta)\right)}{\sin\left(\frac{\pi \Delta^{\mathrm{u}}}{\lambda_c}\sin(\theta)\right)}. \tag{2.23}$$

The angular Rayleigh beamwidth of a ULA is then determined by the null-to-null mainlobe width of the array beampattern. It is given by $\theta^R_{\mathrm{phy}} = 2\arcsin(\lambda_c/L_{\mathrm{phy}})$, which is equivalent to the Rayleigh resolution of a planar antenna except that the array length L_{phy} is used instead of the aperture size D_{phy}. Again, various angular beamwidths of a physical array can be defined. Similar to the angular beamwidth of the planar antenna, the 4 dB angular beamwidth [Richards (2005)] is chosen as

$$\theta_{\mathrm{phy}} \approx \lambda_c / L_{\mathrm{phy}}, \tag{2.24}$$

using a small angle approximation. In contrast to a physical array, a synthetic aperture is built by moving a transceiver with a step size Δ^{A} that sends a signal from each receiving element position of the physical array (here, $p\Delta^{\mathrm{A}} \equiv u\Delta^{\mathrm{u}}$). Consequently, the round-trip delay changes to

$$\tau(p;\theta) = \frac{2(r - p\Delta^{\mathrm{A}}\sin(\theta))}{c}. \tag{2.25}$$

Substituting (2.25) into (2.21) and forming a broadside beam signal leads to the array factor of the synthetic aperture [Richards (2005)] as follows

$$\mathrm{AF}_{\mathrm{syn}}(\theta) = \frac{1}{M_p} \frac{\sin\left(\frac{2\pi L_{\mathrm{syn}}}{\lambda_c}\sin(\theta)\right)}{\sin\left(\frac{2\pi \Delta^{\mathrm{A}}}{\lambda_c}\sin(\theta)\right)}, \tag{2.26}$$

which is identical to (2.23) except for a factor of two in the argument. Note that the number of receiving elements N_{rx} is replaced by the total number of transmission times M_p in (2.26). The factor two results from the two-way phase information of the round-trip delay that is available from each sensor position in the case of synthetic apertures [Richards (2005)]. Therefore, the angular beamwidth of a synthetic aperture is half the one of a physical array, *i.e.*, $\theta_{\mathrm{syn}} = \lambda_c/(2L_{\mathrm{syn}}) \approx \theta_{\mathrm{phy}}/2$ with $L_{\mathrm{syn}} = L_{\mathrm{phy}}$. In other words, to form a beam with the same angular beamwidth, a physical array of twice the length of a synthetic array is required.

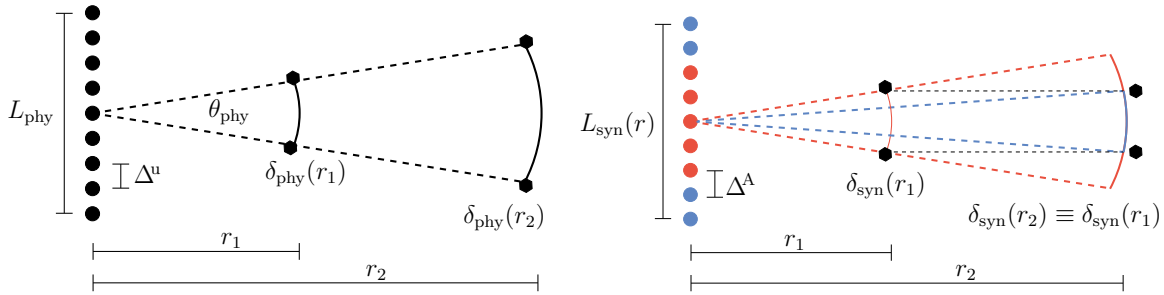
In the following, the difference in cross-range resolution between physical and synthetic arrays is discussed, which is approximately given by the arc length [Jakowatz et al. (1996), Richards (2005)]

$$\delta_{\mathrm{phy}}(r) = 2r\tan(\theta_{\mathrm{phy}}/2) \approx r\,\theta_{\mathrm{phy}} \approx \frac{r\lambda_c}{L_{\mathrm{phy}}} \tag{2.27}$$

for a physical array. As a consequence, the cross-range resolution worsens with range, which leads to a spatially varying resolution of radar or sonar images, which are reconstructed using a real aperture system (RAS). An illustration of cross-range resolution degradation is shown in Figure 2.7a for two different range values. By contrast, the cross-range resolution of the synthetic aperture array can be written due to the dynamically adapting length as follows

$$\delta_{\text{syn}} \approx r\,\theta_{\text{syn}}(r) \approx \frac{r\lambda_c}{2L_{\text{syn}}(r)} \approx \frac{r\lambda_c}{2\,r\,\theta_{\text{BW}}} = \frac{D_{\text{phy}}}{2}, \tag{2.28}$$

which solely depends on the physical aperture size $D_{\text{phy}} \approx \lambda_c/\theta_{\text{BW}}$. Thus, neglecting path losses and SNR considerations, the resolution in synthetic aperture imaging is theoretically constant over range. Moreover, unlike physical arrays where a large aperture size is desirable to improve resolution for large distances, a small transceiver design is preferable for synthetic aperture imaging systems. However, taking into account spatial sampling requirements as outlined in Section 2.3, a trade-off between cross-range resolution and coverage rate of a synthetic aperture system becomes apparent. Figure 2.7b demonstrates the principle of aperture adjustment as well as the resulting capability to form a narrower beam for increasing range values so as to maintain the cross-range resolution over range.



(a) Physical array (b) Synthetic array

Figure 2.7: Cross-range resolution of a physical array (a) and synthetic array (b). The physical array resolution worsens with range and requires double the array size to form the same beamwidth, while the resolution of the synthetic array is constant over range.

2.3 Spatial sampling requirements

While spatial sampling for physical arrays is achieved by spacing the sensors at a distance Δ^{u} along the array dimension, the synthetic aperture is sampled in space by moving the transceiver with a certain step size between two pings. This so-called advance per ping (APP) is denoted by Δ^{A} and depicted in Figure 2.1.

In either case, a violation of the spatial Nyquist criterion leads to cross-range or azimuth image ambiguities due to spatial aliasing [Van Trees (2002), Soumekh (1999)] that can mistakenly be interpreted as targets. Simultaneously, the contrast in the reconstructed image diminishes. Especially for SAS, image ambiguities can destroy shadow information among other degradations, which may lead to a significant performance loss of automatic target recognition (ATR) systems – see Chapter 5.

In the following, an upper bound $\Delta^{\mathrm{A}}_{\max}$ on the APP is provided to correctly sample the synthetic aperture array [Soumekh (1999)]. In order to find an expression for $\Delta^{\mathrm{A}}_{\max}$, the scenario of Figure 2.6 is assumed, where an ideal point target is located at position (r, θ). The relationship between the wavenumber (spatial frequency) in cross-range direction (k_y) and the wavenumber of the wave propagation direction (k_r) is then given by

$$k_y = 2k_r \, \sin(\theta), \tag{2.29}$$

assuming far-field conditions. Note that the factor two in (2.29) results from the two-way phase information that is collected for each sensor position. Given (2.29) and the relation between wavenumber and wavelength as $k_r = (2\pi)/\lambda$, the maximum spatial sampling interval $\Delta^{\mathrm{A}}_{\max}$ according to the Nyquist theorem [Soumekh (1999)] is found as

$$\Delta^{\mathrm{A}}_{\max} \leq \frac{\pi}{k_{y,\max}} = \frac{\lambda_{\min}}{4\sin(\theta_{\max})}, \tag{2.30}$$

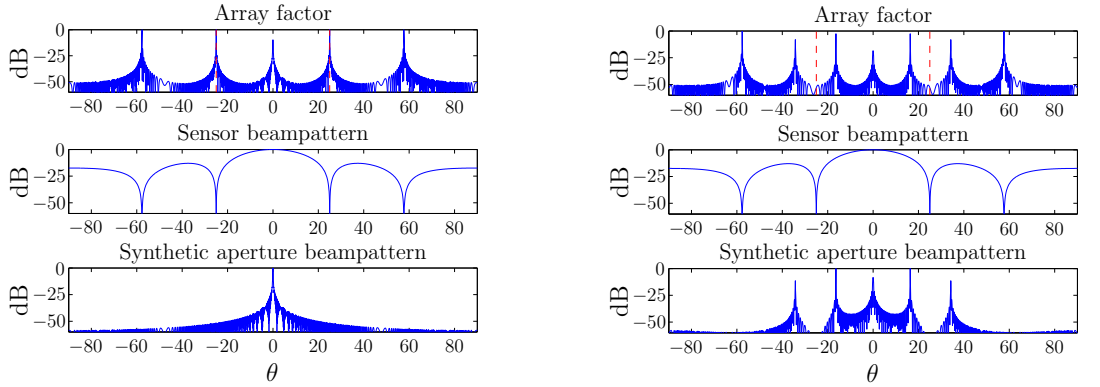
where $\lambda_{\min}$ denotes the minimum wavelength of the transmitted signal. For a stripmap imaging system, the maximum aspect angle $\theta_{\max}$ equals half of the angular beamwidth of the transceiver element, *i.e.*,

$$\theta_{\max} = \theta_R/2 = \arcsin(\lambda/D_{\mathrm{phy}}). \tag{2.31}$$

In order to have a more conservative bound for the spatial sampling requirement, the Rayleigh angular beamwidth θ_R is used in (2.31) instead of the 4 dB beamwidth θ_{BW} as depicted in Figure 2.2b and detailed in Section 2.2.1. Substituting the expression of (2.31) into (2.30) leads to the maximum APP, and the spatial sampling constraint can be denoted by

$$\Delta^{\mathrm{A}}_{\max} \leq \frac{D_{\mathrm{phy}}}{4}. \tag{2.32}$$

Thus, the maximum APP depends on the physical aperture size of the transceiver D_{phy}. Note that simply increasing the physical aperture size D_{phy} to enlarge $\Delta^{\mathrm{A}}_{\max}$ contradicts the design goal of a fine cross-range resolution as stated in (2.28). Violating the condition in (2.32) yields cross-range image ambiguities, which can severely affect image



(a) Correctly sampled (b) Violation of spatial Nyquist criterion

Figure 2.8: Comparison of the beampattern for a spatially correctly sampled synthetic aperture (a) and for violating the spatial Nyquist criterion (b) leading to cross-range image ambiguities.

quality. A beampattern comparison of a correctly sampled and undersampled synthetic aperture is depicted in Figure 2.8a and Figure 2.8b, respectively. Choosing the upper bound $\Delta^{\mathrm{A}}_{\max}$ according to the condition in (2.32) leads to an aliased array factor in Figure 2.8a. However, the array factor is weighted by the transceiver beampattern (element factor), which cancels the occurring ambiguities (also known as grating lobes). In case of spatial undersampling, the grating lobes occur at displaced locations with respect to the nulls of the sensor beampattern, and the synthetic aperture beampattern is affected by aliasing.

2.3.1 Area coverage rate

A key parameter of an imaging system is its area coverage rate as it determines the overall mission time given a fixed size of the search area [Hagen and Hansen (2007)]. The area coverage rate A_{cr} is defined as the product of the maximum imaging range $R_{\max}$ and the speed of the imaging platform v. The latter is linearly related to Δ^{A} via the pulse repetition interval (PRI) denoted by T_{PRI} [Richards (2005)] as follows

$$\Delta^{\mathrm{A}} = v T_{\mathrm{PRI}}. \tag{2.33}$$

The PRI describes the time difference between two transmission times, which is denoted by $T_{\mathrm{PRI}} = t_{p+1} - t_p$. The area coverage rate of a synthetic aperture system assuming a uniformly good image quality over range is then given by [Cook (2007)]

$$A_{\mathrm{cr}} \leq v R_{\max} \leq \frac{v c T_{\mathrm{PRI}}}{2} = \frac{\Delta^{\mathrm{A}} c}{2}. \tag{2.34}$$

Here, $R_{\max} \leq c\,T_{\mathrm{PRI}}/2$ describes the relation between the maximum range and the PRI ignoring the pulse length in order to avoid range ambiguities [Richards (2005)]. Hence, an increase in the advance per ping Δ^{A} is either related to an increase in T_{PRI} or platform speed v. However, a larger T_{PRI} rather affects the maximum range $R_{\max}$, which is at the same time limited by the SNR. Thus, given a maximum range $R_{\max}$ of the imaging system, the area coverage rate A_{cr} is solely determined by the platform speed v. It should be remarked that the area coverage rate in (2.34) is too optimistic for practical synthetic aperture systems, as it ignores parameters such as minimum and maximum ground-range swath distances given by $X_{\min}$ and $X_{\max}$, respectively. The latter usually depend on the altitude, the depression angle of the imaging platform, and the vertical angular beamwidth – see the scenario depicted in Figure 2.1. A redundancy or overlap of the imaging swath of parallel vehicle tracks reduces the area coverage even more. For these reasons, the coverage rate in (2.34) should be interpreted as an upper bound.

Assuming common parameters for SAR (space-borne) and SAS systems, the area coverage rate can be determined to be $A_{\mathrm{cr}}^{\mathrm{SAR}} \approx 1.3 \cdot 10^6\,\mathrm{km}^2/\mathrm{h}$ and $A_{\mathrm{cr}}^{\mathrm{SAS}} \approx 0.03\,\mathrm{km}^2/\mathrm{h}$, respectively, assuming a single transceiver synthetic aperture system. Regarding the SAS coverage rate, it becomes apparent that a single transceiver SAS system is unusable in practice. In order to achieve useful coverage rates for SAS, a single-transmitter and multi-receiver configuration is employed [Hayes and Gough (2009)] that improves the coverage rate by a factor equal to the number of receivers. Typically, the multi-receiver configuration consists of a ULA of N_{rx} elements. The maximum achievable advance per ping $\Delta_{\max}^{\mathrm{ULA}}$ without causing grating lobes is then given by [Hayes and Gough (2009)]

$$\Delta_{\max}^{\mathrm{ULA}} \leq N_{\mathrm{rx}}\,\Delta_{\max}^{\mathrm{A}}. \tag{2.35}$$

Although the improvement is still extremely small compared to SAR systems, it leads to useable SAS systems achieving practical coverage rates of approximately $A_{\mathrm{cr}}^{\mathrm{SAS}} \approx 1.35\,\mathrm{km}^2/\mathrm{h}$. Moreover, the area coverage rate can be further increased by mounting a sonar array onto each side of the imaging platform. The resulting gain in the coverage rate, which mainly depends on the type of survey pattern [Hagen and Hansen (2007)], varies approximately between $20 - 60\%$. However, this improvement comes along with the cost of additional energy consumption.

Chapter 3

Synthetic aperture sonar image reconstruction

Synthetic aperture sonar (SAS) imaging is state-of-the-art for reconstructing high-resolution seafloor images, which are used for various purposes, *e.g.*, mine hunting [Groen et al. (2010), Fandos (2012), Fandos et al. (2013)]. This chapter addresses the involved signal processing techniques to reconstruct high-resolution SAS images for raw sonar measurements collected by a single-transmitter and multi-receiver system over numerous consecutive transmission times. In this chapter, an ideal inertial navigation system (INS) with perfect position information is assumed in order to neglect defocusing effects due to the presence of motion errors. Motion compensation in the case of an imperfect INS is described in Chapter 4. Despite this ideal knowledge, a ground-range based SAS image, which is desirable for SAS interferometry [Bonifant (1999), Sæbø et al. (2007)], defocuses for nonlinear pathways [Jakowatz et al. (1996)] if the bathymetry, *i.e.*, underwater topography, consists of strong variations. Therefore, the reconstruction of height maps using a broadside bathymetry estimation technique is considered in order to obtain focused high-quality SAS images even in environments with difficult bathymetry.

This chapter is organized as follows: Section 3.1 provides a specification of a real SAS system, which has been used for recording various sonar measurements. Section 3.2 introduces an extension of the signal model for a single-transmitter and multi-receiver system followed by a description of the backprojection algorithm in Section 3.3 that is employed as space-time imaging technique. Moreover, aperture shading is discussed in Section 3.4 to reduce occurring sidelobe effects in the reconstructed SAS image. The reconstruction of height maps for ground-range imaging is addressed in Section 3.5.

3.1 Sonar system description

The proposed signal processing techniques together with the developed SAS processing chain have been validated using synthesized data as well as real sonar measurements, which have been recorded during different sea trials. The measurement campaigns have been conducted by ATLAS ELEKTRONIK GmbH, Bremen, Germany and ATLAS ELEKTRONIK UK, Winfrith Newburgh, United Kingdom. The available raw sonar

data has been collected using two different sonar arrays, namely, the MCM-SLS and VISION SAS system. Both arrays have been mounted onto the starboard side of a SeaOtter MKII autonomous underwater vehicle (AUV) as illustrated in Figure 3.1. Typically, the SAS system consists of two vertically displaced sonar arrays to facilitate bathymetry estimation using either traditional real aperture sidescan interferometry or SAS interferometry. Additionally, a second interferometric SAS system is mounted onto the port side of the AUV in order to enhance the area coverage rate. Both SAS systems, MCM-SLS and VISION, differ in their array size, location of the transmitter, center frequency, and bandwidth. While the system parameters of the MCM-SLS sonar array, which has been used as an experimental SAS system, are listed in Table 3.1, the VISION system parameters are classified for commercial reasons.

Figure 3.1: SeaOtter MKII autonomous underwater vehicle carrying an interferometric VISION SAS system at starboard side.

Table 3.1: Sonar system parameters of the MCM-SLS sonar array

Quantity	Value
Rx elements	192
Frequency	206 kHz
Bandwidth	48 kHz
Beamwidth (Tx, horiz.)	7 deg
Beamwidth (Tx, Rx, vert.)	18 deg
Beamwidth (Rx, horiz.)	50 deg
Pulse length	10 ms
Range	100 m
Advance per ping	0.4 m

In order to assess the imaging capabilities of a SAS system, a ground truth about the scene content with its objects is highly desirable. Therefore, two man-made objects

have been placed on the seafloor and are repeatedly shown throughout the provided imaging examples in this thesis. Optical images of the targets at land are provided in Figure 3.2. The first object consists of two metal blades forming a cross, which has four small plastic balls attached to each of the four arms (see Figure 3.2a). The spacing of neighboring plastic balls is 0.1 m and each ball has a diameter of 0.05 m. The second target in Figure 3.2b shows a resolution object that consists of four sectors. Two sectors (east and south) contain multiple parallel metal bars in varying distances to measure the resolvability of an imaging system. While the third sector (north) shows small plastic balls attached to differently spaced strings similar to the first target, the fourth sector (west) comprises an object that is akin to rotor blades of a windmill. The total size of the object is 10 m ×10 m. Both targets are mainly used to test the along-track resolution and the focusing capability of the used SAS systems.

(a) Metall cross (b) Resolution target

Figure 3.2: Man-made objects used to test the resolution capability of both SAS systems as well as the proposed processing chain.

3.2 Multi-receiver signal model extension

In order to relax the constraints given by the spatial sampling criterion as outlined in Section 2.3, a single-transmitter and multi-receiver system is used for SAS to increase the maximum forward speed of the imaging platform. Typically, the multi-receiver configuration consists of a uniform linear array (ULA) with N_{rx} elements. Similar to the signal model of a transceiver system in (2.3), the echo signals $e_p^{\mathrm{BP}}(u,t)$ of the u^{th} receiver during ping p can be expressed as

$$e_p^{\mathrm{BP}}(u,t) = \sum_{d=1}^{D} \sigma_d \, b_{\mathrm{phy}}\big(\theta_{d,p}^{\mathrm{tx}}, \theta_{d,p}^{\mathrm{rx}}(u)\big) \, s^{\mathrm{BP}}\big(t - \tau_{d,p}(u)\big) + v_p^{\mathrm{BP}}(u,t) \tag{3.1}$$

$$p = 0, \ldots, M_p - 1, \quad t \in [0, 2R_{\max}/c + T_{\Pi}],$$

for a total number of M_p pings, a maximum imaging range $R_{\max}$, and a pulse length T_{Π}. The additive noise term $v_p^{\mathrm{BP}}(u,t)$ is assumed to be temporally and spatially white. In (3.1), all components of the signal model become a function of receiver index u except for the target reflectivity σ_d as it is assumed to be independent of the aspect angle. Furthermore, the signal delay $\tau_{d,p}(u)$ for receiver u is related to the two-way distance between the transmitter at position $\boldsymbol{a}_p^{\mathrm{tx}} = [x_p^{\mathrm{tx}}, y_p^{\mathrm{tx}}, h_p^{\mathrm{og}}]^T$ and the d^{th} target as well as between target and the u^{th} receiver element. The latter is located at $\boldsymbol{a}_p^{\mathrm{rx}}(u) = \boldsymbol{a}_p^{\mathrm{tx}} + u\,\boldsymbol{\Delta}^{\mathrm{u}}$ with $u = 1, \ldots, N_{\mathrm{rx}}$, where $\boldsymbol{\Delta}^{\mathrm{u}}$ denotes the interelement spacing vector of the ULA with $\|\boldsymbol{\Delta}^{\mathrm{u}}\|_2 = \Delta^{\mathrm{u}}$. Hence, the round-trip delay of a single-transmitter and multi-receiver system reads as

$$\tau_{d,p}(u) = \frac{r_{d,p}^{\mathrm{tx}} + r_{d,p}^{\mathrm{rx}}(u)}{c}, \tag{3.2}$$

where the single-way distances for the transmitter and receiving element u are given by $r_{d,p}^{\mathrm{tx}} = \left\|\boldsymbol{q}_d - \boldsymbol{a}_p^{\mathrm{tx}}\right\|_2$ and $r_{d,p}^{\mathrm{rx}}(u) = \left\|\boldsymbol{a}_p^{\mathrm{rx}}(u) - \boldsymbol{q}_d\right\|_2$, respectively. The described scenario is depicted in Figure 3.3 for a single ping. Moreover, the composite transmit and receive beampattern function $b_{\mathrm{phy}}(\theta_{d,p}^{\mathrm{tx}}, \theta_{d,p}^{\mathrm{rx}}(u))$ in (3.1) determines whether a certain target is observed by the transmitter and the physical array elements. To this end, the overall beampattern is extended to a function of both azimuth aspect angles, namely, $\theta_{d,p}^{\mathrm{rx}}(u)$ and $\theta_{d,p}^{\mathrm{tx}}$ for the u^{th} receiver and transmitter, respectively. The latter is given by

$$\theta_{d,p}^{\mathrm{tx}} = \arctan\left(\frac{y_d - y_p^{\mathrm{tx}}}{x_d - x_p^{\mathrm{tx}}}\right), \tag{3.3}$$

and the aspect angle $\theta_{d,p}^{\mathrm{rx}}(u)$ is calculated accordingly. Here, the altitude component is not considered as it is only relevant for the vertical angular beamwidth, which is assumed to be flat and to cover the entire range swath. Therefore, it does not affect the field of view. Furthermore, assuming the same angular beamwidth θ_{BW} for the transmitter and receiver, the frequency-independent composite transmit and receive

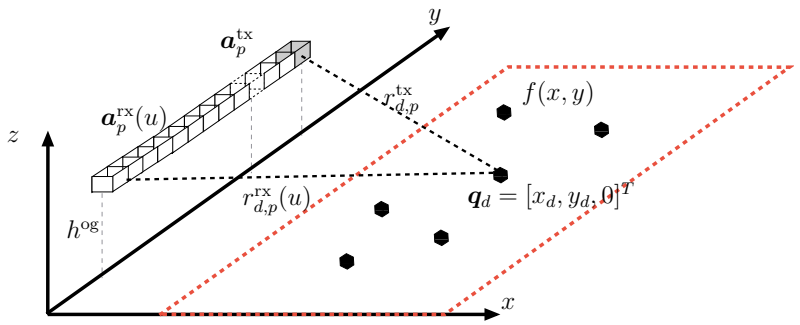


Figure 3.3: Geometry of a SAS system with a single-transmitter and multi-receiver configuration and a target scene $f(x,y)$ consisting of multiple point targets.

beampattern function can be expressed as

$$b_{\mathrm{phy}}\big(\theta^{\mathrm{tx}}_{d,p}, \theta^{\mathrm{rx}}_{d,p}(u)\big) = \begin{cases} w^{\mathrm{tx}}(\theta^{\mathrm{tx}}_{d,p})\, w^{\mathrm{rx}}(\theta^{\mathrm{rx}}_{d,p}(u)) & \text{for } \max\left(\left|\theta^{\mathrm{tx}}_{d,p}\right|, \left|\theta^{\mathrm{rx}}_{d,p}(u)\right|\right) \leq \theta_{\mathrm{BW}}/2 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

where the functions $w^{\mathrm{rx}}(\cdot)$ and $w^{\mathrm{tx}}(\cdot)$ describe an aspect-dependent scaling factor of the beampattern for the receiver elements and the transmitter, respectively. Thus, the extended beampattern function takes into account that a certain target is only observed if it is illuminated by the transmitter beam and sensed by the receiver. This geometrical relation is illustrated in Figure 3.4.

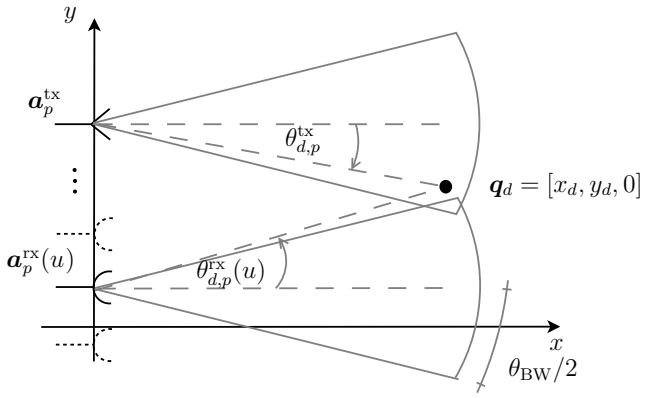


Figure 3.4: Illustration of aspect angles between a point target at location $\boldsymbol{q}_d$ and a single-transmitter ($\boldsymbol{a}_p^{\mathrm{tx}}$) and multi-receiver system. Here, the point target is not observed by the receiving element located at $\boldsymbol{a}_p^{\mathrm{rx}}(u)$.

In order to obtain an equivalent lowpass and pulse-compressed signal model in the discrete-time domain, the pre-processing steps of Chapter 2 such as demodulation, analog-to-digital conversion (ADC) and pulse compression have to be applied to the continuous-time signal model of (3.1). The corresponding pulse-compressed echo signals are then denoted by $e_p(u, n)$ with $p = 0, \ldots, M_p - 1$, $u = 1, \ldots, N_{\mathrm{rx}}$ and $n = 0, \ldots, M_n - 1$, where the number of fast-time samples M_n is given by

$$M_n = \left\lceil \frac{2R_{\max}/c + T_{\Pi}}{T_s} \right\rceil . \tag{3.5}$$

Here, $\lceil \cdot \rceil$ rounds towards the largest integer and T_s denotes the sampling interval. Based on these signals, a technique for ground-range imaging is introduced in the following.

3.3 Ground-range imaging via backprojection

In general, SAS imaging methods are separated into two classes, namely, wavenumber-frequency and space-time based techniques [Hayes and Gough (2009)]. Wavenumber-

frequency imaging techniques such as $\omega - k$ [Stolt (1978), Cafforio et (1991)], range-Doppler imaging techniques [Bamler (1991), Bamler (1992)] and chirp-scaling [Cumming et al. (1992), Raney et al. (1994)] are computationally attractive as they are based on the fast Fourier transform (FFT). However, they inherently assume a regularly spaced rectilinear trajectory, which is disadvantageous for nonlinear pathways and motion compensation techniques such as micronavigation [Bellettini and Pinto (2002)]. By contrast, space-time based techniques, *e.g.*, time-domain correlation (TDC) or the backprojection algorithm [Soumekh (1999)], can handle arbitrary trajectories and are even capable of dealing with a nonlinear sound-speed profile [Hayes and Gough (2009)]. However, on the downside, the computational speed is slow compared to wavenumber-frequency techniques. Although a faster alternative has been developed with the fast-factorized backprojection technique [Basu and Bresler (2000), Ulander et al. (2003)] that trades off image quality with computational speed, the classical backprojection technique (also known as delay-and-sum beamforming) is used in this thesis for SAS image reconstruction accepting the computational disadvantages in order to attain the best image quality. Another advantage over wavenumber-frequency techniques for SAS imaging is the possibility to strip down the reconstruction procedure to the level of processing only single pings, *i.e.*, only the echo data of the physical array is used for image formation. Afterwards, all single ping sonar images also known as real aperture sonar (RAS) images are coherently processed to obtain a SAS image. This characteristic of the space-time image reconstruction is exploited for evaluating the quality of SAS images by using a sequential focus assessment in Chapter 5.

Given pulse-compressed (range focused) echo signals, the aim of image reconstruction techniques is to obtain an estimate of the target scene reflectivity $\hat{f}(x, y)$, with $x \in [X_{\min}, X_{\max}]$ and $y \in [Y_{\min}, Y_{\max}]$. This is achieved by focusing the echo signals collected over the synthetic aperture in cross-range direction. While the target scene is bounded in ground-range direction by $X_{\min} = X_c - X_0$ and $X_{\max} = X_c + X_0$, the boundaries in cross-range direction are given by $Y_{\min} = Y_c - Y_0$ and $Y_{\max} = Y_c + Y_0$. Here, X_c and Y_c denote the center of the target scene in ground-range and cross-range direction, respectively. The corresponding half swath widths of the target scene are given by X_0 and Y_0 – see also Figure 2.1. Due to discrete-time processing, only a discrete reflectivity map $\hat{f}(x_k, y_l)$ at grid points (x_k, y_l) with $k = 1, \ldots, M_x$ and $l = 1, \ldots, M_y$ can be reconstructed by implicitly assuming a ground-plane grid at zero height. Here, M_x and M_y denote the number of grid points in ground-range and cross-range direction, respectively, where each grid point is of size $\Delta x \times \Delta y$ – see Figure 3.5a. Although the pixel size can be chosen arbitrarily coarse or fine to sample the continuous scene reflectivity $\hat{f}(x, y)$, a coarse pixel size leads to an energy

wrapping in the spatial spectrum (wavenumber domain) that hinders post-processing techniques such as image shading. Additionally, potential information about small targets may be lost for a coarse grid. Typically, the pixel size is related to the resolution capability of the synthetic aperture system as discussed in Chapter 2 and hence, to the spectral frequency support B_{k_x} and B_{k_y} of the transformed SAS image $\hat{F}(k_x, k_y)$ in the wavenumber domain. In the case of a stripmap system using a planar aperture [Soumekh (1999)], the spectral support is given by

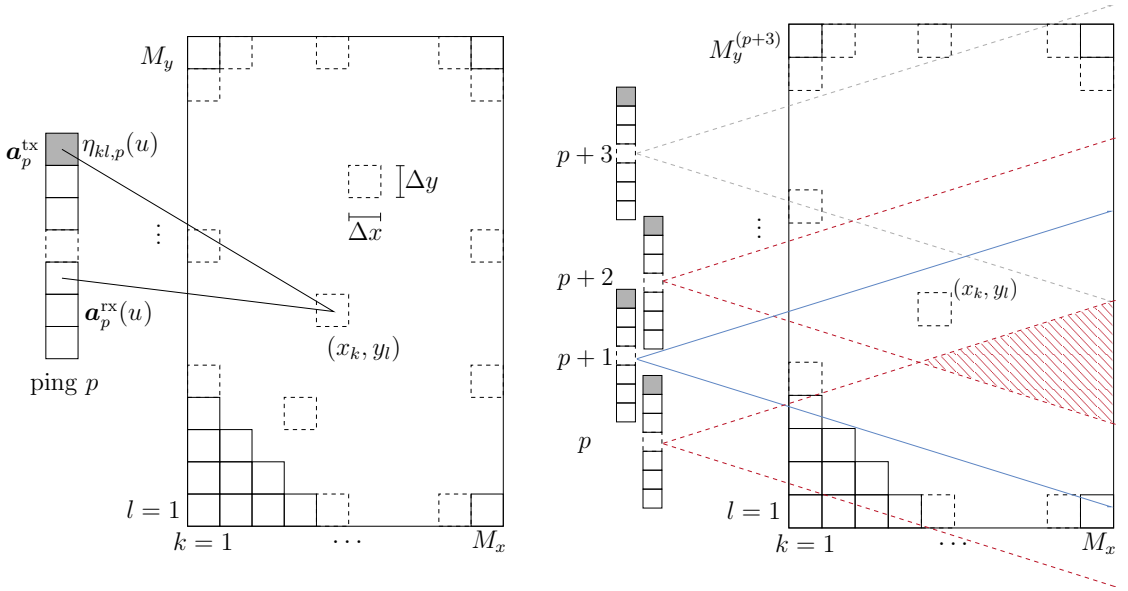
$$B_{k_x} = 2(k_{\max} - k_{\min}) \quad \text{and} \quad B_{k_y} = 4\pi / D_{\text{phy}}, \tag{3.6}$$

in ground-range and cross-range wavenumber dimension, respectively. In (3.6), the minimum wavenumber of the transmitted signal is denoted by $k_{\min} = 2\pi f_{\min}/c$ and the maximum wavenumber is given by $k_{\max} = 2\pi f_{\max}/c$ with $f_{\min} = f_c - f_B/2$ and $f_{\max} = f_c + f_B/2$. According to the spatial sampling criterion – see also Section 2.3, the following conditions have to be fulfilled

$$\Delta x \leq \frac{2\pi}{B_{k_x}} \quad \text{and} \quad \Delta y \leq \frac{2\pi}{B_{k_y}}. \tag{3.7}$$

A minimum grid spacing is then found as $\Delta x \leq c/(2f_B)$ and $\Delta y \leq D_{\text{phy}}/2$, and energy wrapping is avoided in the spatial wavenumber domain.

Using vector notation for a single grid point $\boldsymbol{g}_{kl}$ and assuming a flat target scene, *i.e.*, $\boldsymbol{g}_{kl} = [x_k, y_l, 0]^T$, the scaled focusing delay $\eta_{kl,p}(u)$ between transmitter



(a) Focus delay between physical array and discrete target grid

(b) Illustration of coherent processing of consecutive RAS images

Figure 3.5: Illustration of RAS imaging (a) and coherent processing over several pings to reconstruct a SAS image (b).

location $\boldsymbol{a}_p^{\mathrm{tx}}$, the u^{th} receiver location $\boldsymbol{a}_p^{\mathrm{rx}}(u)$ and grid point $\boldsymbol{g}_{kl}$ reads as

$$\eta_{kl,p}(u) = \frac{\tau_{kl,p}(u)}{T_s} = \frac{\left\|\boldsymbol{g}_{kl} - \boldsymbol{a}_p^{\mathrm{tx}}\right\|_2 + \left\|\boldsymbol{a}_p^{\mathrm{rx}}(u) - \boldsymbol{g}_{kl}\right\|_2}{cT_s}. \tag{3.8}$$

The focusing delay $\tau_{kl,p}(u)$ in (3.8) corresponds to the round-trip delay of the echo signal model in (3.2) with the target location $\boldsymbol{q}_d$ substituted by the grid point $\boldsymbol{g}_{kl}$. Due to sampling, a signal value is typically not available at the focusing delay. Consequently, an interpolation operation is required for the backprojection algorithm [Soumekh (1999), Rosenberg (2004)]. While nearest-neighbor interpolation is inadequate to perform a coherent processing, a linear interpolation such that the signal value at the focusing delay is given by

$$e_p\big(u, \eta_{kl,p}(u)\big) \approx e_p(u, n_0) + \frac{e_p(u, n_1) - e_p(u, n_0)}{n_1 - n_0}\left(\eta_{kl,p}(u) - n_0\right) \tag{3.9}$$

is usually sufficient to obtain good image quality. Here, the discrete-time samples n_0 and n_1 are obtained by $n_0 = \lfloor \eta_{kl,p} \rfloor$ and $n_1 = \lceil \eta_{kl,p} \rceil$, where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ round towards the smallest and largest following integer number, respectively. The single ping RAS image $\hat{f}_p(x_k, y_l)$ at transmission index p is then reconstructed by coherently summing over the echo signals of all N_{rx} receiving elements of the ULA as follows

$$\hat{f}_p(x_k, y_l) = \sum_{u=1}^{N_{\mathrm{rx}}} e_p\left(u, \eta_{kl,p}(u)\right) e^{j\omega_c \eta_{kl,p}(u)} \tag{3.10}$$
$$\text{with } \; k = 1, \ldots, M_x \;\text{ and }\; l = 1, \ldots, M_y,$$

where ω_c denotes the discrete-time angular carrier frequency. The operation in (3.10) is similar to a delay-and-sum interpolation beamformer [Johnson and Dudgeon (1993)]. Assuming a single target scene, *i.e.*, $D = 1$, and substituting the lowpass equivalent pulse-compressed signal of (2.14) for a single-transmitter and multi-receiver system into (3.10) leads to

$$\hat{f}_p(x_k, y_l) = \sigma_1 \sum_{u=1}^{N_{\mathrm{rx}}} \mathrm{r}_{ss}(\eta_{kl,p}(u) - \eta_{1,p}(u)) e^{j\omega_c\left(\eta_{kl,p}(u) - \eta_{1,p}(u)\right)} \tag{3.11}$$
$$\text{with } \; k = 1, \ldots, M_x \;\text{ and }\; l = 1, \ldots, M_y.$$

Thus, if the target position equals a grid point, both focusing delay and round-trip delay are identical and the maximum reflectivity response is attained. Furthermore, the phase term in (3.11) cancels out. Hence, for ground-range imaging, the phase of an image pixel provides information about the range difference between the true target scene and an *a priori* grid. This fact is exploited by interferometry techniques in height map estimation [Sæbø et al. (2007)]. The SAS image of the target scene is then reconstructed using a coherent summation of individual RAS images. Assuming

that the grid point $\boldsymbol{g}_{kl}$ is illuminated at ping $p = p_0$ and $p = p_1$ for the first and last time, respectively, the SAS image reconstruction is given by

$$\hat{f}(x_k, y_l) = \sum_{p=p_0}^{p_1} \hat{f}_p(x_k, y_l)\, b_{\text{phy}}\big(\theta_{kl,p}^{\text{tx}}, \theta_{kl,p}^{\text{rx}}(u)\big) \qquad (3.12)$$
$$\text{for } \; k = 1, \ldots, M_x \text{ and } l = 1, \ldots, M_y,$$

where $b_{\text{phy}}(\theta_{kl,p}^{\text{tx}}, \theta_{kl,p}^{\text{rx}}(u))$ is defined in (3.4). For image reconstruction, the aspect-dependent scaling factors are set to unity within the overlapping beam of the transmitter and receiving elements, *i.e.*, $w^{\text{tx}}(\cdot) = w^{\text{rx}}(\cdot) \equiv 1$. The illustration in Figure 3.5a demonstrates the principle of the backprojection algorithm to obtain the intensity and phase of a respective image pixel by determining a focus delay. After reconstructing individual RAS images, those with a common physical footprint of the target scene are coherently combined to obtain a SAS image. This process requires a common ground grid in order to function as depicted in Figure 3.5b. Concerning computational complexity, a total number of $N_{\text{rx}} M_p M_x M_y$ operations is required, which poses a high computational load. As a consequence, current SAS systems use wavenumber-frequency imaging techniques for large-scale images and switch to space-time techniques such as the introduced backprojection algorithm for smaller areas of interest, *e.g.*, around objects. The latter enables a higher resolution due to the possibility of using more accurate motion compensation [Hansen et al. (2011)].

3.3.1 Grid construction

In stripmap imaging, the illuminated footprint of the target scene moves in cross-range direction according to the advance per ping (APP) of the AUV. Thus, in contrast to spotlight processing, the overlapping section of the target grid changes between subsequent pings. Hence, a procedure is required that coherently combines the overlapping sections of single RAS images so as to obtain a stripmap SAS image. To this end, the boundaries of the illuminated target area of a RAS image in ground-range, $\big[X_{\min}, X_{\max}\big]$, and in cross-range, $\big[Y_{\min}^{\text{RAS}}, Y_{\max}^{\text{RAS}}\big]$, are considered. As the ground-range extent is identical for a SAS and RAS image, the superscript has been omitted, and the swath extent is chosen manually. Given $X_{\min}$ and $X_{\max}$ along with the angular beamwidth θ_{BW}, the boundaries of a RAS image sector in along-track direction are then determined by $Y_{\min}^{\text{RAS}} = -X_{\max} \sin(\theta_{\text{BW}}/2)$ and $Y_{\max}^{\text{RAS}} = X_{\max} \sin(\theta_{\text{BW}}/2)$ relative to the transmitter position, and the sample numbers M_x and M_y are found as follows

$$M_x = \left\lfloor \frac{X_{\max} - X_{\min}}{\Delta x} \right\rfloor + 1 \quad \text{and} \quad M_y = \left\lfloor \frac{2X_{\max} \sin(\theta_{\text{BW}}/2)}{\Delta y} \right\rfloor + 1. \qquad (3.13)$$

In (3.13), $\lfloor \cdot \rceil$ rounds towards the nearest integer neighbor and Δx and Δy denote the pixel size in ground-range and cross-range direction, respectively. As the imaging range is invariant with the ping index $p = 0, \ldots, M_p - 1$, the ground-range grid vector $\boldsymbol{x}$ is written as $\boldsymbol{x} = [x(1), x(2), \ldots, x(k), \ldots, x(M_x)]$ with

$$x(k) = X_{\min} + (k-1)\Delta x. \tag{3.14}$$

Contrarily, the size of the cross-range grid vector $\boldsymbol{y}^p$ and, consequently, the number of cross-range pixels $M_y^{(p)}$ depend on the ping index p due to the moving footprint of the physical aperture. Thus, it becomes necessary to distinguish between the local cross-range vector $\check{\boldsymbol{y}}$ and the global vector $\boldsymbol{y}^p$ at ping p. In general, the global vector contains different grid points of the global reference system for different pings. By contrast, the local cross-range vector $\check{\boldsymbol{y}}$ is identical for each ping, which is denoted by $\check{\boldsymbol{y}} = [\check{y}(1), \check{y}(2), \ldots, \check{y}(l) \ldots, \check{y}(M_y)]$ with

$$\check{y}(l) = Y_{\min}^{\mathrm{RAS}} + (l-1)\Delta y. \tag{3.15}$$

For $p = p_0$, it follows $M_y^{(p_0)} \equiv M_y$. Moreover, note that the notation $x(k) \equiv x_k$ and $y(l) \equiv y_l$ is used interchangeable in the sequel. In order to describe a procedure that guarantees a coherent processing, the matrix notation

$$\boldsymbol{F}_p = \hat{f}_p(x_k, y_l)\Big|_{k=1,\ldots,M_x,\, l=1,\ldots,M_y} \tag{3.16}$$

of a single RAS image at ping p is used. Alternatively, the image matrix can be expressed in terms of stacked range lines $\boldsymbol{f}_l^{(p)} = [\hat{f}_p(x_1, y_l), \hat{f}_p(x_2, y_l), \ldots, \hat{f}_p(x_{M_x}, y_l)]$ as

$$\boldsymbol{F}_p = \begin{bmatrix} \boldsymbol{f}_{M_y}^{(p)} \\ \vdots \\ \boldsymbol{f}_1^{(p)} \end{bmatrix}, \tag{3.17}$$

where the row index is inverted to account for increasing along-track values in positive y direction. Furthermore, a matrix notation $\boldsymbol{F}^{(p+1)}$ of the SAS image is required. Here, the superscript indicates a processing up to ping $p+1$. Hence, for $p = p_0$, the SAS and RAS images are identical and it follows $\boldsymbol{F}^{(p_0)} = \boldsymbol{F}_{p_0}$. For sequential ping processing, a SAS image usually consists of three regions. The first and third region describe the area, which either moves out or into the physical aperture footprint of the current ping. By contrast, the center region of the SAS image overlaps for coherent processing. In order to account for this, the SAS image $\boldsymbol{F}^{(p+1)}$ is defined to consist of three sub-matrix parts as follows

$$\boldsymbol{F}^{(p+1)} = \begin{bmatrix} \check{\boldsymbol{F}}_{p+1} \\ \check{\boldsymbol{F}}_p^{p+1} \\ \check{\boldsymbol{F}}_p \end{bmatrix}, \tag{3.18}$$

where $\check{\boldsymbol{F}}_p^{p+1}$ describes the overlapping area. The sub-matrices $\check{\boldsymbol{F}}_p$ and $\check{\boldsymbol{F}}_{p+1}$ denote the non-overlapping image parts for ping p and $p+1$, respectively. The following procedure is proposed to guarantee pixel overlap from consecutive RAS images and enable a coherent stitching process to construct a SAS image. An illustration of the stitching process is provided in Figure 3.6.

- **Step 1.** Construct the cross-range vectors of two consecutive pings, p and $p+1$, and ensure that the same grid points are used during the reconstruction of both RAS images $\boldsymbol{F}_p$ and $\boldsymbol{F}_{p+1}$ as follows

$$\boldsymbol{y}^p = \left\lfloor \frac{y_p^{\mathrm{tx}} + \check{\boldsymbol{y}}}{\Delta y} \right\rfloor \Delta y \quad \text{and} \quad \boldsymbol{y}^{p+1} = \left\lfloor \frac{y_{p+1}^{\mathrm{tx}} + \check{\boldsymbol{y}}}{\Delta y} \right\rfloor \Delta y,$$

  where y_p^{tx} and y_{p+1}^{tx} denote the along-track coordinates of the transmitter position. For the first ping $p = p_0$: Store $y_{p_0} = y^{p_0}(1)$ and set $\boldsymbol{y}_{p_0}^p = \boldsymbol{y}^{p_0}$.

- **Step 2.** Determine the indices of the first, $\Delta l_{p_0}^{p+1}$, and last range line, Δl_p^{p+1}, of the overlapping image section $\check{\boldsymbol{F}}_p^{p+1}$ as follows

$$\Delta l_{p_0}^{p+1} = \frac{y^{p+1}(1) - y_{p_0}}{\Delta y} \quad \text{and} \quad \Delta l_p^{p+1} = \frac{y^{p+1}\big(M_y\big) - y_{p_0}^p\big(M_y^{(p)}\big)}{\Delta y}.$$

- **Step 3.** Perform an image stitching operation to obtain the SAS image at ping $p+1$ as $\boldsymbol{F}^{(p+1)} = [\check{\boldsymbol{F}}_{p+1},\ \check{\boldsymbol{F}}_p^{p+1},\ \check{\boldsymbol{F}}_p]^T$, where the individual sub-matrices are determined as follows

$$\check{\boldsymbol{F}}_{p+1} = \left[\boldsymbol{f}_{M_y}^{(p+1)}, \ldots, \boldsymbol{f}_{M_y - \Delta l_p^{p+1}+1}^{(p+1)}\right]^T,$$

$$\check{\boldsymbol{F}}_p^{p+1} = \begin{bmatrix} \boldsymbol{f}_{M_y^{(p)}}^{(p)} \\ \vdots \\ \boldsymbol{f}_{\Delta l_{p_0}^{p+1}}^{(p)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{f}_{M_y - \Delta l_p^{p+1}}^{(p+1)} \\ \vdots \\ \boldsymbol{f}_1^{(p+1)} \end{bmatrix},$$

$$\check{\boldsymbol{F}}_p = \left[\boldsymbol{f}_{\Delta l_{p_0}^{p+1}-1}^{(p)}, \ldots, \boldsymbol{f}_1^{(p)}\right]^T.$$

- **Step 4.** Update the joint global cross-range vector $\boldsymbol{y}_{p_0}^p$ to $\boldsymbol{y}_{p_0}^{p+1}$ of the associated SAS image $\boldsymbol{F}^{(p+1)}$ as follows

$$\boldsymbol{y}_{p_0}^{p+1} = \left[\, y_{p_0}^p(1), \ldots, y_{p_0}^p(M_y^{(p)}),\, y^{p+1}(M_y - \Delta l_p^{\,p+1} + 1), \ldots, y^{p+1}(M_y) \,\right],$$

  with size $M_y^{(p+1)} = M_y^{(p)} + \Delta l_p^{\,p+1}$. Increment ping p, continue with Step 1.

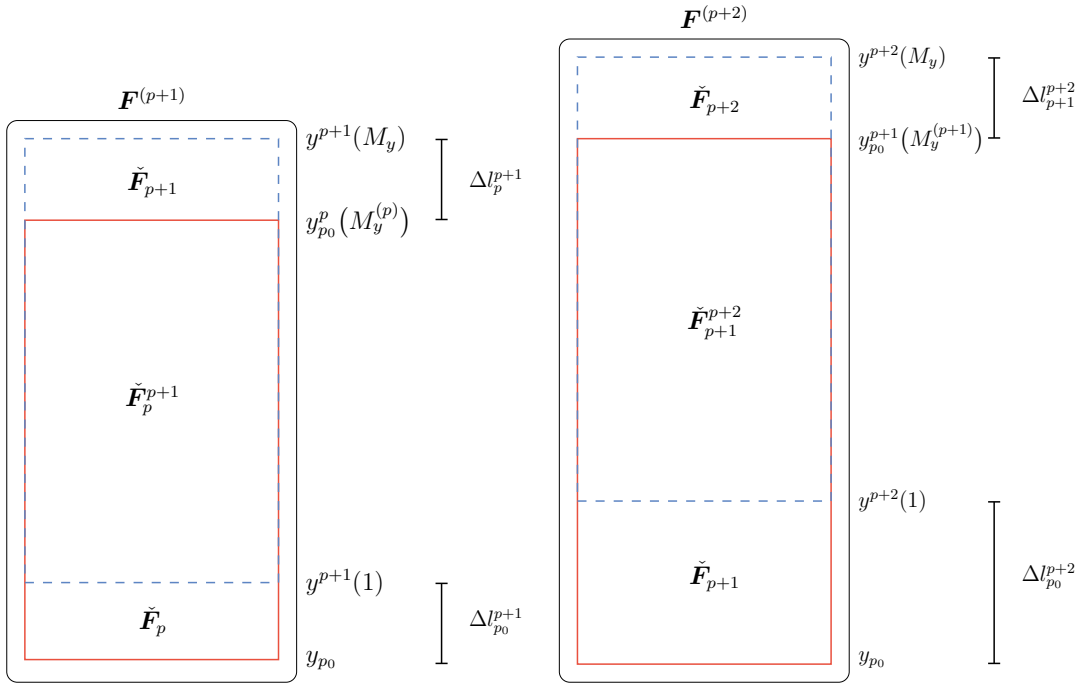


Figure 3.6: Illustration of the coherent stitching process of individual RAS ground grids for pings p, $p+1$, and $p+2$ along with the corresponding sub-matrices.

3.4 Synthetic aperture shading

Aperture shading, also known as aperture apodization or spatial frequency weighting [Jakowatz et al. (1996), Van Trees (2002)], is a technique in array signal processing that trades off sidelobe level (SLL) against mainlobe width by applying different weights to individual array elements. This technique is equivalent to the use of tapers in spectral analysis of time series [Harris (1978)]. Classically, the weights follow a raised cosine or raised cosine-squared function such as a Hanning, Hamming or a Blackman-Harris window, whose shape determines its characteristic SLL and mainlobe width. Another window type is the Taylor window, which approximates the Dolph-Chebyshev window. The latter provides the optimum (spatial) frequency response such that a minimum mainlobe is found given a specified SLL [Jakowatz et al. (1996)]. In imaging, the SLL is typically associated with the contrast of an image, while the mainlobe width is related to its resolution. As the number of receiving elements is invariant, this is a common approach in physical array processing to control the influence of sidelobes.

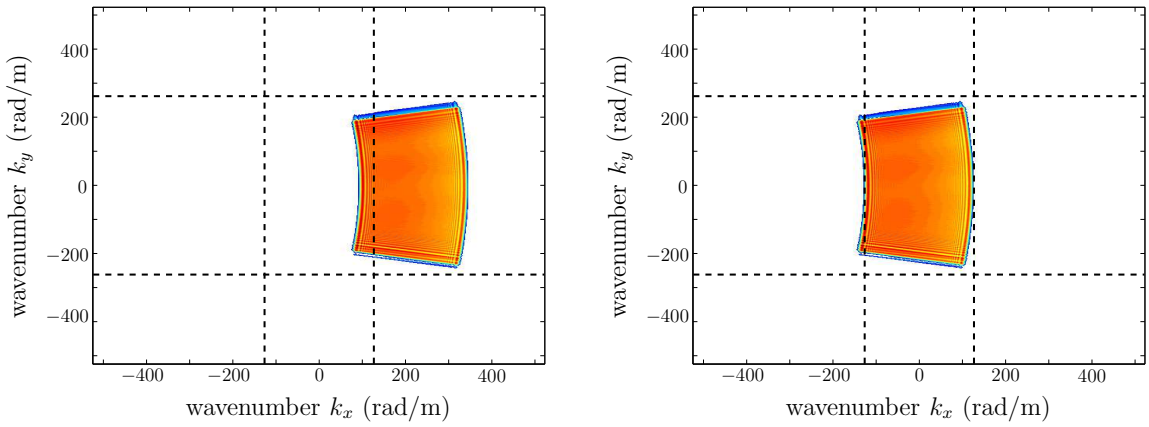
For SAS, the aperture length varies with range and elongates with an increasing number of pings. Thus, determining the correct weights along the synthetic aperture

during space-time image reconstruction is a non-trivial task, especially for unknown motion of the imaging platform. However, in SAR wavenumber-frequency imaging techniques, aperture shading is mainly employed in the 2-D wavenumber domain after polar-to-rectangular interpolation [Stankwitz et al. (1994)] before applying the inverse Fourier transform to the spatial domain. Thus, it is also feasible to use aperture shading as a post-processing technique after SAS image reconstruction as mentioned in [Cook and Brown (2009)]. For this purpose, the SAS image is transformed into the 2-D wavenumber domain using the finite discrete-time Fourier transform (DTFT) as

$$\hat{F}(k_x, k_y) = \sum_{m=1}^{M_x} \sum_{l=1}^{M_y} \hat{f}(x_m, y_l)\, e^{-jk_x(m-1)}\, e^{-jk_y(l-1)}, \tag{3.19}$$

where k_x and k_y are the wavenumbers in ground-range and cross-range directions, respectively. Note that a subscript change from x_k to x_m has been performed here to avoid confusion with the wavenumber symbol k_x. The spatial frequency support of $\hat{F}(k_x, k_y)$ in case of a stripmap system with a planar aperture is given in (3.6) and detailed in [Soumekh (1999)].

An example of a 2-D wavenumber representation is depicted in Figure 3.7, where the backprojection algorithm reconstructs the target scene in the bandpass domain as shown in Figure 3.7a. Therefore, a pixel spacing in ground-range direction of $\Delta x \leq c\,/(2f_c + f_B)$ leads to a spectrum wrapping in k_x dimension. Although this is noncritical for the appearance and quality of the SAS image, it has to be considered for post-processing techniques [Soumekh (1999)] that rely on the phase of the complex image data, such as aperture shading. Thus, the reconstructed SAS image



(a) Bandpass wavenumber spectrum (b) Baseband wavenumber spectrum

Figure 3.7: Wavenumber spectrum of a single target scene in the bandpass region (a) and baseband region (b). Spatial oversampling has been applied in (a) to avoid spectrum wrapping.

is demodulated into baseband, which can be expressed as

$$\hat{f}^{\mathrm{BB}}(x_m, y_l) = \hat{f}(x_m, y_l)\, e^{-j\, 2k_c\, x_m} \quad \begin{array}{l}\text{for } m = 1, \ldots, M_x \\ \text{and } l = 1, \ldots, M_y.\end{array} \tag{3.20}$$

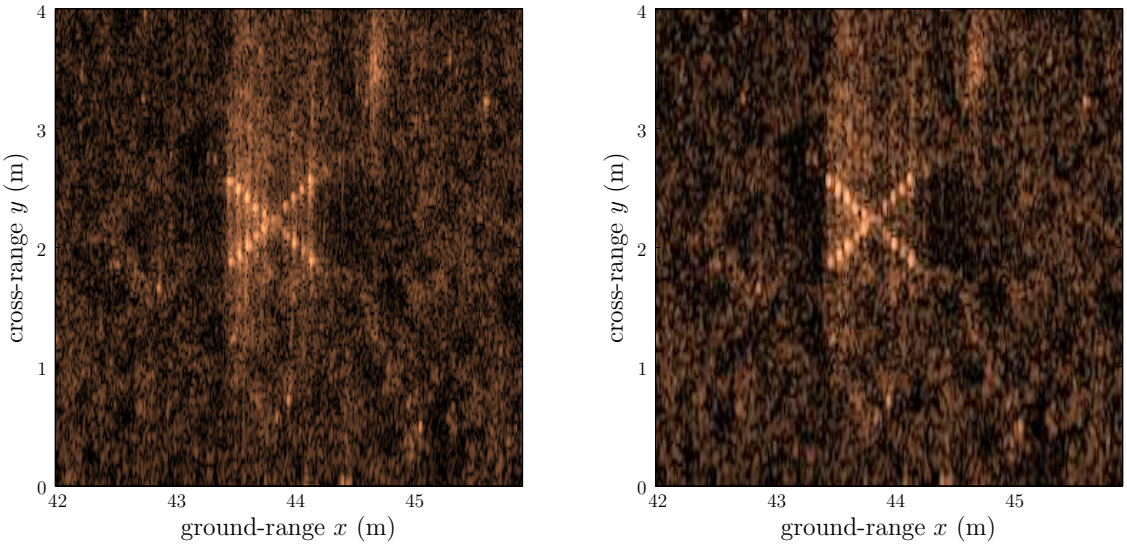
Here, $k_c = 2\pi/\lambda_c$ denotes the wavenumber at the carrier frequency. As this operation is always performed after image reconstruction, the superscript $(\cdot)^{\mathrm{BB}}$ is omitted for notational convenience. The wavenumber representation of a baseband domain SAS image is shown in Figure 3.7b, where the dashed lines indicate the supported bandwidth B_{k_x} and B_{k_y} according to (3.6).

After demodulating the SAS image as provided by (3.20), the wavenumber spectrum is unwrapped for a ground-range spacing of $\Delta x < c/(2f_B)$ and its centroid coincides with the origin of the wavenumber domain. A shading operation can then be applied as follows

$$\hat{f}^{\mathrm{s}}(x_m, y_l) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \hat{F}(k_x, k_y)\, W(k_x, k_y)\, e^{jk_x m}\, e^{jk_y l}\, dk_x\, dk_y \tag{3.21}$$

where $W(k_x, k_y)$ is a 2-D window function in the spatial wavenumber domain and $\hat{f}^{\mathrm{s}}(x_m, y_l)$ denotes the aperture shaded SAS image with $m = 1, \ldots, M_x$ and $l = 1, \ldots, M_y$. In the case of a rectangular window, $W(k_x, k_y)$ is defined as

$$W(k_x, k_y) = \begin{cases} 1, & \text{for } -B_{k_x}/2 \le k_x \le B_{k_x}/2,\ -B_{k_y}/2 \le k_y \le B_{k_y}/2 \\ 0, & \text{otherwise.} \end{cases} \tag{3.22}$$



(a) Rectangular weighted SAS image

(b) Dolph-Chebyshev weighted SAS image

Figure 3.8: Comparison of rectangular weighted SAS image (a) with a Dolph-Chebyshev aperture weighted SAS image (b) using the windowing approach in the wavenumber domain.

A variety of different window functions exists [Harris (1978), Jakowatz et al. (1996), Van Trees (2002)], which have to be defined over the support band analog to (3.22). Typically, a Taylor window or a Dolph-Chebyshev window is of preferred choice. The effect of aperture shading is illustrated in Figure 3.8 on the basis of the man-made target as shown in Figure 3.2a. Both SAS images are depicted with a dynamic range of 50 dB. While the SAS image of the metal cross in Figure 3.8a is smeared in along-direction due to sidelobes of the strongly reflecting plastics balls, this effect has been reduced although still noticeable after applying a Dolph-Chebyshev weighting with a SLL attenuation of 40 dB. Moreover, the contrast of the shadow region right to the target highlight slightly increases. Unless stated otherwise, all SAS images shown in this thesis are weighted with a Dolph-Chebyshev window with a SLL attenuation of 40 dB, and the superscript $(\cdot)^{\mathrm{s}}$ in (3.21) is dropped for notational convenience.

3.5 Height map reconstruction

The introduced space-time image reconstruction technique of Section 3.3 has so far assumed perfect knowledge about transmitter and receiver positions along the entire synthetic aperture so as to avoid image degradation due to unknown motion. In practice sonar motion is not known perfectly and has to be estimated and accounted for in the SAS processing. In this thesis the unknown motion has been estimated using a technique called micronavigation as described in detail in Chapter 4. Nevertheless, even under any of these ideal conditions, image quality may worsen for nonlinear trajectories of the imaging platform in case the seafloor topography, *i.e.*, the bathymetry, is not available during image reconstruction [Jakowatz et al. (1996)]. Hence, besides gathering additional information about the seafloor scenery, bathymetry estimation is motivated by its importance for successful SAS focusing in strongly varying topographic environments [Hansen et al. (2011)].

3.5.1 Depth of focus

In order to assess the role of unknown bathymetry variations, a so-called depth of focus (DOF) criterion [Jakowatz et al. (1996), Hansen et al. (2011)] is investigated for ground-range imaging in the following. In this context, the DOF is defined as the maximum deviation between an *a priori* seafloor height grid $\hat{z}_{kl} \equiv \hat{z}(x_k, y_l)$ and the true spatially sampled bathymetry $z_{kl} \equiv z(x_k, y_l)$ that does not cause a notable defocus

in the SAS image for nonlinear trajectories. The DOF can be written as

$$\Delta z_{kl}^{\mathrm{DOF}} = \left| z_{kl} - \hat{z}_{kl} \right|, \quad \begin{array}{l} \text{for } k = 1, \ldots, M_x \\ \text{and } l = 1, \ldots, M_y, \end{array} \tag{3.23}$$

where the *a priori* seafloor height is usually set to $\hat{z}_{kl} \equiv 0$. In the following, a DOF criterion is derived for ground-range imaging similar to [Jakowatz et al. (1996), Hansen et al. (2011)]. Considering the geometry as depicted in Figure 3.9a, the range difference $\Delta r_{kl} = r'_{kl} - r_{kl}$ can be expressed using a first-order Maclaurin series expansion with $\Delta h_{kl}^{\mathrm{DOF}} \ll r_{kl}$ as follows

$$\Delta r_{kl} \approx \Delta h_{kl}^{\mathrm{DOF}} \sin(\Theta_0), \tag{3.24}$$

where Θ_0 represents the depression angle in the case of a zero-height seafloor. A similar expression of the range difference for the true seafloor height z_{kl} can be derived from the scenario shown in Figure 3.9b as follows

$$\Delta r\big(\Delta z_{kl}^{\mathrm{DOF}}\big) \approx \Delta h_{kl}^{\mathrm{DOF}} \sin(\Theta_z). \tag{3.25}$$

In (3.24) and (3.25), $\Delta h_{kl}^{\mathrm{DOF}}$ describes the height displacement of the imaging platform, which causes a maximum range difference for a given grid point $\boldsymbol{g}_{kl}$. It is found by

$$\Delta h_{kl}^{\mathrm{DOF}} = \max_p h_p^{\mathrm{og}} - \min_p h_p^{\mathrm{og}} \quad \forall\, p = \big\{ p \mid b_{\mathrm{phy}}(\theta_{kl,p}^{\mathrm{tx}}) = 1 \big\} \tag{3.26}$$

where h_p^{og} denotes the altitude coordinate of the transmitter position of ping p and $b_{\mathrm{phy}}(\theta_{kl,p}^{\mathrm{tx}})$ describes the ideal transmitter beampattern similar to (3.4).

Typically, image defocusing is avoided if the maximum range difference scaled by the wavelength is less than a certain phase error tolerance ϕ_{tol} along the synthetic aperture – see Section 4.1 for more details. This condition can be expressed as

$$\frac{4\pi \left| \Delta r_{kl} - \Delta r\big(\Delta z_{kl}^{\mathrm{DOF}}\big) \right|}{\lambda_c} \leq \phi_{\mathrm{tol}}. \tag{3.27}$$

Using the expressions of the range difference in (3.24) and (3.25) along with the relationships for the depression angles

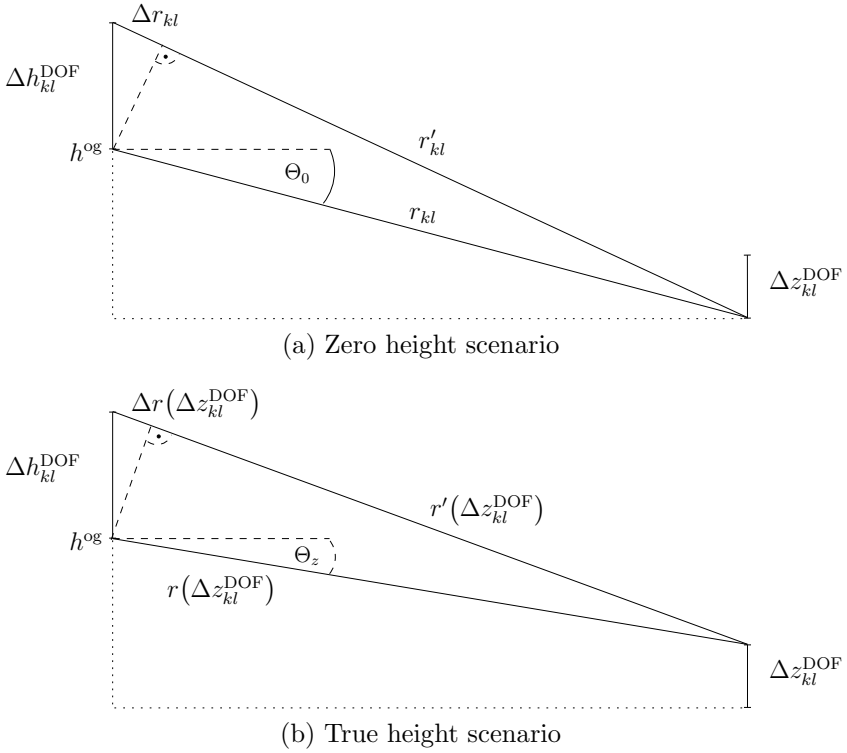
$$\sin \Theta_0 = \frac{h^{\mathrm{og}}}{r_{kl}} \tag{3.28}$$

and

$$\sin \Theta_z = \frac{h^{\mathrm{og}} - z_{kl}^{\mathrm{DOF}}}{r\big(z_{kl}^{\mathrm{DOF}}\big)} \tag{3.29}$$

and assuming that $r_{kl} \approx r\big(z_{kl}^{\mathrm{DOF}}\big)$, the DOF criterion [Hansen et al. (2011)] is given by

$$\Delta z_{kl}^{\mathrm{DOF}} \leq \frac{\lambda_c \, \phi_{\mathrm{tol}} \, r_{kl}}{4\pi \Delta h_{kl}^{\mathrm{DOF}}}. \tag{3.30}$$



(a) Zero height scenario

(b) True height scenario

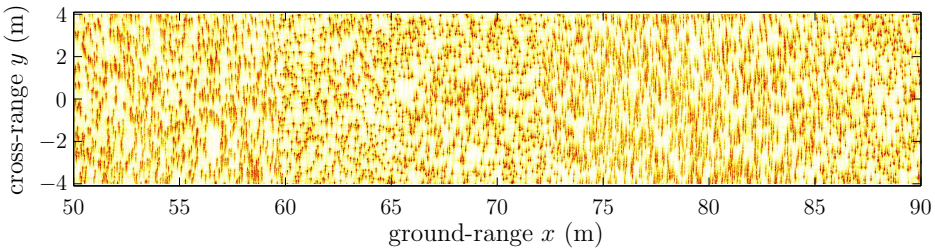
Figure 3.9: Geometry of DOF consideration – adapted from [Hansen et al. (2011)].

Hence, as long as the difference between the true bathymetry variation and the assumed height grid during image reconstruction fulfills the condition in (3.30), the resulting SAS image is focused. Otherwise, blurring may occur in regions where the condition is violated. In case of a straight and level trajectory, *i.e.*, $h_{kl}^{\mathrm{DOF}} \equiv 0 \,\forall\, k, l$, the DOF becomes infinite, and defocusing does not occur independent of the topography variation [Jakowatz et al. (1996)]. Furthermore, it should be remarked that the interaction between image reconstruction and motion estimation as introduced in Chapter 4 is partly capable of compensating blurring effects due to unknown height variations if both assume a flat *a priori* zero height grid [Sæbø (2010)]. The condition in (3.30) is then relaxed, and SAS images with reasonably high image quality can be reconstructed. A theoretical investigation of the interaction between image reconstruction and motion estimation as well as an adaptation of the condition in (3.30) according to this interaction remains subject of future work.

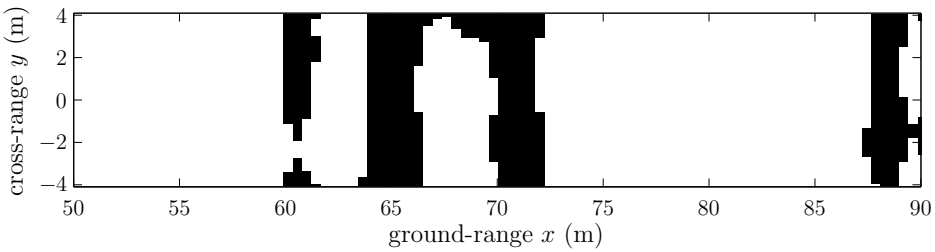
In the sequel, two synthetic data examples are illustrated where the entire trajectory of the imaging platform is assumed to be known and the chosen signal and array parameters are related to the VISION SAS system. However, bathymetry information about the scene is not available. The SAS images of both examples are shown with a dynamic range of 40 dB along with their DOF indicator maps in Figure 3.10 and Figure 3.11, respectively. The DOF maps indicate regions (white) in the SAS image

that suffer from blurring effects due to a deviation between the flat *a priori* height grid with $\hat{z}_{kl} \equiv 0 \; \forall \; k, l$ and the true sampled bathymetry variation z_{kl}. By contrast, the black regions indicate well focused image parts due to a fulfilled DOF condition. The phase tolerance has been set to $\phi_{\text{tol}} = \pi/2$ in both examples – see Figure 4.2 in Section 4.1. While the first SAS image in Figure 3.10a is severely defocused except for small regions of good quality, the second SAS image in Figure 3.11a is split into a good and poor image quality region. The corresponding bathymetry maps of the image scenes are depicted in Figure 3.13.

In order to overcome this defocusing issue, a sidescan bathymetry estimation technique is addressed in the following section to reconstruct a height grid $\hat{z}_{kl}$, which can be incorporated into the imaging method in an ensuing processing iteration. Moreover, the method is used in Chapter 4 to directly provide the micronavigation technique with seafloor height estimates in order to improve motion estimation as it also may suffer from unknown bathymetry. Prior to introducing bathymetry estimation, it has to be remarked that discrepancies occurred between the estimated bathymetry profiles and their corresponding DOF maps for real VISION data. This has not been explicitly shown in this section. According to the DOF prediction, defocusing should occur in the VISION SAS images, however, the resulting image quality is reasonably good for

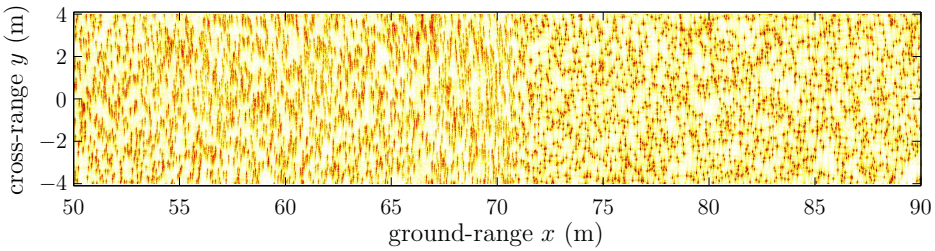


(a) SAS image reconstructed with a zero-height grid

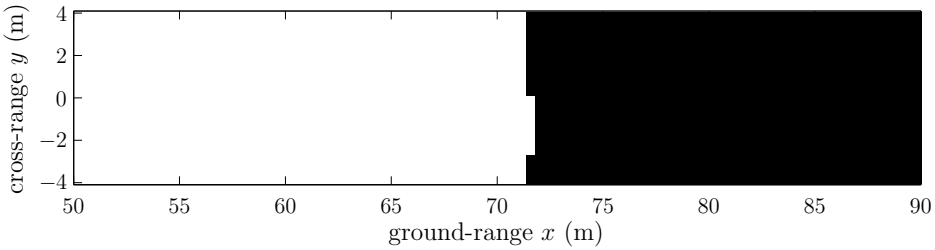


(b) DOF indicator map

Figure 3.10: First synthetic data example of SAS image defocusing (a) due to an unknown bathymetry profile during image reconstruction. Defocused SAS regions are predicted by a DOF binary map in (b), where the black regions indicate good quality. An estimate of the corresponding seafloor topography is depicted in Figure 3.13a.



(a) SAS image with zero-height grid



(b) DOF indicator map

Figure 3.11: Second synthetic data example of SAS image defocusing (a) due to an unknown bathymetry profile during image reconstruction. The defocused SAS regions are predicted by a DOF binary map (b), where the black region indicates good quality. An estimate of the corresponding seafloor topography is depicted in Figure 3.13b.

an *a priori* zero-height grid. The reason for this is the aforementioned interaction between motion estimation and imaging, which partly compensates for the defocusing effect. The corresponding SAS image of the VISION system is illustrated in Chapter 4 after discussing the involved steps for motion estimation.

3.5.2 Broadside bathymetry estimation

Broadside bathymetry estimation, also known as interferometry [Griffiths et al. (1997), Bonifant et al. (2000), Sæbø (2010)], uses single beam signals of two vertically displaced sonar arrays to estimate the seafloor height variation. Typically, two beams are formed by dynamically focusing the echo signals of the bottom and top array individually in slant-range direction, followed by an estimation of their time delay differences as a function of range. Given the spatial displacement between both arrays, a height estimate is obtained by a geometrical conversion of the corresponding time delay difference. In Figure 3.12, a common interferometric setup is depicted. It shows the transmitter located in between two reference points of the top array and bottom array, namely, $\boldsymbol{a}_1^{\text{ref}}$ and $\boldsymbol{a}_2^{\text{ref}}$. Note that the array dimension coincides with the y-axis and is, therefore, not illustrated. As broadside bathymetry estimation is a single ping based coherent processing technique, the ping index p is not considered in this section.

Given the received echo signals $e_i(u,n)$ with $i = 1,2$ for the bottom and top array, respectively, where each array consists of $u = 1,\ldots,N_{\mathrm{rx}}$ receiving elements, the first processing step is to beamform the echo signals with respect to their reference point. Typically, the latter is chosen as the array center. In order to form a beam in azimuth direction θ, a focus delay is required as follows

$$\Delta t_i^{\mathrm{B}}(\theta, r; \boldsymbol{a}_i^{\mathrm{rx}}(u), \boldsymbol{a}^{\mathrm{tx}}, \boldsymbol{a}_i^{\mathrm{ref}}) = \frac{r_i^{\mathrm{rx}}(\theta, r; u) - r_i^{\mathrm{ref}}(\theta, r)}{c} \tag{3.31}$$

where $r_i^{\mathrm{rx}}(\theta, r; u) = \left\| \boldsymbol{a}^{\mathrm{tx}} + \boldsymbol{g}^{\mathrm{fp}}(\theta, r) - \boldsymbol{a}_i^{\mathrm{rx}}(u) \right\|_2$ and $r_i^{\mathrm{ref}}(\theta, r) = \left\| \boldsymbol{a}^{\mathrm{tx}} + \boldsymbol{g}^{\mathrm{fp}}(\theta, r) - \boldsymbol{a}_i^{\mathrm{ref}} \right\|_2$ denote the slant-range of receiver u and of the reference point, respectively. In both range expressions, $\boldsymbol{g}^{\mathrm{fp}}(\theta, r) = [r\cos\theta, r\sin(\theta), 0]^T$ describes the focusing vector in slant-range dimension. Applying the focusing delay of (3.31) in broadside direction with $\theta = \theta_0$ to the echo signals $e_i(u,n)$, with $i = 1,2$, and coherently summing over the receiver dimension yields the beam signals $e_1^{\mathrm{B}}(\theta_0, n)$ and $e_2^{\mathrm{B}}(\theta_0, n)$ of the bottom and top array, respectively. On the basis of these beam signals, it is feasible to estimate the height profile of a narrow range strip in broadside direction. To this end, a short-time windowing approach for time delay estimation (TDE) is applied to obtain estimates of $\Delta\tau(k_{\mathrm{s}})$, which are then scaled by the sound speed c to estimate the range difference $\Delta\hat{r}(k_{\mathrm{s}})$ of a single short-time window with index k_{s}. The latter represents an estimate of the true range difference $\Delta r(r_1^{\mathrm{ref}})$ as exemplarily depicted in Figure 3.12. It should be remarked that each index k_{s} is related to the center range of the respective window. In general, the short-time windowing approach is required to handle the non-stationarity of the time delay differences. The reader is referred to Section 4.2.3 and Section 4.2.5, which provide more details on time delay estimation and beamforming, respectively.

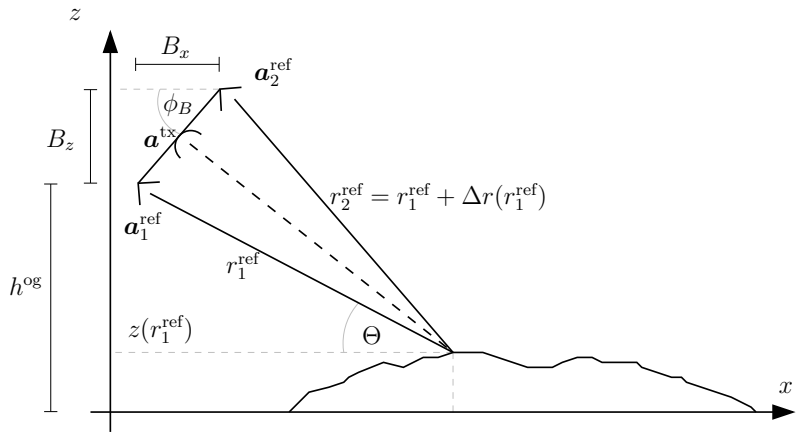


Figure 3.12: Illustration of an interferometry setup for estimating the seafloor height showing the baseline relation between the positions of the reference points of the bottom ($\boldsymbol{a}_1^{\mathrm{ref}}$) and top array ($\boldsymbol{a}_2^{\mathrm{ref}}$). Given the depicted geometry, a range difference $\Delta r(r_1^{\mathrm{ref}})$ is induced into the echo signals, which corresponds to the seafloor height $z(r_1^{\mathrm{ref}})$.

Considering again the geometry of Figure 3.12, the seafloor height $z(r_1)$ with $r_1 \equiv r_1^{\text{ref}}$ is determined based on the slant-range difference $\Delta r(k_{\text{s}})$ and known baseline parameters relating both reference points with each other [Hein (2003)]. These parameters are the baseline distance B and the baseline angle ϕ_B, which are given by

$$B = \sqrt{B_x^2 + B_z^2} \quad \text{and} \quad \phi_B = \operatorname{atan}\left(\frac{B_z}{B_x}\right), \tag{3.32}$$

where B_x and B_z denote the horizontal and vertical baseline displacements, respectively. In situations where the imaging platform is rolling during forward motion, the current roll angle ϕ_{roll} has to be considered – see Figure 4.1 in Section 4.1 for a description of motion degrees of freedom. Otherwise, a systematic error is induced into the calculation of the seafloor height [Griffiths et al. (1997), Bonifant et al. (2000)]. The seafloor height is then geometrically derived with respect to the altitude h^{og} of the reference point $\boldsymbol{a}_1^{\text{ref}}$ similar to [Hein (2003)] as follows

$$z(r_1) = h^{\text{og}} - r_1 \sin\big(\Theta(r_1, \Delta r(r_1); \phi_{\text{roll}})\big) \tag{3.33}$$

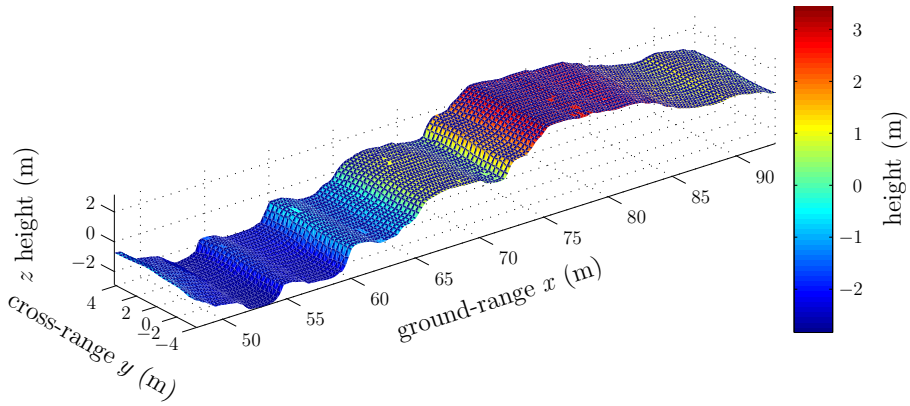
with

$$\Theta\big(r_1, \Delta r(r_1); \phi_{\text{roll}}\big) = \operatorname{acos}\left(\frac{r_1^2 + B^2 - \big(r_1 + \Delta r(r_1)\big)^2}{2Br_1}\right) - \big(\phi_B + \phi_{\text{roll}}\big). \tag{3.34}$$

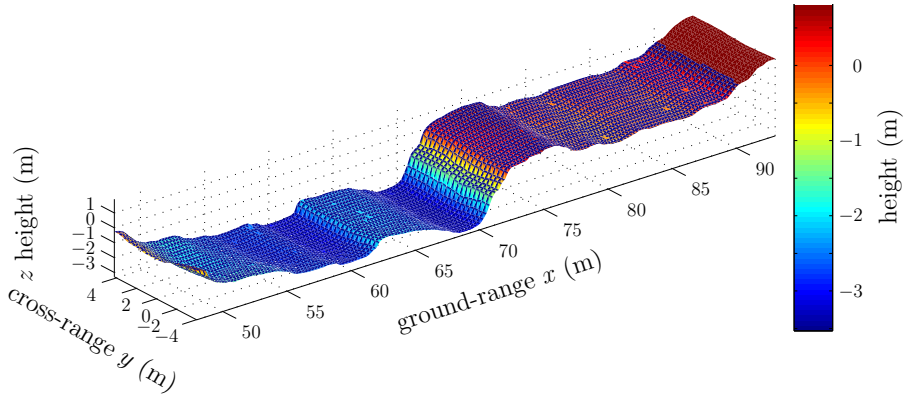
Here, $\Theta(r_1, \Delta r(r_1); \phi_{\text{roll}})$ describes a varying depression angle. In case of roll-free motion, *e.g.*, for synthetic data, the roll angle is set to $\phi_{\text{roll}} = 0$. Two synthetic data examples of estimated height profiles of artificial seafloors are depicted in Figure 3.13. They represent an estimate of the underlying bathymetry variation of the artificial target scenes for which the SAS images in Figure 3.10a and Figure 3.11a have been reconstructed. While the bathymetry profile in Figure 3.13a only consists of three narrow cross-range stripes, namely, around $x_1 \approx 65$ m, $x_1 \approx 70$ m and $x_1 \approx 90$ m, where the seafloor height is roughly $z \approx 0$ m, the profile in Figure 3.13b features two plateaus. The second plateau is located in a ground-range interval of $70\,\text{m} \leq x \leq 90\,\text{m}$ and has approximately a zero height level. For both bathymetry profiles, the mentioned regions are in accordance with the focused parts of the SAS images and their corresponding DOF maps. Rerunning the SAS processing chain with an estimated height grid in a second iteration only leads to focused imagery in case the platform motion is known. The combination of using height grids for space-time imaging and, simultaneously, providing the motion compensation technique with the estimated bathymetry is discussed in Section 4.2.7. Moreover, the focused counterpart of the SAS image in Figure 3.10a is depicted in Figure 4.23 of that section.

An estimated bathymetry profile using real sonar measurements is illustrated in Figure 3.14a for $M_p = 80$ processed pings and a ground-range swath between

$X_{\min} = 60$ m and $X_{\max} = 110$ m. The height variation is approximately $z = 1$ m. For bathymetry estimation, only time delay estimates with a correlation value larger than $\rho_{\min} = 0.67$ have been considered to guarantee a sufficiently high signal-to-noise ratio (SNR), and therefore, reliable height estimates. In Figure 3.14a, the absence of blank spots in the estimated bathymetry profile indicates a high correlation between the beam signals of the bottom and top array. Although the roll of the imaging platform is incorporated in (3.33), sinusoidal height variations are still observable in cross-range direction (ping direction) of the depicted bathymetry profile that stem from a continuous movement of the AUV. In the case of a significant change in the roll angle between consecutive transmission times, the applied stop-and-hop assumption in (3.33) is violated. Therefore, even small roll differences may cause a noticeable height estimation error due to the significantly long lever arm, *i.e.*, for a long imaging range [Griffiths et al. (1997), Bonifant et al. (2000)]. As a consequence, the continuous roll movement has to be considered in (3.33) rather than a fixed roll angle during transmission time. This is achieved by replacing ϕ_{roll} with $\phi_{\text{roll}}(t)$.

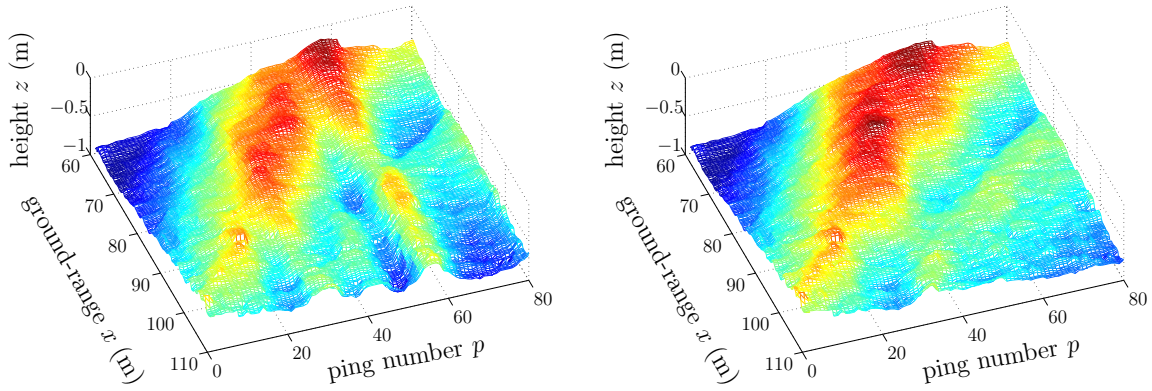


(a) Bathymetry profile of the SAS image shown in Figure 3.10a



(b) Bathymetry profile of the SAS image shown in Figure 3.11a

Figure 3.13: Illustration of estimated bathymetry profiles of a synthetic seafloor using the broadside bathymetry estimation technique.



(a) Without roll compensation

(b) With roll compensation

Figure 3.14: Illustration of an estimated height profile of the seafloor using the broadside bathymetry estimation technique without continuous roll compensation (a) and with roll interpolation (b). The raw sonar measurements have been recorded by the VISION SAS system.

A straightforward approach to obtain $\phi_{\text{roll}}(t)$ without taking into account additional samples of an INS unit is via interpolation. In the case of stop-and-hop bathymetry estimation, a roll angle $\phi_{\text{roll}}(t_p)$ is already available for each transmission time t_p with $p = 0, \ldots, M_p - 1$. Thus, using linear interpolation, the change in roll angle with respect to a given fast-time value t (slant-range value r) can be denoted by

$$\hat{\phi}_{\text{roll}}(t) = m_{\text{roll}}\, t + b_{\text{roll}} \quad \text{with} \quad m_{\text{roll}} = \frac{\phi_{\text{roll}}(t_{p+1}) - \phi_{\text{roll}}(t_p)}{t_{p+1} - t_p} \tag{3.35}$$

and $b_{\text{roll}} = \phi_{\text{roll}}(t_p) - m_{\text{roll}}\, t_p$. Provided that short-time windows are used to estimate time delays, only the time samples t_{k_s} with $k_\text{s} = 1, \ldots, K$ are available, representing the center of a sliding short-time window of length M_k. Hence, a roll angle interpolation is only required for the time instants given by

$$t_{k_\text{s}} = T_s \left(\left(k_\text{s} - 1\right) M_k + \frac{M_k - 1}{2} \right), \tag{3.36}$$

where M_k is assumed an odd sample number. Substituting $\hat{\phi}_{\text{roll}}(t_{k_\text{s}})$ into (3.33) yields a continuous roll compensated bathymetry profile as shown in Figure 3.14b, where the sinusoidal height variations are clearly reduced. Although the INS typically samples the continuous movement of the AUV more than once per ping, a single roll value combined with linear interpolation is sufficient to overcome the stop-and-hop effects. In order to reconstruct bathymetry profiles that are directly applicable as a height grid for SAS imaging, all pings of available sonar measurements have to be processed prior to SAS processing using the introduced technique. Alternatively, a sequential stitching technique based on RAS bathymetry estimation is feasible to instantly build a height

grid [Kronig (2014)], which is usable for SAS imaging without the need of a second processing iteration.

3.6 Conclusions

In this chapter, a signal model extension as well as an space-time image reconstruction technique for SAS systems have been considered assuming perfect motion knowledge. After the introduction of a real sonar system mounted onto an AUV, a technique for sequentially constructing a ground-range grid has been proposed to be used with the space-time imaging method. Moreover, a method called synthetic aperture shading has been described. It trades off resolution with image contrast by controlling the mainlobe and sidelobes of point scatterers via apodization with a window function to improve image quality. The effect of synthetic aperture shading has been illustrated for a real data example of a ground-range SAS image, which has been reconstructed using the proposed processing chain. Afterwards, the necessity of knowing the underlying bathymetry of a target scene is highlighted based on synthetic data in order to obtain focused SAS images in environments with a strongly varying topography. An agreement with a theoretical DOF criterion has been demonstrated in the case of synthetic data. The criterion predicts the occurrence of image blurring due to a flat bottom assumption during ground-range imaging.

Further, it has been pointed out that a significant ping-to-ping roll severely affects the estimated bathymetry profile. This effect has been compensated using a linear interpolation of available stop-and-hop roll angle values leading to a significant improvement. Besides height grid construction, the introduced broadside bathymetry technique is a preprocessing stage that is used to provide height estimates for motion compensation. Therefore, broadside bathymetry estimation has to solely rely on position information provided by an INS unit that may be another potential source of biased height estimates. In order to cope with unknown height displacements of the imaging platform, a procedure is required that switches iteratively between broadside bathymetry estimation and motion estimation. Furthermore, additional bathymetry estimation techniques should be employed to facilitate a sequential reconstruction of focused ground-range SAS images without applying an entire second processing routine. Finally, a validation of improving image quality is still ongoing for real sonar measurements recorded under harsh bathymetric situations.

Chapter 4

Data-driven compensation techniques

In real-life scenarios, the reconstruction of high-resolution synthetic aperture sonar (SAS) images demands precise knowledge of the exact sensor location at each transmission and reception time along the entire synthetic aperture. Otherwise, the mismatch between focusing delay and true round-trip delay yields a distorted SAS image if a series of real aperture sonar (RAS) images is coherently combined. Typically, an accuracy of more than a tenth of a wavelength [Jakowatz et al. (1996)] in lateral range direction is required to obtain focused non-degraded high-quality synthetic aperture imagery. However, current on-board inertial navigation systems (INS) used in autonomous underwater vehicles (AUVs) do not have sufficient accuracy to track translational motion deviations from an ideal trajectory. Furthermore, an INS is not able to account for medium turbulences [Hayes and Gough (2009)]. Consequently, additional means in terms of data-driven approaches are employed to meet the stringent requirements. For example, the displaced phase center antenna (DPCA) [Bellettini and Pinto (2002)] is a micronavigation technique making use of the common multi-receiver configuration in SAS by exploiting the temporal and spatial coherence of the seafloor backscattering. The DPCA estimates the true sensor position by evaluating time delays between received echo signals corresponding to redundant phase centers (RPC) [Bellettini and Pinto (2002), Cook et al. (2006)] of two consecutive transmission times. Therefore, a sequential motion estimate can be obtained on a ping-to-ping basis that is directly usable in space-time image reconstruction such as the backprojection algorithm.

In addition to micronavigation techniques, autofocus techniques exist, which estimate the motion-induced phase error based on the reconstructed synthetic aperture image data. Typically, they can be categorized according to the operation mode of the synthetic aperture system, namely, spotlight or stripmap. While spotlight autofocus techniques such as the phase gradient autofocus (PGA) algorithm are advantageous due to their direct applicability in the image domain, its stripmap counterpart works iteratively. It switches between image data and raw echo measurements as the point spread functions (PSF) are no longer spatially invariant. Hence, stripmap PGA leads to a high computational load. In this chapter, a translational motion error model is introduced in Section 4.1. Thereupon, the data-driven motion compensation technique is explained in detail in Section 4.2 along with proposed extensions such as a correction technique to handle phase wrap errors and the integration of bathymetry information.

Afterwards, a proposed autofocusing technique is addressed in Section 4.3 that aims at overcoming existing problems when mosaic autofocus techniques are applied to stripmap SAS images. Finally, Section 4.4 describes a calibration technique to estimate the sound speed based on a quality assessment of SAS image data.

4.1 Motion error model

The advantage of maintaining a constant cross-range resolution of synthetic aperture systems comes along with the prerequisite of subwavelength accuracy on the position of the transmitter and receiver array during transmission and reception time. Otherwise, phase errors are induced, which cause image degradation. The phase errors originate from a mismatch between the focusing delay $\tau_{kl,p}(u)$ used inside the image reconstruction for each grid point $\boldsymbol{g}_{kl}$ of the reconstructed target scene and the true round-trip delay $\tau_{d,p}(u)$ of the echo data. While current INS units are capable of accurately measuring the rotational motion (yaw, pitch and roll) of an AUV [Cook and Brown (2009)], they lack precision for translational motion. The latter are characterized by the ground-range sway Δa_p^{s}, the surge Δa_p^{y}, and the heave Δa_p^{h}. All six motion degrees of freedom are illustrated in Figure 4.1 for a rigid body in 3-D space.

In the sequel, the error displacement vector $\Delta\boldsymbol{a}_p = [\Delta a_p^{\mathrm{s}}, \Delta a_p^{\mathrm{y}}, \Delta a_p^{\mathrm{h}}]^T$ as a function of the current ping p models the translational path deviations from an ideal trajectory presumed by the INS unit. As rotational motion errors are not considered, the error displacement vector $\Delta\boldsymbol{a}_p$ is independent of the receiver index u. Note that this model assumes a stop-and-hop scenario [Cook (2007)], which has been found to be accurate enough to describe the unknown platform motion. Given the unknown positions $\Delta\boldsymbol{a}_p$, the true round-trip delay $\tau_{d,p}(u)$ of (3.2) changes to

$$\tilde{\tau}_{d,p}(u) = \frac{\left\|\boldsymbol{q}_d - (\boldsymbol{a}_p^{\mathrm{tx}} + \Delta\boldsymbol{a}_p)\right\|_2 + \left\|\boldsymbol{q}_d - (\boldsymbol{a}_p^{\mathrm{rx}}(u) + \Delta\boldsymbol{a}_p)\right\|_2}{c}. \tag{4.1}$$

By contrast, the focusing delay $\tau_{kl,p}(u)$ remains as stated in (3.8) using the ideal transmitter and receiver locations with $\boldsymbol{a}_p^{\mathrm{tx}}$ and $\boldsymbol{a}_p^{\mathrm{rx}}(u)$, respectively. This leads to a time delay difference $\Delta\tau_{d,p}(u) = \tilde{\tau}_{d,p}(u) - \tau_{d,p}(u)$ that causes image degradation during the reconstruction process. In order to relate the induced phase error to the time delay difference $\Delta\tau_{d,p}(u)$, the time-domain correlation (TDC) method [Soumekh (1999)] is considered for reconstructing a synthetic aperture image. For ease of notation, the single transceiver model of Chapter 2 is used and the pulse-compressed echo data is written as $e(p,n) \equiv e_p(n)$, with $p = 0,\ldots,M_p-1$ and $n = 0,\ldots,M_n-1$, where M_p and M_n denote the number of slow-time and fast-time samples, respectively. The

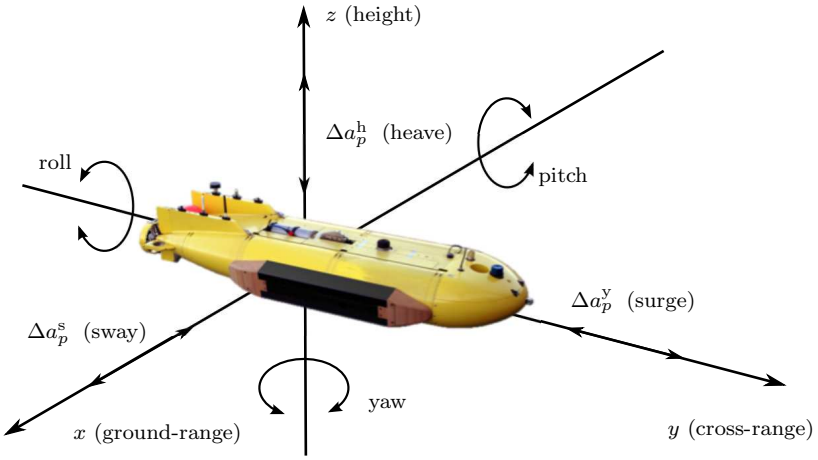


Figure 4.1: Six motion degrees of freedom of a rigid body in 3-D space.

TDC reconstruction technique correlates the received phase history data with replicas of the transmitted signal $s(n - \eta_{kl}(p))$. To this end, a single replica with delay $\eta_{kl}(p)$ is constructed for each grid point $\boldsymbol{g}_{kl} = [x_k, y_l]^T$ as a function of the ping index p (cross-range transceiver position). The reconstruction rule of the TDC method then reads

$$\hat{f}(x_k, y_l) = \sum_p \sum_n e\big(p, n\big)\; s\big(n - \eta_{kl}(p)\big)^* \qquad (4.2)$$
$$\text{with} \;\; k = 1, \ldots, M_x \;\; \text{and} \;\; l = 1, \ldots, M_y.$$

The expression in (4.2) can be interpreted as a spatially varying 2-D correlation, although the reference signal is not directly shifted in ground-range dimension x_k and cross-range dimension y_l but rather via the delay function $\eta_{kl}(p)$. As a consequence of this, both dimensions are coupled, which is known as range migration [Richards (2005)]. Assuming a sufficiently small or already compensated range migration [Bamler (1992), Gough and Hawkins (1997)], separate 1-D processing is feasible. Following the derivation in [Cook and Brown (2009)], and assuming already pulse-compressed echo signals, the remaining cross-range focusing operation is a 1-D correlation between the received phase history data and a reference signal $s(p)$ as a function of the ping index p (cross-range transceiver position). Limiting this operation to a single range line n_1, which coincides with the ground-range position $\boldsymbol{q}_1 = [x_1, 0]^T$ of a single point target with unit reflectivity such that the echo signals with $e(p, n_1) \equiv e(p)$ contain the entire target signature, the reconstruction process ignoring amplitude considerations can be described for a single frequency by [Cook and Brown (2009)]

$$\hat{f}(x_1, y_l) = e(p) *_p s(-p)^* = \sum_p e^{-j\omega_c \tilde{\eta}(p)}\; e^{j\omega_c \eta(y_l - p)}, \qquad (4.3)$$

with $e(p) = e^{-j\omega_c\tilde{\eta}(p)}$ – see also (2.14), and $\eta(p) = 2\sqrt{x_1^2+p^2}/(cT_s)$. In (4.3), ω_c denotes the discrete-time angular carrier frequency, the delay function $\eta(p)$ describes a symmetric function, and only those pings p are processed for which the single point target is observed. Provided that the true target delay $\tilde{\eta}(p) = \tilde{\tau}(p)/T_s$ arising from unknown platform positions is given by an ideal delay term plus some unknown delay term as follows $\tilde{\eta}(p) = \eta(p) + \Delta\eta(p)$, the 1-D reconstruction can be reformulated as

$$\begin{aligned}\tilde{f}(x_1, y_l) &= \sum_p e^{-j\omega_c(\eta(p)+\Delta\eta(p))}\; e^{j\omega_c\eta(y_l-p)} \\ &= \sum_p e^{-j\omega_c\eta(p)}\; e^{-j\phi_{\mathrm{err}}(p)}\; e^{j\omega_c\eta(y_l-p)}, \text{ with } l = 1,\ldots,M_y. \end{aligned} \tag{4.4}$$

In (4.4), the phase error term $\phi_{\mathrm{err}}(p) = \omega_c\,\Delta\eta(p)$ is multiplied with the ideal phase history data before performing the correlation (matched-filter) operation in cross-range direction. Hence, this results in a phase mismatch that spreads the ideal PSF as depicted in Figure 4.2. Typically, a maximum phase error of $\left|\phi_{\mathrm{err}}(p)\right| \leq \phi_{\mathrm{tol}}\ \forall\ p$ with $\phi_{\mathrm{tol}} = \pi/2$ is still acceptable for reasonable image quality [Hansen et al. (2011)] although the sidelobe level (SLL) is already increased by approximately 3 dB. However, the latter can be compensated by an appropriate synthetic aperture shading as outlined in Section 3.4.

Assuming the imaging geometry of a common operating situation of a SAS system, *i.e.*, relatively small altitude (to avoid wave transmission through various water columns) and broad swath width, the amplitude value of the ground-range sway error Δa_p^{s} is most critical to obtain good image quality. However, the actual degradation effect

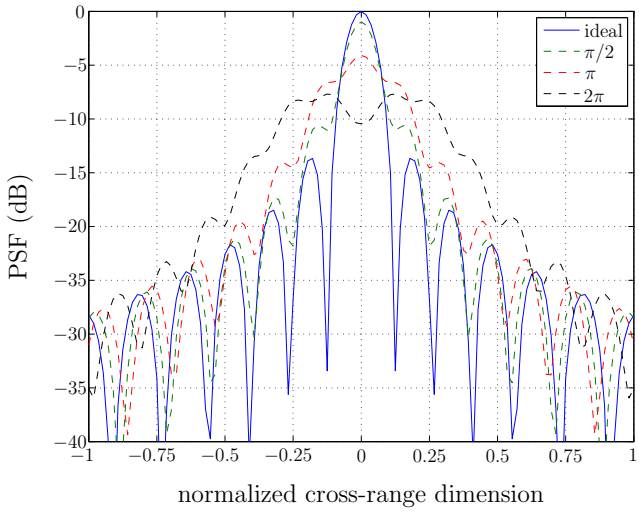


Figure 4.2: Illustration of PSF degradation for different maximum phase errors – after [Cook and Brown (2009)].

is not determined by the degree of motion, *e.g.*, ground-range sway or heave, but rather by the functional behavior of the round-trip delay mismatch $\Delta\tau_{d,p}(u)$ along the synthetic aperture. A thorough analytical study is provided by [Fornaro (1999), Cook and Brown (2009)]. In the following, the focus lies on sinusoidal path deviations caused by AUV motion and quadratic phase errors caused by an incorrect measurement of the sound speed along the acoustical path. While the sinusoidal path deviations are interesting due to their frequent occurrence in practice, sound speed phase errors arise if the sound-speed profile varies along the acoustical path, keeping in mind that water is an inhomogeneous medium. Both error types are detailed subsequently as they are used throughout the thesis to intentionally degrade the SAS reconstruction process. This is done to either validate proposed compensation techniques or to study the influence of image quality on automatic target recognition (ATR) systems – see Chapter 5.

4.1.1 Sound speed errors

A common source of defocus in SAS imagery is due to a mismatch between a measured sound speed value and the true sound speed, which is denoted by c_0 in the following [Hansen et al. (2007), Hansen et al. (2011)]. This mismatch occurs since sound speed may vary spatially with temperature, salinity, pressure and density [Lurton (2002)]. Unfortunately, it can only be measured easily in the vicinity of the AUV position. In order to investigate its effect on SAS images, the maximum difference in the round-trip delay along the synthetic aperture has to be determined. It occurs between the broadside position $\boldsymbol{a}(\theta_0) = [0,0,0]^T$ to the target located at $\boldsymbol{q} = [x,0,0]^T$, and the last synthetic aperture position $\boldsymbol{a}(\theta_{\mathrm{BW}}/2) = [0,y,0]^T$ for which the target is seen. For simplicity, a transceiver system is considered here. Assuming an erroneous sound speed $\tilde{c} = c_0 + \Delta c$ with a mismatch of Δc, the maximum delay difference [Hansen et al. (2007)] can then be written as

$$\Delta\tau_{\max}(x,y;\Delta c) = 2\Big(\big\|\boldsymbol{a}(\theta_0)-\boldsymbol{q}\big\|_2 - \big\|\boldsymbol{a}(\theta_{\mathrm{BW}}/2)-\boldsymbol{q}\big\|_2\Big)\left(\frac{1}{c_0}-\frac{1}{c_0+\Delta c}\right). \tag{4.5}$$

where Δc is assumed to be an average constant offset along the acoustical pathway. An expression for the phase error $\phi_{\mathrm{err}}(x,y;\Delta c) = \omega_c\,\Delta\tau_{\max}(x,y;\Delta c)$ can then be derived using a first-order Maclaurin series approximation of the range difference expression in (4.5) as follows [Cook and Brown (2009)]

$$\phi_{\mathrm{err}}(x,y;\Delta c) \approx \frac{4\pi}{\lambda_c}\left(x+\frac{y^2}{2x}\right)\frac{\Delta c}{c_0+\Delta c}. \tag{4.6}$$

In (4.6), two effects become apparent. First the error of the sound speed causes a linear shift of the target scene in ground-range direction, and second, the quadratic

along-track term affects the cross-range compression as discussed in Section 4.1. This leads to a spreading of the point scatterer response and, consequently, to a blurring of the SAS imagery [Jakowatz et al. (1996)] that is proportional to the target range [Hansen et al. (2007)]. This degradation effect is used in Section 4.4 and Chapter 5 to verify a proposed compensation technique and for studying its influence on ATR systems. However, in order to avoid any approximations made in the derivation of the phase error, the true sound speed c_0 is typically modified inside the SAS processing chain for the purpose of image blurring to

$$\tilde{c} = (1 + q_\epsilon)\, c_0. \tag{4.7}$$

Here, $q_\epsilon = \Delta c / c_0$ represents a mismatch ratio. In general, the mismatch value is in the range of $\pm 2\%$ in practical scenarios [Hansen et al. (2011)], however, smaller variations are more common.

4.1.2 Sinusoidal path deviation

Often translational path deviations that occur in practice can be described by a sinusoidal function or a superposition of different sinusoidal functions [Fornaro (1999), Cook and Brown (2009)]. Consequently, the common model used within this thesis for translational ground-range (s - sway) and height (h - heave) motion errors with $\epsilon \in \{s, h\}$ is given by

$$\Delta a_p^\epsilon = A^\epsilon \sin(2\pi f_\mathrm{p}\, p),\ p = p_0, \ldots, p_1. \tag{4.8}$$

Typically, the motion amplitude A^ϵ is given in terms of the carrier wavelength λ_c, and f_p denotes the cycles per synthetic aperture length frequency [Fornaro (1999)] over $p = p_0, \ldots, p_1$ transmission times. A sinusoidal phase error leads to replicas of the ideal PSF, which are weighted and spaced as a function of A^ϵ and f_p, respectively. By choosing small values of f_p, a blurring can be induced in the reconstructed SAS image. For more details on deriving the phase error expression in case of sinusoidal motion, the reader is referred to [Fornaro (1999), Cook and Brown (2009)].

4.2 Displaced phase center antenna algorithm

In order to overcome the inaccuracies of current INS units and to avoid a defocusing during the reconstruction of high-resolution SAS images, additional data-driven means

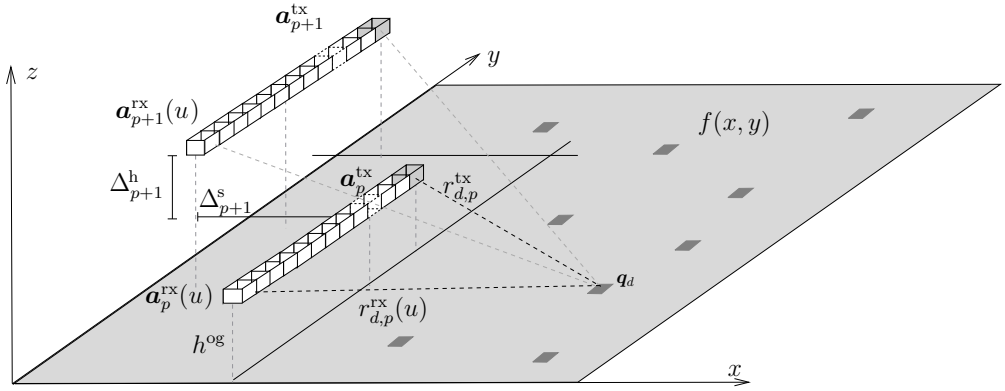


Figure 4.3: Geometry of a SAS system with a single-transmitter and multi-receiver configuration for two transmission times, t_p and t_{p+1}. The scene of interest $f(x, y)$ consists of point targets at positions $\boldsymbol{q}_d$, $d = 1, \ldots, D$. For ping $p + 1$, the sonar is displaced in x (ground-range) and z (height) directions by $\Delta^{\mathrm{s}}_{p+1}$ and $\Delta^{\mathrm{h}}_{p+1}$ with respect to ping p. The nominal altitude of the imaging system is given by h^{og}.

have to be utilized to estimate the true sensor positions along the synthetic aperture. These estimated positions are directly applicable in space-time image reconstruction techniques, *e.g.*, the backprojection algorithm, to determine a correct focusing delay. The displaced phase center antenna (DPCA) technique [Bellettini and Pinto (2002)] is a data-driven micronavigation approach, which sequentially operates on the echo data of two consecutive pings taking into account the position information of the INS. Its fundamental idea is based on the temporal and spatial coherence of the seafloor backscattering. By constructing phase center arrays and determining their redundant elements, the spatial and temporal correlation structure in the echo data is exploited to estimate translational motion errors on a ping-to-ping basis. Therefore, the DPCA avoids an iterative procedure as opposed to classical stripmap autofocusing techniques. This is beneficial in terms of computational load.

A common geometrical scenario for the displacement of a single-transmitter and multi-receiver SAS system of two consecutive pings is illustrated in Figure 4.3. Assuming that the ideal traveling path of the synthetic aperture is located along the extension of the rectilinear trajectory of the array at ping p, the motion between ping p and $p + 1$ can be modeled by $\boldsymbol{\Delta}_{p+1} = [\Delta^{\mathrm{s}}_{p+1}, \Delta^{\mathrm{y}}_{p+1}, \Delta^{\mathrm{h}}_{p+1}]^T$, which can be interpreted as a velocity vector. Hence, the ping-to-ping motion has to be integrated in order to get the absolute path deviation $\Delta\boldsymbol{a}_{p+M_p}$ from the assumed ideal trajectory at ping $p+M_p+1$. An estimate of the absolute array displacement with respect to the ideal trajectory is consequently given by

$$\Delta\hat{\boldsymbol{a}}_{p+M_p} = \left[\sum_{p'=p}^{p+M_p} \hat{\Delta}^{\mathrm{s}}_{p'}, \sum_{p'=p}^{p+M_p} \hat{\Delta}^{\mathrm{y}}_{p'}, \sum_{p'=p}^{p+M_p} \hat{\Delta}^{\mathrm{h}}_{p'} \right]^T, \tag{4.9}$$

where $p' = p$ denotes an initial transmission index with $\boldsymbol{\Delta}_p = [0, 0, 0]^T$. The latter is arbitrarily set, as the first available ping-to-ping motion estimate is obtained for ping $p + 1$. Note that in Figure 4.3 perfect knowledge about the along-track displacement has been assumed, *i.e.*, $\Delta a_p^{\mathrm{y}} = 0 \; \forall \; p$.

A general overview of the DPCA micronavigation technique is depicted in the block diagram in Figure 4.4. It illustrates that the DPCA method operates on the echo data $e_p(u, n)$ and $e_{p+1}(u, n)$. Furthermore, it is highlighted that the corresponding position information provided by the INS is considered, and that the outcome is an estimate of the absolute position deviation $\Delta\hat{\boldsymbol{a}}_{p+1}$. The technique itself comprises five major blocks, namely, the determination of phase center pairs, surge estimation, and near-field and widebeam corrections followed by a time delay estimation procedure. Based on the obtained time delays, the ping-to-ping ground-range sway $\Delta_{p+1}^{\mathrm{s}}$ and heave $\Delta_{p+1}^{\mathrm{h}}$ motion is determined. Each of the individual processing blocks is addressed in detail in the following sections.

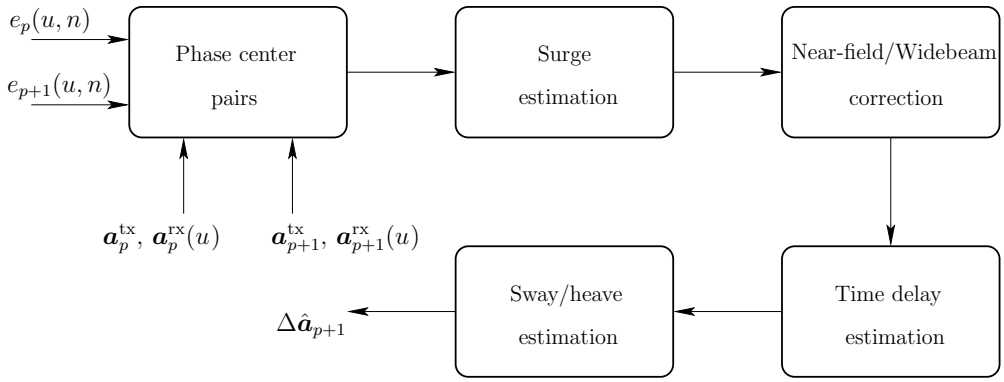


Figure 4.4: Overview diagram of the DPCA micronavigation technique.

4.2.1 Phase center arrays

The concept of a phase center describes the point on an aperture from which the transmitted signal seems to originate given its far-field phase curvature. Therefore, the phase center describes the best geometrical reference point for the recorded phase history data [Carrara et al. (1995)]. In the case of a spatially separated transmit and receive aperture, the phase center can be considered as a virtual transceiver that is located at the geometric center of both physical apertures [Bellettini and Pinto (2002)]. Replacing all receiving elements by their corresponding phase centers leads to a phase

center array, which is located at

$$\boldsymbol{a}_p^{\mathrm{pc}}(u) = \frac{\boldsymbol{a}_p^{\mathrm{tx}} + \boldsymbol{a}_p^{\mathrm{rx}}(u)}{2}, \quad \text{with} \quad u = 1, \ldots, N_{\mathrm{rx}}. \tag{4.10}$$

In order to estimate motion, the DPCA exploits the redundancy between overlapping phase centers of two consecutive pings. These so-called redundant phase center (RPC) pairs with index $u_{ij} = (u_i, u_j)$ are defined to be displaced only in slant-range direction. Mathematically, this is denoted by

$$\boldsymbol{a}_{p+1}^{\mathrm{pc}}(u_j) = \boldsymbol{a}_p^{\mathrm{pc}}(u_i) + [\Delta_{p+1}^{\mathrm{s}}, 0, \Delta_{p+1}^{\mathrm{h}}]^T, \tag{4.11}$$

with $u_i, u_j \in \{1, \ldots, N_{\mathrm{rx}}\}$. Note that index u_j is typically determined by the phase center index u_i and the along-track speed of the AUV as discussed in more detail in Section 4.2.2. Figure 4.5 illustrates exemplarily the conversion of physical arrays into equivalent phase center arrays for a 2-D geometry, *i.e.*, $\Delta_{p+1}^{\mathrm{h}} = 0$. The RPC pairs are highlighted within the gray-shaded box. The characteristic of echo signals corresponding to an RPC pair, also called RPC signals, is their strong correlation behavior. This is explained by the fact that each RPC pair illuminates the stationary target scene with an identical overlapping virtual beam footprint at the same transmission and reception times. However, these RPC signals are not coherent given the different transmitter and receiver positions during consecutive pings. Typically, the phase center approximation assumes a stop-and-hop scenario, far-field conditions as well as a narrowbeam SAS system. However, SAS imaging usually takes place in the near-field, which violates this assumption. Moreover, state-of-the-art SAS systems are widebeam rather than narrowbeam [Callow et al. (2009)]. Therefore, occurring near-field and widebeam effects may have to be considered using the compensation

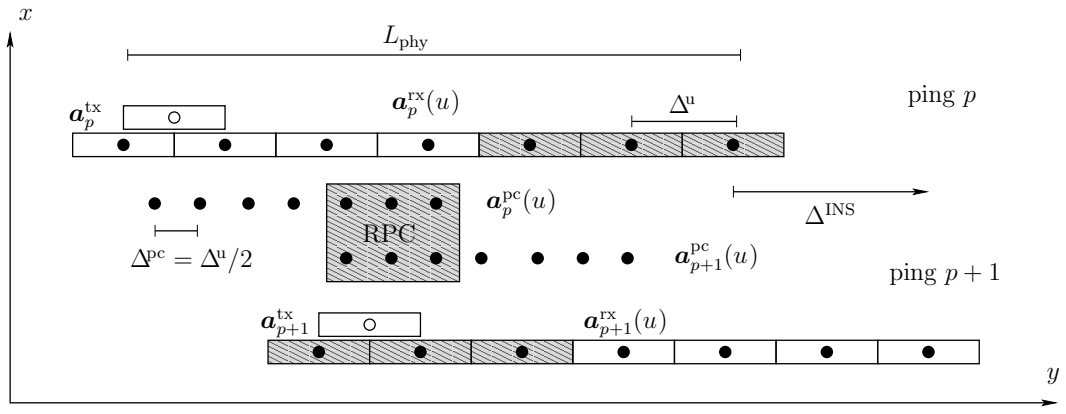


Figure 4.5: Conversion of single-transmitter and multi-receiver arrays into their corresponding phase center arrays for two consecutive pings. The redundant phase center (RPC) pairs are highlighted within the gray-shaded box. Figure is adapted from [Bellettini and Pinto (2002)].

technique as outlined in Section 4.2.5. The definition of RPC pairs in (4.11) requires a perfect along-track alignment of the phase center arrays. As motion without surge variation hardly occurs in reality, the next processing block of the DPCA algorithm is to identify the true RPC pairs based on the correlation structure in the echo signals.

4.2.2 Surge estimation

Surge describes a motion error that models the uncertainties in measuring the forward speed of the imaging platform. For a constant bias in the presumed advance per ping (APP), quadratic phase errors are induced along the synthetic aperture, possibly affecting the image reconstruction process [Jakowatz et al. (1996)]. Simultaneously, a surge leads to an incorrect selection of RPC pairs if solely relying on phase center positions $\boldsymbol{a}_p^{\mathrm{pc}}(u)$ and the APP provided by the INS unit. The latter is denoted by Δ^{INS} and is similar to Δ^{A} in the case of a transceiver system – see Section 2.1. It describes the traveling distance in along-track direction between two pings for a single receiving element, which is expressed as

$$\Delta^{\mathrm{INS}}(u) = \left(\boldsymbol{a}_{p+1}^{\mathrm{rx}}(u) - \boldsymbol{a}_p^{\mathrm{rx}}(u)\right)^T \boldsymbol{v}_{\mathrm{A}}, \text{ with } u = 1, \ldots, N_{\mathrm{rx}}. \tag{4.12}$$

Here, $\boldsymbol{v}_{\mathrm{A}}$ describes a unit vector pointing in the along-track direction. Since the array is mounted onto a rigid body, the APP is the same for each receiving element, *i.e.*, $\Delta^{\mathrm{INS}} \equiv \Delta^{\mathrm{INS}}(u)$. The RPC pair u_j of phase center u_i is then found via the relationship

$$u_j = \left\lfloor \frac{2\Delta^{\mathrm{INS}}}{\Delta^{\mathrm{u}}} \right\rceil + u_i, \text{ with } u_i = 1, \ldots, N_{\mathrm{pc}}, \tag{4.13}$$

where N_{pc} denotes the maximum number of RPC pairs and $\lfloor \cdot \rceil$ rounds towards the nearest integer. It is given by

$$N_{\mathrm{pc}} = \max\left(0, N_{\mathrm{rx}} - \lfloor 2\Delta^{\mathrm{INS}}/\Delta^{\mathrm{u}} \rceil\right). \tag{4.14}$$

However, given an unknown surge motion of $\Delta_{p+1}^{\mathrm{y}}$, the position $\tilde{\boldsymbol{a}}_{p+1}^{\mathrm{pc}}(\tilde{u}_j)$ of the true RPC index $\tilde{u}_j$ changes to the location

$$\tilde{\boldsymbol{a}}_{p+1}^{\mathrm{pc}}(\tilde{u}_j) = \boldsymbol{v}_{\mathrm{A}}\left(\Delta^{\mathrm{INS}} + \Delta_{p+1}^{\mathrm{y}}\right) + \boldsymbol{a}_p^{\mathrm{pc}}(u_i), \tag{4.15}$$

such that the RPC signals of the pair $u_{ij} = (u_i, u_j)$ decorrelate in contrast to the pairs $\tilde{u}_{ij} = (u_i, \tilde{u}_j)$ that become correlated. For the sake of clarity, sway and heave motion is neglected in (4.15). In order to estimate the surge component $\Delta_{p+1}^{\mathrm{y}}$, the correct

RPC pair index $\tilde{u}_j$ has to be found. This is done via the normalized cross-covariance function [Groen (2006), Oeschger (2006)] using neighboring RPC signals, where

$$\hat{c}^{p+1}_{u_i u'_j}(\kappa) = \frac{1}{M_n\, \hat{\sigma}_{e_p} \hat{\sigma}_{e_{p+1}}} \sum_{n=0}^{M_n-1-\kappa} \bar{e}_p(u_i, n+\kappa)\bar{e}_{p+1}(u'_j, n)^*, \quad 0 \le \kappa \le M_n - 1 \tag{4.16}$$

$$\text{with} \quad u_i = 1, \ldots, N_{\text{pc}} \quad \text{and} \quad u'_j = u_j - \frac{N^{\text{y}}_{\max} - 1}{2}, \ldots, u_j + \frac{N^{\text{y}}_{\max} - 1}{2}$$

describes an estimator of the normalized cross-covariance function for positive lags, *i.e.*, $\kappa \ge 0$. By exploiting the property $\hat{c}_{u_i u'_j}(\kappa) = \hat{c}_{u'_j u_i}(-\kappa)^*$ [Böhme (1998)] where $(\cdot)^*$ denotes complex conjugate, the cross-covariance function can be estimated for negative lags with $\kappa < 0$. In (4.16), $\hat{\sigma}_{e_p}$ is the sample standard deviation of the corresponding echo signal, $\bar{e}_p(u, n) = e_p(u, n) - \hat{\mu}_p$ describes the echo data with sample mean removed, and $N^{\text{y}}_{\max}$ denotes the maximum expected surge motion such that the following condition is fulfilled

$$u_j - \frac{N^{\text{y}}_{\max} - 1}{2} \le \tilde{u}_j \le u_j + \frac{N^{\text{y}}_{\max} - 1}{2}. \tag{4.17}$$

By introducing the maximum expected surge $N^{\text{y}}_{\max}$, the computational load can be significantly reduced by limiting the number of estimates of cross-covariance functions. Note that in (4.16), the cross-correlation is performed in the fast-time dimension for each possible combination of u_i and u'_j as long as $1 \le u'_j \le N_{\text{rx}}$. In a next step, the maximum peak value of each cross-covariance function $\hat{c}^{p+1}_{u_i u'_j}(\kappa)$ is determined as follows

$$\hat{\rho}^{p+1}_{u_i u'_j} = \max_{\kappa} \left| \hat{c}^{p+1}_{u_i u'_j}(\kappa) \right|. \tag{4.18}$$

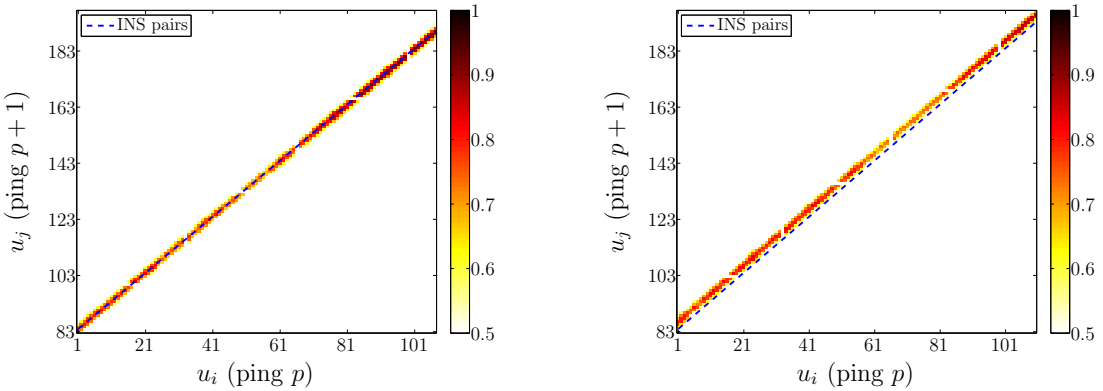
Thereupon, the surge is estimated as the difference between the presumed INS APP, Δ^{INS}, and the data-driven estimated APP, $\hat{\Delta}^{\text{y}}_{p+1}$, as follows

$$\hat{\Delta}^{\text{y}}_{p+1} = \frac{1}{N_{\text{pc}}} \sum_{u_i=1}^{N_{\text{pc}}} \hat{\Delta}^{\text{y}}_{p+1}(u_i) - \Delta^{\text{INS}}, \tag{4.19}$$

where the individual surge estimates $\hat{\Delta}^{\text{y}}_{p+1}(u_i)$ with $u_i = 1, \ldots, N_{\text{pc}}$ are given by

$$\hat{\Delta}^{\text{y}}_{p+1}(u_i) = \arg \max_{u'_j} \hat{\rho}^{p+1}_{u_i u'_j} - u_i. \tag{4.20}$$

It should be noted that the correlation in the fast-time dimension of (4.16) covers the entire received signal. However, in practice, the correlation operation is performed on short-time windows of the RPC signals in order to account for the varying time delays due to the 3-D geometry of the imaging scenario. This is discussed in more detail in Section 4.2.3 and Section 4.2.6. A surge estimation example using a real data set recorded by the MCM-SLS system is depicted in Figure 4.6, where all correlation



(a) Motion without surge variation

(b) Motion with surge variation

Figure 4.6: Surge estimation example between two pings taken from MCM-SLS data.

values below a threshold of $\rho_{\min} = 0.5$ have been ignored. While in Figure 4.6a the presumed APP by the INS (blue dashed line) is in agreement with the estimated APP $\hat{\Delta}^{\mathrm{y}}_{p+1}$, they differ in Figure 4.6b indicating a surge variation in the forward motion. Here, the magnitude of the shift difference between both diagonals determines the amount of surge shift, and the location of the estimated diagonal, *i.e.*, below or above the INS diagonal, identifies the surge direction. Neglecting the average operation over the number of RPC pairs in (4.19), in principle, the introduced correlation method is only able to provide surge estimates as integer multiples of the phase center interelement spacing $\Delta^{\mathrm{pc}} = \Delta^{\mathrm{u}}/2$. Although quadratic peak interpolation approaches exist to obtain fractional surge estimates [Cook (2007)], the additional accuracy is not relevant considering the small quadratic phase error it induces along the synthetic aperture.

Moreover, alternative surge estimation methods exist, *e.g.*, replacing the cross-correlation function in (4.16) by the inner product given a vector representation of the RPC signals or using an eigendecomposition approach [Cook (2007)]. While both methods are computationally more attractive, they require, however, that the time delay difference between RPC signals is less than a resolution cell. Consequently, this class of surge estimation techniques is limited to a maximum sway and heave motion. Therefore, these techniques are not considered in the proposed SAS processing chain. Although a larger number of RPC pairs provides a higher estimation accuracy, it reduces the APP of the imaging system at the same time. Hence, the coverage rate of the system decreases. This trade-off has to be considered in the design process of a SAS system. Alternatively, an adaptive scheme could be introduced, which lowers the forward speed in difficult motion compensation scenarios to produce a larger number of RPC pairs.

4.2.3 Time delay estimation

Correcting the surge influence and identifying the true RPC pairs with $u_{ij} \equiv \tilde{u}_{ij}$ guarantees that the RPC signals are highly correlated, which is a prerequisite for estimating their time delay difference $\Delta\tau(u_{ij})$. The latter is induced into the RPC signals due to a ping-to-ping sway and heave motion between the RPC arrays as follows

$$\Delta\tau(u_{ij}) = 2 \left\| \boldsymbol{a}_{p+1}^{\mathrm{pc}}(u_j) - \boldsymbol{a}_{p}^{\mathrm{pc}}(u_i) \right\|_2 / c. \tag{4.21}$$

In order to obtain estimates of the sway and heave displacement, a model function is fitted to the retrieved time delay differences in Section 4.2.6. Hence, time delay estimation (TDE) is a crucial step inside the motion compensation processing chain as its accuracy mainly dictates the estimation accuracy of the translational motion errors. Typically, the reconstruction of high-resolution and high-quality SAS images requires subwavelength accuracy [Jakowatz et al. (1996), Hayes and Gough (2009)] in the knowledge of transmitter and receiver positions along the synthetic aperture, which necessitates the use of subsample TDE techniques. The TDE technique, which is employed inside the proposed SAS processing chain, is addressed in the following.

The general approach to estimate a time delay difference or simply a time delay between two signals $\mathcal{X}_i(n) = e_p(u_i, n) + \mathcal{V}_i(n)$ and $\mathcal{Y}_j(n) = e_{p+1}(u_j, n) + \mathcal{V}_j(n)$ is based on their cross-covariance function [Carter (1987), Böhme (1998)]. Both signals $\mathcal{X}_i(n)$ and $\mathcal{Y}_j(n)$ are assumed to be wide-sense stationary (WSS) random processes with zero-mean, *i.e.*, $E\{\mathcal{X}_i(n)\} = E\{\mathcal{Y}_j(n)\} = 0$, where $E\{\cdot\}$ denotes expectation. Furthermore, the random noise processes $\mathcal{V}_i(n)$ and $\mathcal{V}_j(n)$ are assumed to be mutually uncorrelated $\forall\, i, j$ and uncorrelated with the transmitted signal $s(n)$. Expressing the received echo signals as delayed versions of $s(n)$, neglecting any attenuation factor, as follows $e_p(u_i) = s(n - \eta_i)e^{-j\omega_c\eta_i}$ and $e_{p+1}(u_j) = s(n - \eta_j)e^{-j\omega_c\eta_j}$, the cross-covariance function of the random processes $\mathcal{X}_i(n)$ and $\mathcal{Y}_j(n)$ is given by

$$c_{\mathcal{X}_i\mathcal{Y}_j}(\kappa) = E\big\{ \mathcal{X}_i(n+\kappa)\, \mathcal{Y}_j(n)^* \big\} = e^{-j\omega_c\Delta\eta_{ij}} c_{ss}(\kappa - \Delta\eta_{ij}). \tag{4.22}$$

Here, $\Delta\eta_{ij} = \eta_i - \eta_j$ is the delay difference and $c_{ss}(\kappa)$ is the auto-covariance function of the transmitted signal. Exploiting the property $c_{ss}(0) \geq c_{ss}(\kappa)$ for $\kappa > 0$ [Hayes (1996), Böhme (1998)] yields a coarse time delay as

$$\Delta\tau_c(u_{ij}) = T_s \, \lfloor \Delta\eta_{ij} \rceil = T_s \, \left\lfloor \arg\max_{\kappa} \left| c_{\mathcal{X}_i\mathcal{Y}_j}(\kappa) \right| \right\rceil . \tag{4.23}$$

Thus, the accuracy of $\Delta\tau_c(u_{ij})$ depends, among other things, mainly on the sampling interval T_s of the analog-to-digital converter (ADC). However, micronavigation requires subwavelength precision due to stringent requirements on the sensor position accuracy.

Hence, additional phase information is considered. In the sequel, a two-step estimation approach [Sæbø et al. (2007), Cook (2007)] is outlined that exploits the magnitude information given by (4.23) as well as the phase information of the cross-covariance function in (4.22). First, the estimator is constructed as follows

$$\Delta\tau_f(m, u_{ij}) = -\frac{\alpha(u_{ij})}{2\pi f_c} + \frac{m}{f_c}, \quad m \in \mathbb{Z}, \tag{4.24}$$

where f_c is the carrier frequency, $\alpha(u_{ij}) = \angle\left\{c_{\mathcal{X}_i\mathcal{Y}_j}(\lfloor\Delta\eta_{ij}\rceil)\right\}$ describes the phase at the sampled peak location of the cross-covariance function $c_{\mathcal{X}_i\mathcal{Y}_j}(\kappa)$, and m denotes a variable describing the number of full carrier phase cycles. Thus, the expression in (4.24) consists of two parts: the first determines a subsample delay and the second, a sample delay. Unfortunately, the number of full phase cycles m_f is typically unknown but can be obtained by solving the subsequent minimization problem

$$m_f(u_{ij}) = \arg\min_m \left|\Delta\tau_c(u_{ij}) - \Delta\tau_f(m, u_{ij})\right|, \quad m \in \mathbb{Z}, \tag{4.25}$$

by comparing the fine delay estimate as a function of full phase cycles with the coarse time delay estimate of (4.23). Thus, a subsample time delay is obtained as follows

$$\Delta\tau_f(u_{ij}) = \Delta\tau_f(m, u_{ij})\big|_{m=m_f(u_{ij})}. \tag{4.26}$$

Alternatively, a quadratic peak interpolation [Moddemeijer (1991)], adaptive filters [Youn et al. (1982), Leier and Zoubir (2012a)] or the approach by Cleveland and Parzen [Cleveland and Parzen (1975)] can be employed to obtain subsample precision. However, extensive testing for real sonar measurements has shown superior performance of the introduced two-step TDE approach with respect to precision and computational complexity. Therefore, it is used inside the motion compensation processing chain. The method is also applied for broadside bathymetry estimation as outlined in Section 3.5.

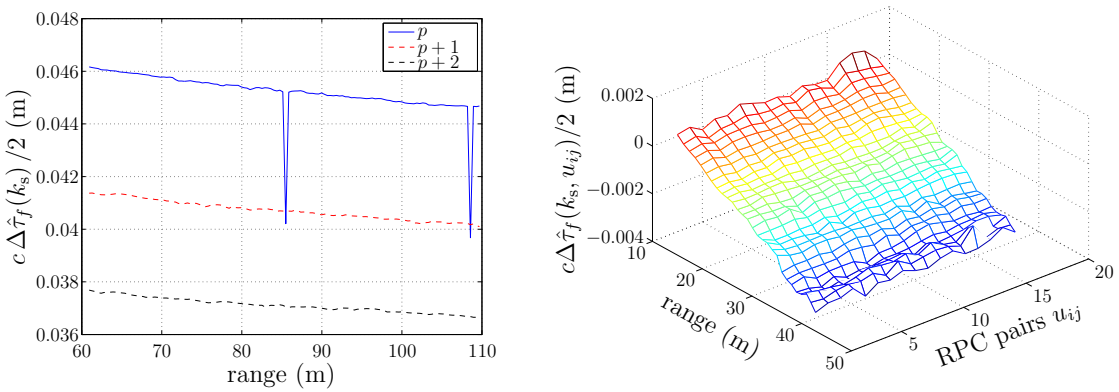
The above procedure assumes a stationary behavior [Böhme (1998)] of the time delay difference between RPC signals. However, given the common 3-D geometry of the imaging scenario – see Figure 4.3, the RPC pairs are typically displaced by ground-range sway $\Delta^{\mathrm{s}}_{p+1}$ and heave $\Delta^{\mathrm{h}}_{p+1}$ such that the resulting time delay difference varies in the fast-time dimension of the received signals [Bellettini and Pinto (2002)]. Therefore, TDE becomes a non-stationary problem. A common approach to cope with a slowly varying non-stationarity is to divide the signal into short-time windows, for which piece-wise or local stationarity can be assumed [Adak (1998)]. Dividing the echo signals of both pings into K short-time windows each of length M_k leads to

$$e_p\big(u_i, n; k_{\mathrm{s}}\big) = e_p\big(u_i, n + (k_{\mathrm{s}} - 1)\, M_k\big), \tag{4.27}$$

$$e_{p+1}\big(u_j, n; k_{\mathrm{s}}\big) = e_{p+1}\big(u_j, n + (k_{\mathrm{s}} - 1)\, M_k\big), \tag{4.28}$$

$$\text{for} \quad k_{\mathrm{s}} = 1, \ldots, K \quad \text{and} \quad n = 0, \ldots, M_k - 1.$$

The introduced TDE method is then performed for each short-time window with index k_s, which enables the time delay difference estimate of $\Delta\tau(k_\mathrm{s}, u_{ij})$ to be expressed as a function of fast-time. This procedure is repeated for all RPC pairs $u_{ij} = (u_i, u_j)$ with $u_i = 1, \ldots, N_\mathrm{pc}$ and $u_j = \left\lfloor (\Delta^\mathrm{INS} + \hat{\Delta}^\mathrm{y}_{p+1})/\Delta^\mathrm{pc} \right\rceil + u_i$, where $\hat{\Delta}^\mathrm{y}_{p+1}$ denotes the estimated ping-to-ping surge, and $\Delta^\mathrm{pc} = \Delta^\mathrm{u}/2$ is the interelement spacing between phase centers. Notice that the time delay estimates are only unbiased using the raw echo signals in the case of negligible near-field and widebeam effects. Otherwise, broadside beamforming has to be applied and the TDE method has to operate on the beamformed RPC signals. More details on how to beamform RPC signals are discussed in Section 4.2.5.



(a) Single RPC

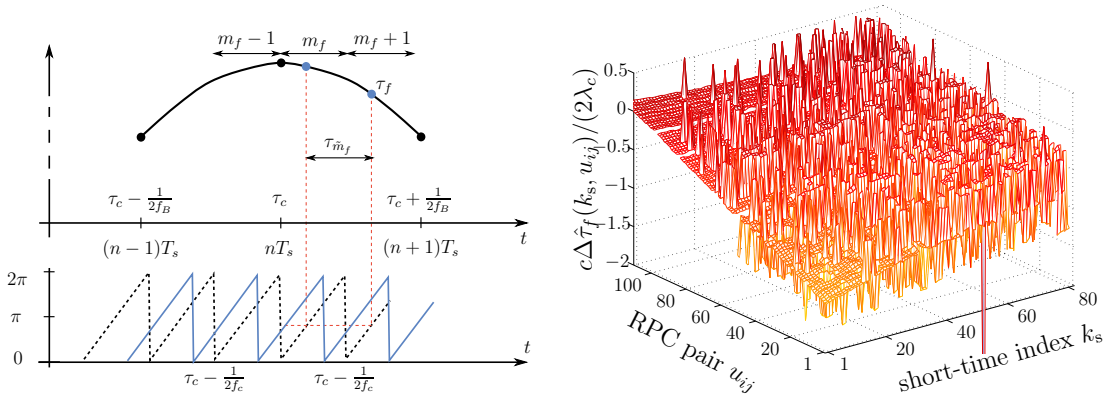
(b) All RPC pairs

Figure 4.7: Examples of time delay estimation using two different SAS systems. Single TDE curves are shown for three consecutive pings in (a), and a single ping TDE curve for all RPC pairs u_{ij} is depicted in (b).

A real sonar data example of range-variant (fast-time variant) time delay estimates $\Delta\hat{\tau}_f(k_\mathrm{s})$ scaled by a factor $c/2$ is depicted in Figure 4.7a for three consecutive pings. A short-time window length of $L_k = c\,M_k\,T_s = 1.8$ m has been used with an overlapping factor of 50% between windows. In Figure 4.7b, the estimated time delays $\Delta\hat{\tau}_f(k_\mathrm{s}, u_{ij})$ are depicted as a function of range and RPC pair index u_{ij}. The sonar measurements originate from two sonars with different frequencies, which also cover a different minimum and maximum ground-range swath. While the time delays for ping $p+1$ and $p+2$ in Figure 4.7a show a smooth behavior, the estimates for ping p are corrupted by two outliers due to an erroneous selection of the full phase cycle number m_f that is denoted by $\tilde{m}_f$. These outliers arise as a consequence of a high ratio of carrier frequency to bandwidth, and lead to an error in the time delay of $\tau_{\tilde{m}_f} = c/(2f_c) \approx 0.005$ m, which may affect the sway and heave estimation. In order to cope with these biased estimates, a phase wrapping compensation technique is proposed in the next section.

4.2.4 Phase wrap error correction

In general, the two-step TDE approach as outlined in Section 4.2.3 estimates accurate time delays, which are the basis for a precise correction of sway and heave motion errors. However, given a high ratio of carrier frequency to bandwidth, phase wrap errors are likely to occur that may lead to biased position estimates. Typically, ADC takes place after the demodulation of continuous-time signals by sampling the baseband signal with $f_s \geq f_B$, rather than sampling at a rate $f_s \geq 2f_{\max}$ with $f_{\max} = f_c + f_B/2$, where f_B denotes the bandwidth. Although baseband sampling relaxes hardware constraints of the ADC, phase wraps in the carrier phase term influence the identification of the full phase cycle number m_f in (4.25). Here, the identification mainly relies on the fact that the carrier phase term cycles less than once between two sample points. Otherwise, a wrong selection of m_f leads to a biased time delay estimate, especially for larger Q-factor systems with $Q = f_c/f_B$. An example where the carrier phase term cycles twice during a sampling interval $T_s = 0.5/f_B$ is depicted in Figure 4.8a. Thus, given a true time delay τ_f and picking the phase at the peak location τ_c yields a wrong number of full phase cycles $\tilde{m}_f$, which induces a biased time delay estimate (here, $\tau_{\tilde{m}_f} = -1/f_c$). A real data example of time delay estimates $\Delta\hat{\tau}_f(k_\mathrm{s}, u_{ij})$ scaled by $c/(2\lambda_c)$ is depicted in Figure 4.8b as a function of short-time index k_s and RPC pair index u_{ij}. It highlights the severe impact of an incorrect selection of phase cycles on the two-step TDE approach. The data has been recorded using the MCM-SLS SAS system, which has a Q-factor of $Q \approx 4.3$. In the sequel, a technique is proposed to compensate for these phase wrap errors. The individual steps of the proposed algorithm [Leier and Zoubir (2012)] are as follows:



(a) Phase wrap explanation

(b) Biased time delay estimates

Figure 4.8: Illustration of an explanation why phase wrap errors occur (a) in case of a high ratio of carrier frequency to bandwidth (Q-factor). A TDE example with phase wrap errors is shown in (b) for real sonar measurements.

1. Given time delay estimates $\Delta\hat{\tau}_f(k_\mathrm{s}, u_{ij})$ for each short-time window k_s with $k_\mathrm{s} = 1, \ldots, K$ and RPC pair $u_i \equiv u_{ij}$, with $u_i = 1, \ldots, N_\mathrm{pc}$, determine the subsample time delay as

$$\Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_i) = \Delta\hat{\tau}_f(k_\mathrm{s}, u_i) - \tilde{m}_f(k_\mathrm{s}, u_i)/f_c,$$

where $\tilde{m}_f(k_\mathrm{s}, u_i)$ is the erroneous ambiguous number due to a high ratio of carrier frequency to bandwidth.

2. Take the first difference of $\Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_i)$ in both dimensions such that

$$\begin{aligned}\Delta\hat{\tau}'_{\alpha,1}(k_\mathrm{s}, u_i) &= \Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_i) - \Delta\hat{\tau}_\alpha(k_\mathrm{s}+1, u_i) \quad \text{and}\\ \Delta\hat{\tau}'_{\alpha,2}(k_\mathrm{s}, u_i) &= \Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_i) - \Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_i+1).\end{aligned}$$

3. Obtain a binary image $\boldsymbol{I}_B = \boldsymbol{I}_{B_1} \vee \boldsymbol{I}_{B_2}$ via thresholding with $j = 1, 2$ as follows

$$\boldsymbol{I}_{B_j}(k_\mathrm{s}, u_i) = \begin{cases} 1 & \text{if } \Delta\hat{\tau}'_{\alpha,j}(k_\mathrm{s}, u_i) \geq \pi \\ 0 & \text{otherwise,} \end{cases}$$

where $\vee$ denotes an element-wise logical or-operation.

4. Perform a morphological closing operation with structuring element S_E on binary image $\boldsymbol{I}_B$ to fuse narrow breaks and eliminate small holes with

$$\boldsymbol{I}_C = (\boldsymbol{I}_B \oplus S_\mathrm{E}) \ominus S_\mathrm{E},$$

where $\oplus$ and $\ominus$ denote dilation and erosion, respectively.

5. Identify phase cycle regions $\boldsymbol{I}_l^\mathrm{PCR}$ with $l = 1, \ldots, N_\mathrm{pcr}$ in binary image $\boldsymbol{I}_C$, which are delimited by the detected phase wraps, *e.g.*, by determining the label connected components [Haralick and Shapiro (1992)].

6. Assuming that the majority of phase cycle numbers are correct, estimate the new phase cycle numbers as

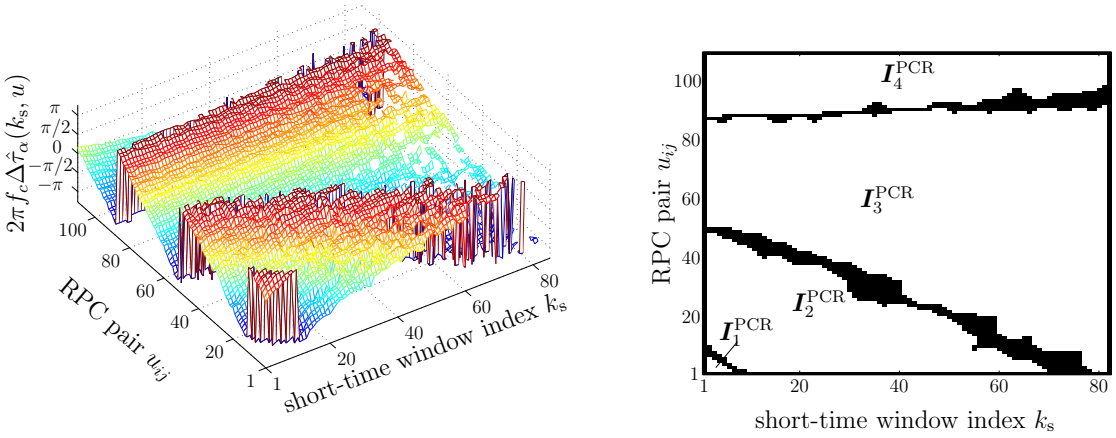
$$\hat{m}_c(k_\mathrm{s}, u_i) = \underset{(k_\mathrm{s}, u_i) \in \boldsymbol{I}_l^\mathrm{PCR}}{\mathcal{M}\big\{\tilde{m}_f(k_\mathrm{s}, u_i)\big\}},\ l = 1, \ldots, N_\mathrm{pcr}$$

where $\mathcal{M}\{\cdot\}$ describes the mode operation, which selects the most frequent value in a set. Obtain the time delay estimates with corrected phase wrap errors by applying

$$\Delta\hat{\tau}(k_\mathrm{s}, u_i) = \Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_i) + \hat{m}_c(k_\mathrm{s}, u_i)/f_c.$$

The principle of the proposed technique is to identify phase cycle regions using binary image processing methods [Gonzalez and Woods (2002)] such as morphological operations and label connected components [Haralick and Shapiro (1992)]. For each phase cycle region, it then determines the most frequent occurrence of full phase cycles to correct the individual time delays. The proposed technique relies on the assumption that the majority of phase cycle numbers are estimated correctly within a region. However, this assumption has been validated for a wide-variety of different data sets.

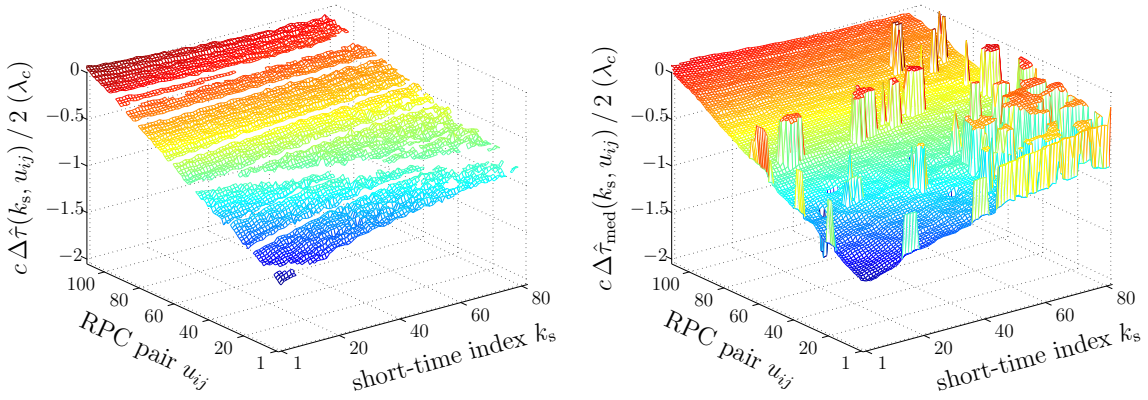
The input of the proposed algorithm is given by the biased estimates $\Delta\hat{\tau}_f(k_\mathrm{s}, u_{ij})$ as depicted in Figure 4.8b. After subtracting the full cycle number in Step 1, the subsample time delay $\Delta\hat{\tau}_\alpha(k_\mathrm{s}, u_{ij})$ is obtained as illustrated in Figure 4.9a. It is expressed in terms of the phase angle. The binary image of detected phase cycle regions $\boldsymbol{I}_l^{\mathrm{PCR}}$ is shown in Figure 4.9b, which is obtained after Step 5 of the proposed algorithm. Here, the black lines indicate the location of phase jumps as depicted in Figure 4.9a. The compensated time delays $\Delta\hat{\tau}(k_\mathrm{s}, u_{ij})$ are then depicted in Figure 4.10a, which demonstrate an enormous improvement over the initial estimates of Figure 4.8b. In order to compare exemplarily the bias correction capabilities of the proposed algorithm, a 2-D median filter is applied with a mask size of $M'_x \times M'_y$, where $M'_x = M'_y = 5$. The corresponding estimates $\Delta\hat{\tau}_{\mathrm{med}}(k_\mathrm{s}, u_{ij})$ are depicted in Figure 4.10b. While biased estimates are still observable for the median filter approach, the proposed algorithm is able to correct all biased estimates at the cost of introducing some missing data values. The latter arise due to guard bands around detected phase jumps. However, missing data points can more easily be handled during sway and heave estimation than biased time delays.



(a) Carrier phase

(b) Phase cycle regions

Figure 4.9: Illustration of carrier phase (a) and binary image $\boldsymbol{I}_C$ in (b) showing phase jumps (black lines) as well as phase cycle regions $\boldsymbol{I}_l^{\mathrm{PCR}}$ for $l = 1, \ldots, N_{\mathrm{pcr}}$ with $N_{\mathrm{pcr}} = 4$.



(a) Proposed algorithm

(b) 2D median filter

Figure 4.10: A comparison of corrected range difference estimates for the proposed phase wrap error compensation algorithm (a) and a 2D median filter (b).

Besides the provided example in Figure 4.10, performance of the proposed algorithm is evaluated using sonar measurements of $M_p = 500$ consecutive pings that have been recorded during a sea trial by the MCM-SLS SAS system. For the two-step TDE approach, a fixed short-range window (short-time window) of length $L_k = 0.6$ m is used with an overlap of 50% between neighboring windows. Moreover, a fast-time frame is considered according to a minimum and maximum slant-range of $R_{\min} = 20$ m and $R_{\max} = 50$ m, respectively. Additionally, only estimates are considered where the peak value of the normalized cross-covariance function is larger than $\rho_{\min} = 0.5$. Note that this is equivalent to a signal-to-noise ratio (SNR) of 0 dB given the relationship between coherence and SNR [Böhme (1998), Sæbø (2010)]. Thus, the subsequent evaluation of the proposed approach is performed in a low SNR region. First, the time delay estimates are cleaned as exemplarily depicted in Figure 4.10a using the proposed compensation technique. Then, the displacement technique for estimating the sway and heave [Cook et al. (2006)] as outlined in Section 4.2.6 is applied to reconstruct reference time delays using a flat bottom assumption. In order to limit the influence of biased time delays, which appear as outliers in the displacement estimation of sway and heave parameters, a robust linear regression is carried out with a bisquare weighting function [Cleveland (1979)]. Subsequently, reference delays are calculated based on the estimated parameters to construct residuals of the proposed method and the median filter approach to determine the number of successfully corrected biased estimates. For this purpose, a threshold of $\tau_{\text{thres}} = 1/(2f_c)$ is set. On average, a contamination of 8.95% biased estimates have occurred in the used sonar measurements due to a high ratio of carrier frequency to bandwidth. While the median filter approach achieves a successful correction rate of 83.91%, the proposed method shows superior performance by correcting 98.11%. Note that the contamination rate as well as both correction

rates depend on the SNR, and therefore, on the preset minimum correlation value $\rho_{\min}$. Given that the proposed technique performs very well for a SNR of 0 dB, it becomes a default improvement block inside the SAS processing chain.

4.2.5 Near-field and widebeam correction

Besides exploiting the temporal and spatial coherence properties of the seafloor backscattering such that the corresponding RPC signals are correlated [Bellettini and Pinto (2002)], the principle of DPCA motion compensation relies on the relation between estimated time delays and translational displacements of the RPC arrays. The latter is constructed using the phase center approximation – see Section 4.2.1. It assumes (i) the target scene to be located in the far-field and (ii) a small along-track extent of the target scene compared to the imaging range. This is known as the narrowbeam assumption [Soumekh (1999), Callow et al. (2009)]. However, both conditions are hardly fulfilled for common state-of-the-art SAS systems, such as the MCM-SLS and VISION SAS. Typically, these systems cover a maximum slant-range of $R_{\max} \leq 200$ m. Inserting the system parameters into the Fraunhofer distance [Richards (2005)], which provides the boundary between near-field and far-field, as follows

$$R_{\text{far}} = \frac{2(L_{\text{phy}})^2}{\lambda}, \tag{4.29}$$

yields $R_{\text{far},1} \approx 1000$ m and $R_{\text{far},2} \approx 290$ m for the MCM-SLS and the VISION SAS systems, respectively. As these values clearly exceed the maximum slant-range $R_{\max}$, imaging and motion compensation takes place in a near-field scenario. However, in [Bellettini and Pinto (2002)], a narrowbeam condition is provided that has to be fulfilled for each RPC pair in near-field scenarios such that the DPCA technique is still applicable. The condition is given by

$$\frac{(u_i \Delta^{\text{u}})^2}{4 r \lambda_c}\big(1 - \cos^2(\theta_{\text{BW}}/2)\big) \ll 1, \text{ for } u_i = 1, \ldots, N_{\text{pc}}, \tag{4.30}$$

where $\theta_{\text{BW}}/2$ denotes half of the angular transmission beamwidth and $u_i \Delta^{\text{u}}$ describes the along-track position of the phase center array. The expression is derived from a Taylor series approximation of the range error between true and virtual round-trip distances, which is denoted by

$$\Delta r^{\text{PCA}}(u_i) = \left\| \boldsymbol{a}_p^{\text{tx}} - \boldsymbol{q}_d \right\|_2 + \left\| \boldsymbol{a}_p^{\text{rx}}(u_i) - \boldsymbol{q}_d \right\|_2 - 2 \left\| \boldsymbol{a}_p^{\text{pc}}(u_i) - \boldsymbol{q}_d \right\|_2 . \tag{4.31}$$

According to [Bellettini and Pinto (2002)], the condition in (4.30) ensures that the excess in round-trip delay is identical for all targets inside the angular transmission

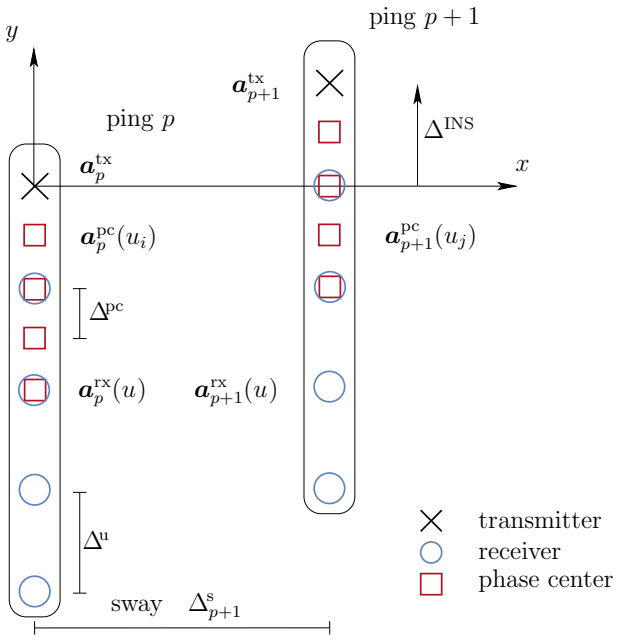


Figure 4.11: Displacement illustration of receiver arrays and their corresponding RPC arrays for two pings in a 2-D scenario. The sway displacement is denoted by $\Delta^{\mathrm{s}}_{p+1}$.

beamwidth for a given range. Although it is fulfilled for both SAS systems (*e.g.* $\sim 10^{-4}$ for $u_i = 1$ and $2.1 \cdot 10^{-3}$ for $u_i = N_{\mathrm{pc}}$ in the case of VISION SAS), the time delay difference varies between echo signals over the RPC array, thereby, potentially introducing a bias into the displacement estimation of motion errors.

In order to illustrate this, a 2-D geometry is considered in Figure 4.11. It shows a single-transmitter and multi-receiver array configuration for two consecutive pings as well as their RPC pairs $u_{ij} = (u_i, u_j)$. The two array positions are displaced by

$$\Delta^{\mathrm{s}}_{p+1} = \left\| \boldsymbol{a}^{\mathrm{pc}}_{p+1}(u_j) - \boldsymbol{a}^{\mathrm{pc}}_{p}(u_i) \right\|_2, \ \forall\, u_i,\, u_j. \tag{4.32}$$

Assuming a single target at $\boldsymbol{q}_d = [x_d, y_d]^T$, the range difference of the corresponding RPC pairs u_{ij} for a single-transmitter and multi-receiver SAS system reads

$$\begin{aligned} \Delta r^{\mathrm{rx}}_{d,p}(u_{ij}) = \big(\left\| \boldsymbol{a}^{\mathrm{tx}}_{p} - \boldsymbol{q}_d \right\|_2 + \left\| \boldsymbol{a}^{\mathrm{rx}}_{p}(u_i) - \boldsymbol{q}_d \right\|_2 \big) - \\ \big(\left\| \boldsymbol{a}^{\mathrm{tx}}_{p+1} - \boldsymbol{q}_d \right\|_2 + \left\| \boldsymbol{a}^{\mathrm{rx}}_{p+1}(u_j) - \boldsymbol{q}_d \right\|_2 \big), \end{aligned} \tag{4.33}$$

and its counterpart for RPC arrays is given by

$$\Delta r^{\mathrm{pc}}_{d,p}(u_{ij}) = 2 \left(\left\| \boldsymbol{a}^{\mathrm{pc}}_{p}(u_i) - \boldsymbol{q}_d \right\|_2 - \left\| \boldsymbol{a}^{\mathrm{pc}}_{p+1}(u_j) - \boldsymbol{q}_d \right\|_2 \right). \tag{4.34}$$

Using a first-order Maclaurin series approximation for $(1+x)^b \approx 1 + bx$, the range difference $\Delta r^{\mathrm{pc}}_{d,p}(u_{ij})$ of RPC pairs u_{ij} can be expressed as (see Appendix A.1)

$$\Delta r^{\mathrm{pc}}_{d,p}(u_{ij}) \approx 2\Delta^{\mathrm{s}}_{p+1} + \frac{\left(y_d + u_i \Delta^{\mathrm{pc}}\right)^2 - \left(y_d - \Delta^{\mathrm{INS}} + u_j \Delta^{\mathrm{pc}}\right)^2}{x_d}. \tag{4.35}$$

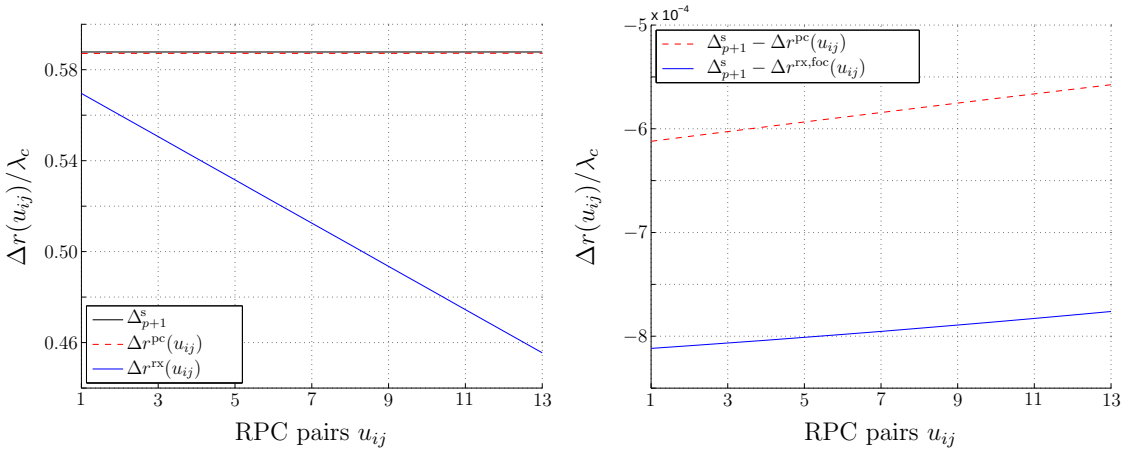
Substituting the index relation $u_j = u_i + \Delta^{\mathrm{INS}}/\Delta^{\mathrm{pc}}$ between RPC pairs into the expression of (4.35) associates the range difference with the sway displacement as

$$\Delta r^{\mathrm{pc}}_{d,p}(u_{ij}) \approx 2\Delta^{\mathrm{s}}_{p+1}. \tag{4.36}$$

Note that the expression is independent of the RPC pair index u_{ij} and that Δ^{INS} is assumed to be an integer multiple of the interelement spacing Δ^{pc} of an RPC array. Hence, theoretically, the conversion of a time delay into a range difference provides a direct relation to the sway displacement $\Delta^{\mathrm{s}}_{p+1}$ between two RPC arrays. However, performing the same derivation for the range difference of a single-transmitter and multi-receiver configuration yields (see Appendix A.2)

$$\Delta r^{\mathrm{rx}}_{d,p}(u_{ij}) \equiv \Delta r^{\mathrm{rx}}_{d,p}(u_i) \approx 2\Delta^{\mathrm{s}}_{p+1} - \frac{\Delta^{\mathrm{INS}}(\Delta^{\mathrm{INS}} + u_i\,\Delta^{\mathrm{u}})}{x_d}, \quad u_i = 1, \ldots, N_{\mathrm{pc}}, \tag{4.37}$$

which is still a function of the phase center index u_i. Only in far-field situations, where x_d is large, does the second term vanish, and the range difference equals the sway displacement. Otherwise, it varies over the RPC array, which may cause biased estimates of the sway component $\Delta^{\mathrm{s}}_{p+1}$ although the condition in (4.30) is met. This behavior is highlighted in Figure 4.12a using a synthetic data example with VISION parameters. It shows the differences of round-trip distances normalized by the carrier wavelength λ_c for (i) the true displacement $\Delta^{\mathrm{s}}_{p+1}$ as stated in (4.32), (ii) the virtual RPC pairs, $\Delta r^{\mathrm{pc}}(u_{ij})$, and (iii) the single-transmitter and multi-receiver system, $\Delta r^{\mathrm{rx}}(u_{ij})$. While the RPC range difference $\Delta r^{\mathrm{pc}}(u_{ij})$ approximates the displacement $\Delta^{\mathrm{s}}_{p+1}$, the range difference $\Delta r^{\mathrm{rx}}(u_{ij})$ of the single-transmitter and multi-receiver configuration



(a) Differences of round-trip distances between two consecutive pings without compensation.

(b) Differences of round-trip distances between two pings after beamforming compensation with respect to the spatial displacement.

Figure 4.12: Time delay comparison of virtual RPC pairs and a single-transmitter and multi-receiver system before (a) and after beamforming correction (b).

excites a linear trend as a function of the RPC index u_i. Its maximum deviation from the true displacement $\Delta^{\mathrm{s}}_{p+1}$ is given approximately by $0.12\lambda_c$, which exceeds the required accuracy limit along the synthetic aperture to obtain focused SAS images [Jakowatz et al. (1996)]. The curves in Figure 4.12a depict average round-trip distances for point targets placed inside an area with an along-track extent of $Y_0 = 1$ m and a ground-range swath of $X_{\min} = 30$ m and $X_{\max} = 70$ m.

In the sequel, a technique is proposed to compensate for the varying time delays in order to estimate unbiased motion errors. Simultaneously, it increases the SNR, which leads to higher correlation values of the TDE method as depicted in Figure 4.13 and therefore, to more reliable time delay estimates. To this end, the echo signals are beamformed in broadside direction θ_0 to lower the influence of backscattering targets outside the array normal direction [Callow et al. (2009)]. The beam signals $e^{\mathrm{B}}_p(\theta, n)$, given as a function of azimuth direction θ, are formed by calculating an angle- and range-dependent time delay similar to the focusing delay of the backprojection algorithm – see Section 3.3. The beamforming delay $\Delta t^{\mathrm{B}}(\theta, r; \boldsymbol{a}^{\mathrm{rx}}_p(u), \boldsymbol{a}^{\mathrm{tx}}_p, \boldsymbol{a}^{\mathrm{ref}}_p)$ compensates the varying delays in the echo signals for each receiver element. It is given by

$$\Delta t^{\mathrm{B}}(\theta, r; \boldsymbol{a}^{\mathrm{rx}}_p(u_i), \boldsymbol{a}^{\mathrm{tx}}_p, \boldsymbol{a}^{\mathrm{ref}}_p) = \frac{r + r^{\mathrm{rx}}(\theta, r; u_i) - 2\, r^{\mathrm{ref}}(\theta, r)}{c},\ u_i = 1, \ldots, N_{\mathrm{pc}}, \tag{4.38}$$

where the slant-range distances are $r^{\mathrm{rx}}(\theta, r; u_i) = \left\| \boldsymbol{a}^{\mathrm{tx}}_p + \boldsymbol{g}^{\mathrm{fp}}(\theta, r) - \boldsymbol{a}^{\mathrm{rx}}_p(u_i) \right\|_2$ and $r^{\mathrm{ref}}(\theta, r) = \left\| \boldsymbol{a}^{\mathrm{tx}}_p + \boldsymbol{g}^{\mathrm{fp}}(\theta, r) - \boldsymbol{a}^{\mathrm{ref}}_p \right\|_2$. Here, $\boldsymbol{g}^{\mathrm{fp}}(\theta, r) = [r\cos(\theta), r\sin(\theta), 0]^T$ describes a focus vector in the slant-range domain, and $\boldsymbol{a}^{\mathrm{ref}}_p$ denotes the position of the beamforming reference point. The latter is chosen as $\boldsymbol{a}^{\mathrm{ref}}_p = \boldsymbol{a}^{\mathrm{pc}}_p(u_i)\big|_{u_i=1}$. Then, the broadside RPC beam signals $e^{\mathrm{B}}_p(\theta_0, n)$ can be written as follows

$$\begin{aligned} e^{\mathrm{B}}_p(\theta_0, n) = \sum_{u_i=1}^{N_{\mathrm{pc}}} & e_p\big(u_i, n + \Delta t^{\mathrm{B}}(\theta_0, r; \boldsymbol{a}^{\mathrm{rx}}_p(u_i), \boldsymbol{a}^{\mathrm{tx}}_p, \boldsymbol{a}^{\mathrm{ref}}_p)/T_s\big) \\ & \times \exp\big(j\Omega_c \Delta t^{\mathrm{B}}(\theta_0, r; \boldsymbol{a}^{\mathrm{rx}}_p(u_i), \boldsymbol{a}^{\mathrm{tx}}_p, \boldsymbol{a}^{\mathrm{ref}}_p)\big). \end{aligned} \tag{4.39}$$

Given discrete-time echo signals $e_p(u, n)$, an interpolation beamformer is required [Johnson and Dudgeon (1993)] to facilitate coherent processing by taking into account the correct signal values at the beamforming delays. Typically, a linear interpolation of the echo signals shows a good trade-off between accuracy and computational load. Note that (4.39) describes a stop-and-hop beamformer, which is adequate for estimating the sway and heave displacement for common swath widths. Alternatively, intermediate position samples provided by the INS unit can be used to interpolate the positions over the entire reception time for a single ping. After beamforming the RPC signals of both pings, time delay estimation is performed on $e^{\mathrm{B}}_p(\theta_0, n)$ and $e^{\mathrm{B}}_{p+1}(\theta_0, n)$ using short-time windows. This leads to a compensation of the varying time delay differences,

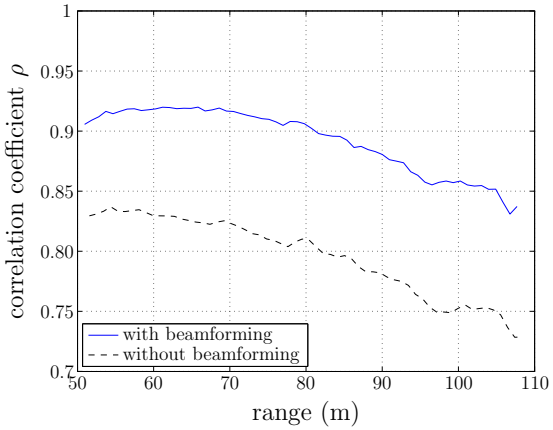


Figure 4.13: Comparison of average correlation coefficients of $M_p = 90$ pings for TDE based on beam signals (blue solid line) and on echo signals (black dashed line).

which then approximate $2\Delta^{\mathrm{s}}_{p+1}$, as illustrated in Figure 4.12b. The figure compares the error between the true displacement $\Delta^{\mathrm{s}}_{p+1}$ and the range difference of the virtual transceivers $\Delta r^{\mathrm{pc}}(u_{ij})$ with the error between the true displacement and the range difference $\Delta r^{\mathrm{rx,foc}}(u_{ij})$ of the beamformed RPC signals. As can be seen in Figure 4.12b, the difference between both error functions is on the order of $2 \cdot 10^{-4}\lambda_c$, which is significantly smaller than the accuracy requirement for SAS imaging. Note that only a single range bin has been focused in this example.

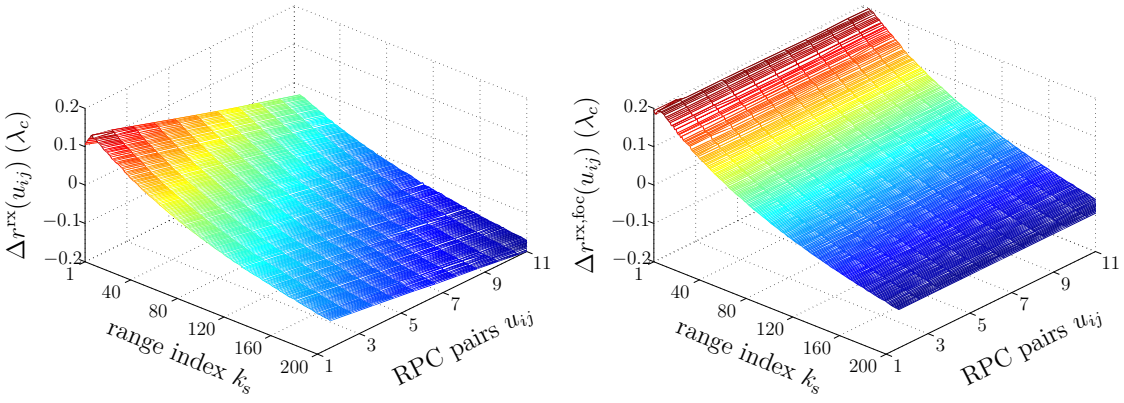
In general, there are two options how to beamform the RPC signals. Either the entire RPC array is used leading to a single time delay $\Delta\tau(k_{\mathrm{s}})$ as a function of the short-time index k_{s}; or a sliding window approach is employed. The latter yields time delays $\Delta\tau(k_{\mathrm{s}}, u_{ij})$ that additionally depend on the RPC pair index u_{ij}. This approach enables the use of the phase wrapping error compensation technique as outlined in Section 4.2.4. Given a maximum number of RPC pairs N'_{pc} in a sliding window frame with $N'_{\mathrm{pc}} \leq N_{\mathrm{pc}}$, the beamformed RPC signals can be expressed as

$$e^{\mathrm{B}}_p(\theta_0, n; u'_i) = \sum_{u_i = u'_i - \frac{N'_{\mathrm{pc}}-1}{2}}^{u'_i + \frac{N'_{\mathrm{pc}}-1}{2}} e_p\big(u_i, n + \Delta t^{\mathrm{B}}(\theta_0, r; \boldsymbol{a}^{\mathrm{rx}}(u_i), \boldsymbol{a}^{\mathrm{tx}}_p, \boldsymbol{a}^{\mathrm{ref}}_p)/T_s\big) \times \exp\big(j\Omega_c \Delta t^{\mathrm{B}}(\theta_0, r; \boldsymbol{a}^{\mathrm{rx}}(u_i), \boldsymbol{a}^{\mathrm{tx}}_p, \boldsymbol{a}^{\mathrm{ref}}_p)\big) \tag{4.40}$$

for $u'_i = (N'_{\mathrm{pc}} + 1)/2, \ldots, N_{\mathrm{pc}} - (N'_{\mathrm{pc}} - 1)/2$. Here, u'_i is the center index of the sliding window over the RPC dimension, and N'_{pc} is assumed to be an odd number.

A synthetic data example for this approach is illustrated in Figure 4.14, which highlights the impact of the focus compensation. The synthetic data has been generated using a 3-D imaging geometry, where only sway motion errors are injected. While time

delay estimation has been performed directly on the echo signals in Figure 4.14a, the proposed beamforming compensation has been applied prior to estimating the delay difference in Figure 4.14b using $N'_{\mathrm{pc}} = 3$. This leads to the same time delay difference for each RPC pair u_{ij} in contrast to Figure 4.14a. However, irrespective of this variation, the displacement estimation technique as addressed in Section 4.2.6 is capable of correctly estimating the sway motion due to model fitting in the range dimension – see also Figure 4.19a. By contrast, the resulting estimates erroneously suggest the presence of pitch and heave motion errors due to the varying delay in the RPC dimension that typically causes biased heave estimates – see also Figure 4.19b. The latter are usually not crucial for SAS image quality taking into account a common side-looking SAS geometry, but may affect the accuracy of bathymetry estimation.



(a) Uncompensated delay differences

(b) Compensated delay differences

Figure 4.14: Illustration of varying time delays of the RPC array dimension before (a) and after (b) applying the near-field and widebeam beamforming compensation.

4.2.6 Sway and heave estimation

The objective of the last block inside the motion compensation processing chain is to estimate the sway and heave displacement for each RPC pair based on time delays – see Figure 4.4. In general, this is described by a displacement estimator $\mathcal{D}\{\cdot\}$ as follows

$$\left[\hat{\Delta}^{\mathrm{s}}_{p+1}(u_{ij}), \hat{\Delta}^{\mathrm{h}}_{p+1}(u_{ij})\right] = \mathcal{D}\big\{c\,\Delta\hat{\tau}(k_{\mathrm{s}}, u_{ij})/2\big\}, \tag{4.41}$$

where both translational displacements as well as the time delay estimates are a function of the RPC pair index $u_{ij} = (u_i, u_j)$. As the phase center index u_j can always be determined via the relation $u_j = u_i + \Delta^{\mathrm{INS}}/\Delta^{\mathrm{pc}}$, the RPC pair index u_{ij} is

rewritten as $u_{ij} \equiv u_i$ for notational convenience. In the absence of rotational motion errors, the underlying sway and heave motion process can be assumed to be stationary. Hence, the ping-to-ping sway estimate is obtained by

$$\hat{\Delta}^{\mathrm{s}}_{p+1} = \frac{1}{N_{\mathrm{pc}}} \sum_{u_i=1}^{N_{\mathrm{pc}}} \hat{\Delta}^{\mathrm{s}}_{p+1}(u_i). \tag{4.42}$$

The ping-to-ping heave is estimated in a similar way. In the presence of rotational motion errors, a line fitting method has to be applied instead of simple averaging, where the slope is related to the rotational error, *i.e.*, yaw or pitch [Cook (2007)]. However, the developed data-driven motion compensation technique is currently only employed to handle translational motion errors since the INS is considered to be accurate enough to measure rotation.

In the sequel, the estimator $\mathcal{D}\{\cdot\}$ is addressed in more detail for a single RPC pair using the range difference values $\Delta\hat{r}(k_{\mathrm{s}}, u_i) = c\,\Delta\hat{\tau}(k_{\mathrm{s}}, u_i)/2$ of $k_{\mathrm{s}} = 1, \ldots, K$ short-time estimates of the time delay difference $\Delta\hat{\tau}(k_{\mathrm{s}}, u_i)$. In order to obtain motion estimates, a model fitting approach based on least-squares estimation is applied [Kay (1993)], which solves the minimization problem

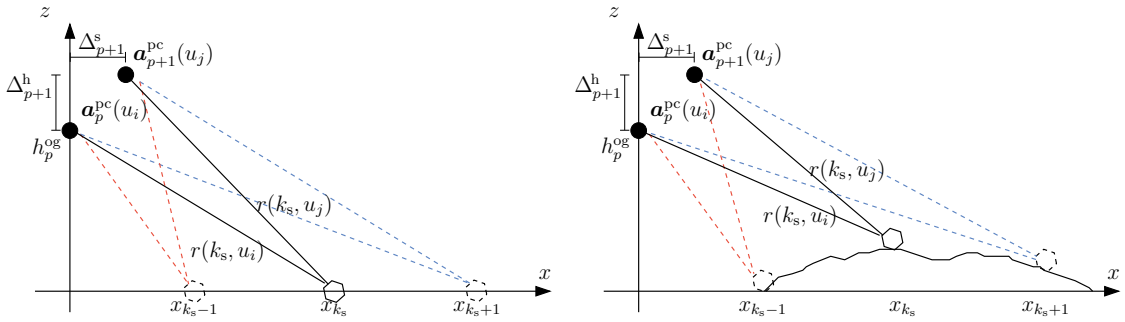
$$\hat{\boldsymbol{\Delta}}_{p+1}(u_i) = \arg \min_{\boldsymbol{\Delta}_{p+1}} \left| \Delta r(\boldsymbol{g}_{k_{\mathrm{s}}}; \boldsymbol{a}^{\mathrm{pc}}_p(u_i), \boldsymbol{\Delta}_{p+1}) - \Delta\hat{r}(k_{\mathrm{s}}, u_i) \right|^2, \tag{4.43}$$

given a model function $\Delta r(\boldsymbol{g}_{k_{\mathrm{s}}}; \boldsymbol{a}^{\mathrm{pc}}_p(u_i), \boldsymbol{\Delta}_{p+1})$ with $\boldsymbol{\Delta}_{p+1} = [\Delta^{\mathrm{s}}_{p+1},\ \Delta^{\mathrm{h}}_{p+1}]^T$. Two types of model functions are possible as depicted in Figure 4.15. While the scenario in Figure 4.15a assumes a flat bottom as proposed in [Cook et al. (2006)], the height profile is taken into account in Figure 4.15b. In general, the model function can be described by

$$\Delta r(\boldsymbol{g}_{k_{\mathrm{s}}}; \boldsymbol{a}^{\mathrm{pc}}_p(u_i), \boldsymbol{\Delta}_{p+1}) = \underbrace{\left\| \boldsymbol{a}^{\mathrm{pc}}_p(u_i) - \boldsymbol{g}_{k_{\mathrm{s}}} \right\|_2}_{= r(k_{\mathrm{s}}, u_i)} - \underbrace{\left\| \boldsymbol{a}^{\mathrm{pc}}_p(u_i) + \boldsymbol{\Delta}_{p+1} - \boldsymbol{g}_{k_{\mathrm{s}}} \right\|_2}_{= r(k_{\mathrm{s}}, u_j)}, \tag{4.44}$$

where $\boldsymbol{g}_{k_{\mathrm{s}}}$ represents a ground-range grid point vector for the center of a sliding window with $k_{\mathrm{s}} = 1, \ldots, K$ and $\boldsymbol{a}^{\mathrm{pc}}_{p+1}(u_j) = \boldsymbol{a}^{\mathrm{pc}}_p(u_i) + \boldsymbol{\Delta}_{p+1}$. For the flat bottom assumption, the grid point vector is set to $\boldsymbol{g}_{k_{\mathrm{s}}} = [x_{k_{\mathrm{s}}}, 0, 0]^T$, where $x_{k_{\mathrm{s}}}$ is the ground-range center of the sliding window. In the case of available estimates of the seafloor height, *e.g.*, via broadside interferometry, the grid point vector is expressed as $\boldsymbol{g}_{k_{\mathrm{s}}} \equiv \hat{\boldsymbol{g}}_{k_{\mathrm{s}}} = [x_{k_{\mathrm{s}}}, 0, \hat{z}_{k_{\mathrm{s}}}]^T$. For the sake of completeness, the method suggested in [Cook et al. (2006)] is addressed in the sequel, which solves the minimization problem as stated in (4.43), taking into account height estimates of the seafloor.

Given the model function $\Delta r(\boldsymbol{g}_{k_{\mathrm{s}}}; \boldsymbol{a}^{\mathrm{pc}}_p(u_i), \boldsymbol{\Delta}_{p+1})$, the minimization problem is nonlinear in its parameters $\boldsymbol{\Delta}_{p+1}$ but can be solved iteratively, *e.g.*, using the Newton-Raphson method [Finckenstein et al. (2006)]. Omitting the ping index p for notational



(a) Flat bottom (b) Height profile

Figure 4.15: RPC displacement scenario as described in [Cook (2007)] for a flat bottom assumption (a) and for a height variation of the seafloor (b).

convenience in the sequel, the iterative procedure reads

$$\boldsymbol{\Delta}^{(i+1)} = \boldsymbol{\Delta}^{(i)} - \boldsymbol{H}^{-1}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta}^{(i)})^2\right)\ \boldsymbol{J}^T\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta}^{(i)})^2\right). \tag{4.45}$$

In (4.45), the residuals are given by $\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta}) = \Delta r(\boldsymbol{g}_{k_{\mathrm{s}}}; \boldsymbol{a}^{\mathrm{pc}}_p(u_i), \boldsymbol{\Delta}) - \Delta\hat{r}(k_{\mathrm{s}}, u_i)$, and the iteration index is denoted by i. Furthermore, $\boldsymbol{H}(\cdot)$ and $\boldsymbol{J}(\cdot)$ represent the Hessian and Jacobian matrix with respect to the multi-dimensional parameter $\boldsymbol{\Delta}$, respectively. The transposed Jacobian matrix of a two dimensional parameter $\boldsymbol{\Delta} = [\Delta_1, \Delta_2]^T$ is given by

$$\boldsymbol{J}^T\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) = \begin{bmatrix} \frac{\partial}{\partial\Delta_1} \\ \frac{\partial}{\partial\Delta_2} \end{bmatrix} \nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2 = 2\begin{bmatrix} \nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})\frac{\partial}{\partial\Delta_1}\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta}) \\ \nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})\frac{\partial}{\partial\Delta_2}\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta}) \end{bmatrix}, \tag{4.46}$$

and the Hessian matrix can be written as

$$\boldsymbol{H}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) = \begin{bmatrix} H_{11}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right), & H_{12}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) \\ H_{21}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right), & H_{22}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) \end{bmatrix}. \tag{4.47}$$

The elements of the Hessian matrix are found as follows

$$H_{11}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) = 2\left(\frac{\partial\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_1}\right)^2 + 2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})\frac{\partial^2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_1\partial\Delta_1}, \tag{4.48}$$

$$H_{12}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) = 2\frac{\partial\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_1}\frac{\partial\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_2} + 2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})\frac{\partial^2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_1\partial\Delta_2}, \tag{4.49}$$

$$H_{21}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) = 2\frac{\partial\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_2}\frac{\partial\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_1} + 2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})\frac{\partial^2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_2\partial\Delta_1}, \tag{4.50}$$

$$H_{22}\left(\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})^2\right) = 2\left(\frac{\partial\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_2}\right)^2 + 2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})\frac{\partial^2\nu(x_{k_{\mathrm{s}}}, \boldsymbol{\Delta})}{\partial\Delta_2\partial\Delta_2}. \tag{4.51}$$

Alternatively, the iteration method in (4.45) can be formulated in matrix-vector notation as follows

$$\boldsymbol{\Delta}^{(i+1)} = \boldsymbol{\Delta}^{(i)} - \left(1/2\,\boldsymbol{J}\big(\boldsymbol{\Delta}^{(i)}\big)^T\boldsymbol{J}\big(\boldsymbol{\Delta}^{(i)}\big) + \boldsymbol{S}\big(\boldsymbol{\Delta}^{(i)}\big)\right)^{-1}\ \boldsymbol{J}\big(\boldsymbol{\Delta}^{(i)}\big)^T\boldsymbol{\nu}\big(\boldsymbol{\Delta}^{(i)}\big), \tag{4.52}$$

where the residual vector is denoted by $\boldsymbol{\nu}\big(\boldsymbol{\Delta}^{(i)}\big) = [\nu(x_1, \boldsymbol{\Delta}^{(i)}), \ldots, \nu(x_K, \boldsymbol{\Delta}^{(i)})]^T$ and the matrix $\boldsymbol{S}\big(\boldsymbol{\Delta}^{(i)}\big)$ is obtained by

$$\boldsymbol{S}\big(\boldsymbol{\Delta}^{(i)}\big) = 2\sum_{k_\mathrm{s}=1}^{K} \nu\big(x_{k_\mathrm{s}}, \boldsymbol{\Delta}^{(i)}\big)\boldsymbol{H}\big(x_{k_\mathrm{s}}, \boldsymbol{\Delta}^{(i)}\big). \tag{4.53}$$

Note that in (4.53), the Hessian matrix $\boldsymbol{H}\big(x_{k_\mathrm{s}}, \boldsymbol{\Delta}^{(i)}\big) = \boldsymbol{H}\big(\nu(x_{k_\mathrm{s}}, \boldsymbol{\Delta})\big)|_{\boldsymbol{\Delta}=\boldsymbol{\Delta}^{(i)}}$ has to be evaluated for each discrete value of x_{k_s}, with $k_\mathrm{s} = 1, \ldots, K$. This general formulation can now be used to estimate sway and heave motion. To this end, the model function in (4.44) is considered with $\boldsymbol{g}_{k_\mathrm{s}} \equiv \hat{\boldsymbol{g}}_{k_\mathrm{s}} = [x_{k_\mathrm{s}}, 0, \hat{z}_{k_\mathrm{s}}]^T$ along with the phase center position $\boldsymbol{a}_p^{\mathrm{pc}} = [a_p^{\mathrm{x}}, 0, h_p^{\mathrm{og}}]^T$ as depicted in Figure 4.15b. It can be rewritten as

$$\begin{aligned}\Delta r\big(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}}\big) =& \sqrt{(a_p^{\mathrm{x}} - x_{k_\mathrm{s}})^2 + (h_p^{\mathrm{og}} - \hat{z}_{k_\mathrm{s}})^2} \\ &- \underbrace{\sqrt{(a_p^{\mathrm{x}} + \Delta_{p+1}^{\mathrm{s}} - x_{k_\mathrm{s}})^2 + (h_p^{\mathrm{og}} + \Delta_{p+1}^{\mathrm{h}} - \hat{z}_{k_\mathrm{s}})^2}}_{= r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}.\end{aligned} \tag{4.54}$$

The first-order partial derivatives can then be derived as follows

$$\frac{\partial \Delta r(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}{\partial \Delta_{p+1}^{\mathrm{s}}} = -\frac{a_p^{\mathrm{x}} + \Delta_{p+1}^{\mathrm{s}} - x_{k_\mathrm{s}}}{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}, \tag{4.55}$$

$$\frac{\partial \Delta r(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}{\partial \Delta_{p+1}^{\mathrm{h}}} = -\frac{h_p^{\mathrm{og}} + \Delta_{p+1}^{\mathrm{h}} - \hat{z}_{k_\mathrm{s}}}{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}. \tag{4.56}$$

Moreover, the second-order partial derivates can be determined by

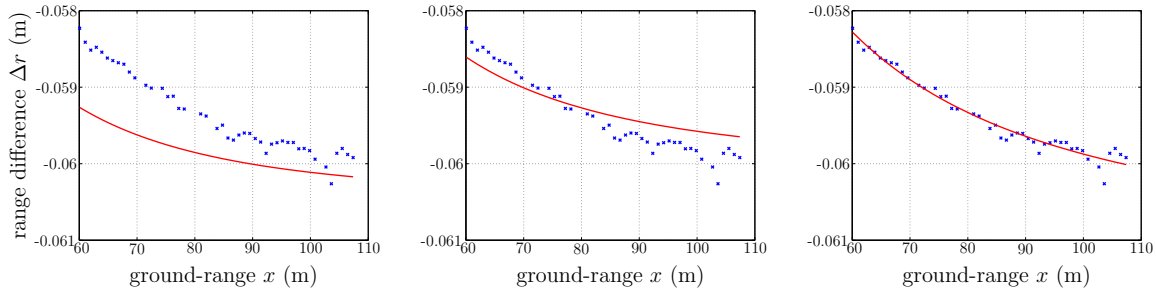
$$\frac{\partial^2 \Delta r(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}{\partial \Delta_{p+1}^{\mathrm{s}} \partial \Delta_{p+1}^{\mathrm{s}}} = -\frac{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})^2 - (a_p^{\mathrm{x}} + \Delta_{p+1}^{\mathrm{s}} - x_{k_\mathrm{s}})^2}{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})^3}, \tag{4.57}$$

$$\frac{\partial^2 \Delta r(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}{\partial \Delta_{p+1}^{\mathrm{h}} \partial \Delta_{p+1}^{\mathrm{h}}} = -\frac{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})^2 - (h_p^{\mathrm{og}} + \Delta_{p+1}^{\mathrm{h}} - \hat{z}_{k_\mathrm{s}})^2}{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})^3}, \tag{4.58}$$

$$\frac{\partial^2 \Delta r(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}{\partial \Delta_{p+1}^{\mathrm{s}} \partial \Delta_{p+1}^{\mathrm{h}}} = \frac{(a_p^{\mathrm{x}} + \Delta_{p+1}^{\mathrm{s}} - x_{k_\mathrm{s}})(h_p^{\mathrm{og}} + \Delta_{p+1}^{\mathrm{h}} - \hat{z}_{k_\mathrm{s}})}{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})^3}, \tag{4.59}$$

$$\frac{\partial^2 \Delta r(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})}{\partial \Delta_{p+1}^{\mathrm{h}} \partial \Delta_{p+1}^{\mathrm{s}}} = \frac{(a_p^{\mathrm{x}} + \Delta_{p+1}^{\mathrm{s}} - x_{k_\mathrm{s}})(h_p^{\mathrm{og}} + \Delta_{p+1}^{\mathrm{h}} - \hat{z}_{k_\mathrm{s}})}{r_2(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}}, \Delta_{p+1}^{\mathrm{h}})^3}. \tag{4.60}$$

Substituting the partial derivatives into the expression for the Hessian and Jacobian matrices and applying the iteration method as stated in (4.52) yields an estimate of the ground-range sway $\hat{\Delta}_{p+1}^{\mathrm{s}}$ and heave $\hat{\Delta}_{p+1}^{\mathrm{h}}$. An example of the iterative fitting procedure is depicted in Figure 4.16 showing three iterations to estimate the displacement parameters. While the blue data points represent the scaled time delay estimates $c\,\Delta\hat{\tau}(x_{k_\mathrm{s}})/2$, the red solid line indicates the current adaptation of the model function $\Delta r\big(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta_{p+1}^{\mathrm{s}\,(i)}, \Delta_{p+1}^{\mathrm{h}\,(i)}\big)$ according to the parameters $\Delta_{p+1}^{\mathrm{s}\,(i)}$ and $\Delta_{p+1}^{\mathrm{h}\,(i)}$ with $i = 1, \ldots, 3$ in Figure 4.16a-c. After three iterations, a reasonable fit is attained. Typically, a fitting error and a maximum number of iterations are used as a terminating condition.



(a) First iteration (b) Second iteration (c) Third iteration

Figure 4.16: Results of the iterative fitting procedure are depicted in (a)-(c). The model function $\Delta r\left(x_{k_\mathrm{s}}, \hat{z}_{k_\mathrm{s}}; \Delta^{\mathrm{s}}_{p+1}, \Delta^{\mathrm{h}}_{p+1}\right)$ with broadside bathymetry estimates $\hat{z}_{k_\mathrm{s}}$ is fitted to the time delay differences $\Delta\hat{r}(k_\mathrm{s})$ to estimate sway and heave motion.

4.2.7 Motion estimation and SAS image examples

In this section, a selected collection of defocused and focused SAS images that have been reconstructed using the proposed processing chain is illustrated. The image quality difference highlights the importance of accurately estimating translational motion. Additionally, a comparison of data-driven estimates of ping-to-ping sway and heave motion with their INS counterparts points out the imprecision of the INS in tracking translations. All SAS image examples using real sonar measurements have been reconstructed with a flat bottom height grid for the backprojection algorithm. Moreover, the same assumption has been applied for the introduced motion compensation technique. At this point in time, a validation of SAS image quality improvement by using available bathymetry information is still open. This is mainly due to a current shortage of collected interferometric data sets that contain a rough seafloor topography. Instead, synthetic data is used to verify the theoretical considerations about the need of height information as discussed in Sections 3.5 and 4.2.6. For both synthetic data and real sonar measurements, the imaging process applies the introduced concepts such as synthetic aperture shading, phase wrap compensation as well as near-field and widebeam correction via broadside beamforming.

The first example of a defocused SAS image is depicted in Figure 4.17 with a dynamic range of 30 dB. It shows a seafloor area of size 40×60 m^2 that partly consists of sand ripples. The corresponding raw data measurements have been recorded by the VISION SAS system. The general subjective impression about the image is a lack of detail along with a low contrast. However, sand ripple details are still clearly visible. Here, the image reconstruction process has solely relied on rotational and translational motion information provided by the INS unit. Contrarily, the SAS image in Figure 4.18, which

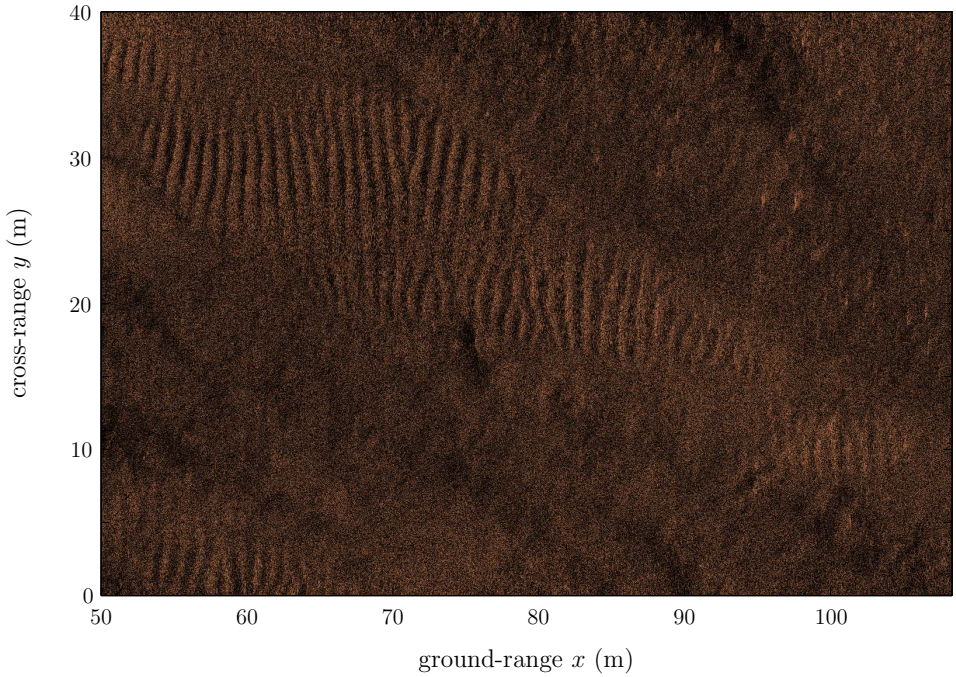


Figure 4.17: Defocused SAS image of the VISION SAS system. Image reconstruction uses only position information provided by the INS.

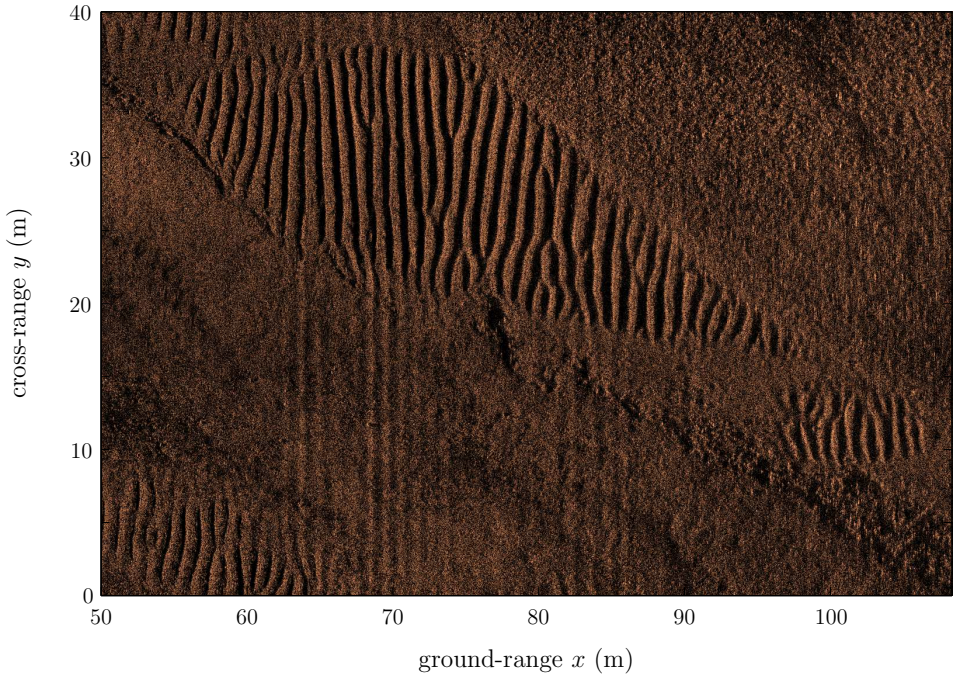
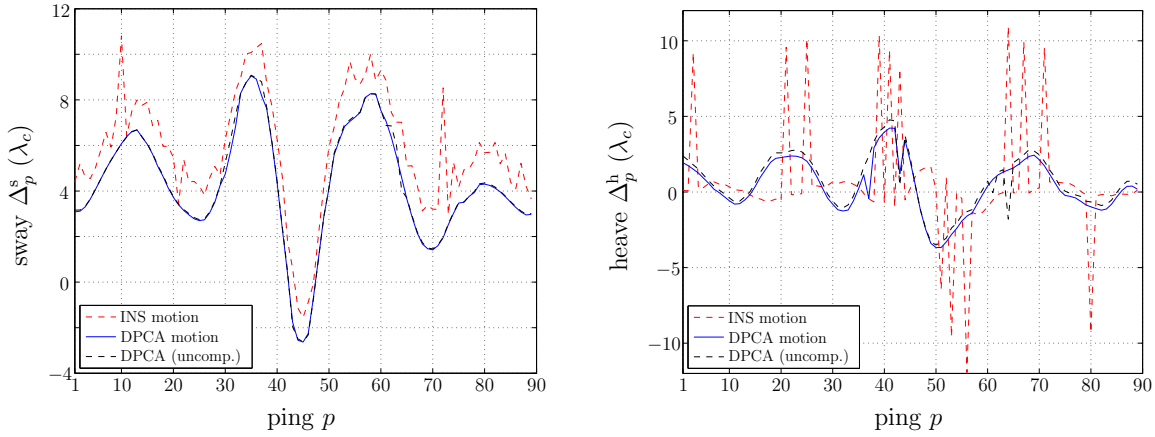


Figure 4.18: Focused SAS image of the VISION SAS system. Imaging uses rotational position information provided by the INS and translational estimates using DPCA.

shows the same scene content, is focused and reveals a significant gain in detail, *e.g.*, all contour lines of the sand ripples appear sharp and the contrast is increased. The latter is especially noticeable between the crest and trough region of sand ripples. Furthermore, the overall structure of the remaining seafloor parts becomes apparent with fine textural details.

For the second SAS image example, micronavigation estimates the translational motion degrees of freedom, while rotational motion information is still provided by the INS. The corresponding sway and heave estimates are illustrated in Figure 4.19 in terms of the carrier wavelength λ_c (blue line) along with the INS counterparts (red dashed line) for a total number of $M_p = 90$ pings. Furthermore, motion estimates are shown (black dashed line) that have been obtained without applying the beamforming compensation technique in order to compensate near-field and widebeam effects. As mentioned in Section 4.2.5, the fitting approach is capable of handling the varying time delays over the RPC dimension for estimating sway. This is noticeable by a nearly identical behavior of the sway motion between the black and blue line in Figure 4.19a. Both sway estimates differ significantly from the INS sway motion, which explains the defocusing of the SAS image in Figure 4.17. For heave motion, the estimates are slightly biased without beamforming compensation. This is demonstrated by a small constant offset between the black and blue line in Figure 4.19b. As the heave estimates represent ping-to-ping motion, the absolute altitude positions are obtained after integration. Hence, the constant offset becomes a linear trend, and the heave discrepancy increases up to a maximum of approximately 0.2 m in this example. Note that the outlier



(a) Sway comparison (b) Heave comparison

Figure 4.19: VISION example for comparing ping-to-ping sway (a) and heave motion (b) of the INS (red dashed line) with the estimated motion using near-field and widebeam compensation (blue line) or without compensation (black dashed line).

values in the INS heave motion result from fusing INS data with fluctuating altitude measurements. In summary, this means that although the time delay differences vary over the RPC dimension, their impact is negligible for estimating translational motion without affecting SAS image quality. However, applying the beamforming compensation yields higher correlation values as shown in Figure 4.13. Further, the method should be considered when estimating rotational motion errors by a data-driven micronavigation approach or if larger physical arrays are employed for SAS systems.

Another image example is depicted in Figure 4.20, where the raw echo measurements have been recorded by the experimental MCM-SLS SAS system – see Section 3.1. The scene content has an extent of 30 m in along-track and 35 m in ground-range direction and is mainly showing several small rocks together with their corresponding shadows. The latter occur as the regions behind objects are not captured by the sonar. Moreover, the metal cross target has been laid in the area to test the resolution capability of the SAS system – see Figure 3.2a for an optical image of the object. The target is located at the approximate coordinates of $(55, 26)$ m. A zoom-out around the object is additionally shown in Figure 4.20, which points out that the individual

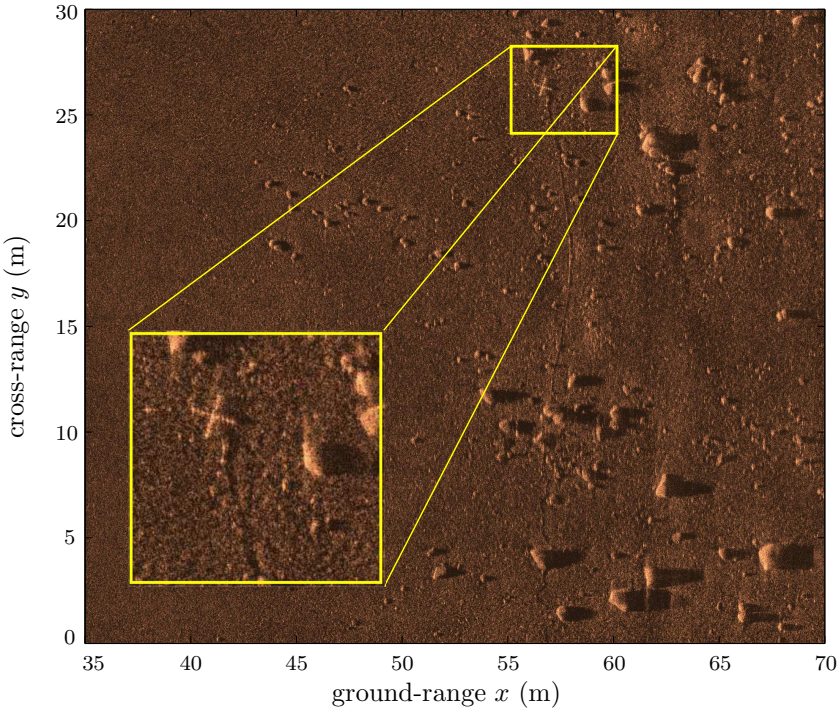


Figure 4.20: Focused SAS image example of the experimental MCM-SLS SAS system. The scene content consists of small rocks and their associated shadows. A small area is zoomed out to enlarge the metal cross target in order to validate whether the attached plastic balls are clearly resolved.

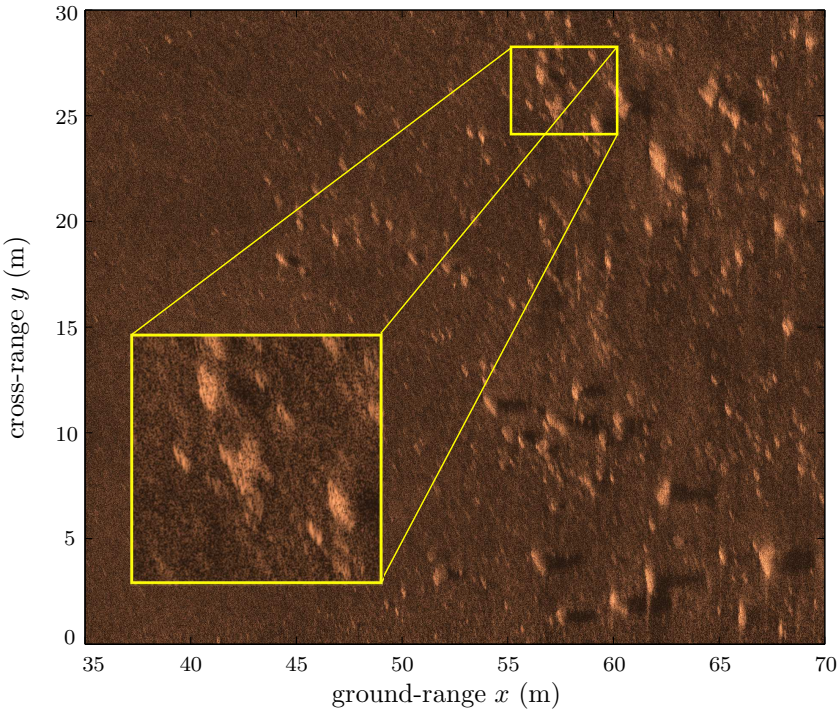


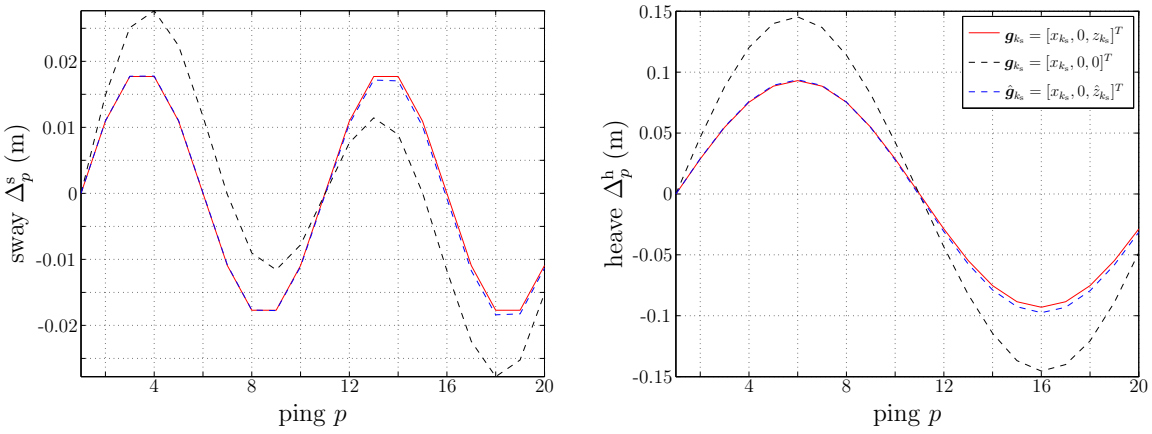
Figure 4.21: Defocused SAS image example of the experimental MCM-SLS SAS system. A small area is zoomed out to enlarge the metal cross resolution target. The entire object is blurred and high-resolution is not attained due to inaccurate motion tracking.

plastic balls attached to the cross are clearly resolved. The general subjective quality impression of the image is very high. By closer inspection, even a cord attached to the object is recognizable throughout the entire ground-range swath between $[55, 60]$ m. The cord is used to pull the object out of the water. While for the SAS reconstruction of Figure 4.20 the introduced motion compensation approach has been applied, the imaging process of the SAS image depicted in Figure 4.21 has solely relied on INS information. In contrast to Figure 4.20, the entire scene is totally defocused and the details of the man-made target are no longer identifiable. This becomes especially apparent in the zoom-out highlighted in Figure 4.21.

A major aspect that influences the quality of SAS images is a strong bathymetry variation of the target scene. As outlined in Section 3.5, it is essential to have at least rough knowledge about the seafloor height for ground-range SAS imaging algorithms instead of simply assuming a flat bottom for the entire scenery. Otherwise, even under perfect trajectory knowledge, image degradation may occur for nonlinear pathways [Jakowatz et al. (1996)]. Besides the construction of an *a priori* height grid, it is also important to use bathymetry information for the nonlinear least squares method in order to correctly adapt the model function – see Section 4.2.6, and therefore, to

avoid biased motion estimates. The latter occur due to a wrong modeling of the underlying time delay differences between redundant phase centers. Subsequently, the consequences of an incorrect modeling are exemplarily demonstrated based on synthetic data, which has been generated using system parameters similar to the VISION SAS system as described in Section 3.1. To investigate the impact, two scenarios are assumed: (i) a flat bottom at height level zero and (ii) the availability of height estimates via the broadside bathymetry estimation technique as outlined in Section 3.5. The outcome for this example is illustrated in Figure 4.22a and Figure 4.22b for sway and heave motion, respectively. For both motion directions, a sinusoidal path deviation function has been introduced into the synthetic data generation process. While the cycle per synthetic aperture length frequency is set to $f_{\mathrm{p}} = 2$ for sway motion, it is chosen as $f_{\mathrm{p}} = 1$ for the heave. The corresponding motion amplitudes are given by $A^s = \lambda_c$ and $A^h = 5\lambda_c$, respectively. A comparison with the ground truth of the ping-to-ping motion reveals a bias in the estimation procedure for both motion directions if a flat bottom with $z_{kl} \equiv 0 \ \forall \ k, l$ is assumed. By contrast, including bathymetry information yields accurate estimates for ground-range sway and heave motion over the entire number of processed pings. The bathymetry variation of the artificial seafloor used in this example is depicted in Figure 3.13a.

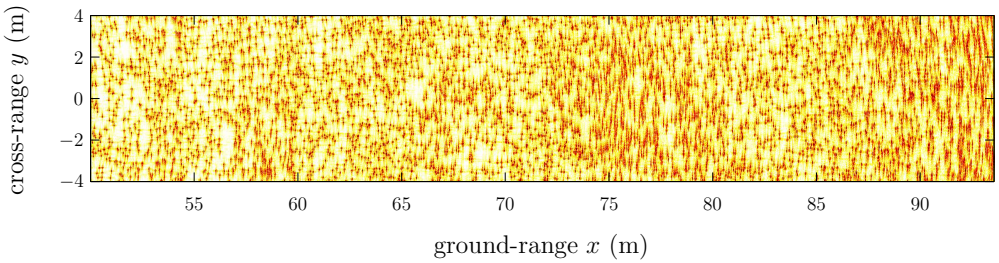
The consequences of missing bathymetry information on SAS image quality are depicted in Figure 4.23a, where two regions show a degradation of the PSF in along-track direction, namely, in the intervals $x \in [75, 80]$ m and $x \geq 90$ m. This degradation effect becomes significantly worse if a height grid is available for the



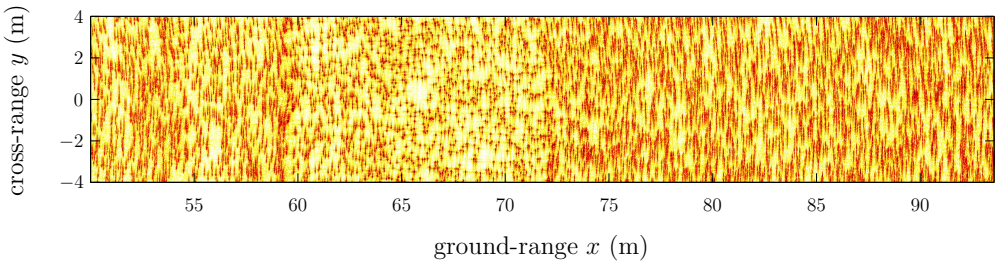
(a) Sway estimates (b) Heave estimates

Figure 4.22: Synthetic data example for comparing ping-to-ping motion estimates of sway (a) and heave (b) given the ground truth (red line), under the assumption of a flat bottom (black dashed line) and using the broadside bathymetry estimates of the seafloor height variation (blue dashed line).

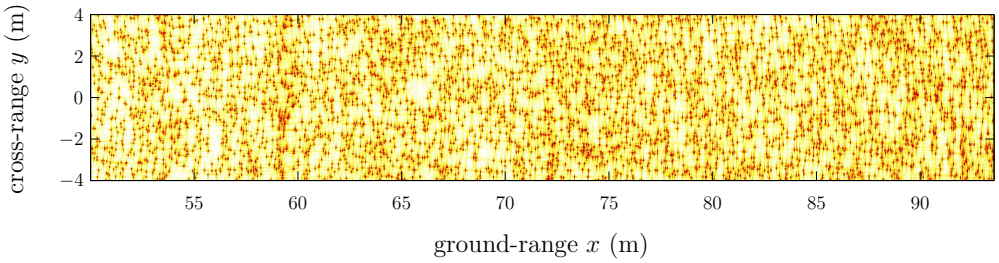
image reconstruction but its use inside the motion compensation is neglected. This is highlighted in Figure 4.23b. In conclusion, it appears that the assumption of a flat bottom in both, the backprojection algorithm and DPCA, is able to compensate missing bathymetry knowledge up to certain extent. However, finding an exact relation between biased motion estimates and its (positive) influence on compensating missing height information for image reconstruction as well as an upper uncertainty bound similar to the depth of focus (DOF) in (3.30) remains the subject of future work. Finally, the combined use of bathymetry information leads to the SAS image depicted in Figure 4.23c that shows an improved image quality in terms of well focused point targets for the entire scene including the above-mentioned regions.



(a) Flat bottom assumption for DPCA and image reconstruction



(b) Flat bottom assumption for DPCA



(c) Using bathymetry for DPCA and image reconstruction

Figure 4.23: A comparison of SAS images using synthetic VISION data, which have been reconstructed using different assumptions about available bathymetry information for imaging and motion compensation in (a)-(c).

4.3 Stripmap autofocus modifications

The objective of autofocus techniques is to enhance degraded synthetic aperture images, which lack quality due to residual motion errors. Typically, autofocus methods are divided into two groups, depending on the two common operating modes of synthetic aperture systems, namely, stripmap and spotlight [Carrara et al. (1995), Soumekh (1999)]. For the latter, the well-known phase gradient autofocus (PGA) [Wahl et al. (1994), Jakowatz et al. (1996)] algorithm nearly attains a diffraction limited restoration of defocused imagery by compensating residual phase errors directly in the image domain. As the raw echo measurements remain untouched, the use of the PGA is computationally attractive and its applicability has been proven for a wide-variety of scenes in synthetic aperture radar (SAR). However, it assumes (i) a range-invariant blurring model of the PSF and (ii) a visibility of each target along the entire synthetic aperture. Both assumptions restrict the use of the PGA algorithm mainly to spotlight synthetic aperture images.

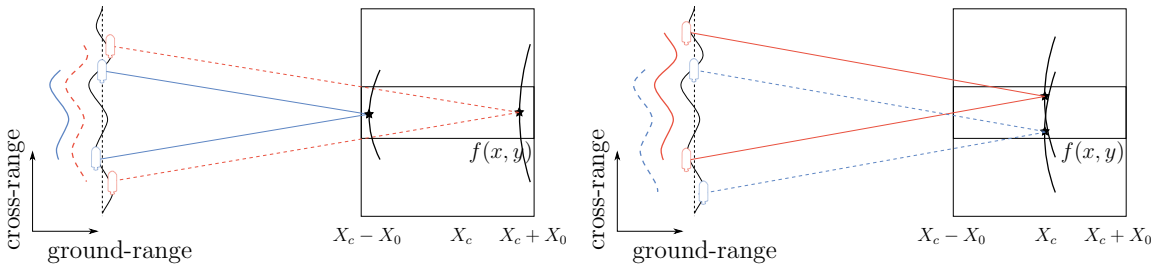
For stripmap imaging, the phase history data is limited by the angular beamwidth of the physical aperture, therefore, capturing only parts of the target area instead of the entire scene as in spotlight imaging – see Chapter 2 and Chapter 3. Consequently, the effect of motion errors on the PSF is spatially varying as depicted in Figure 4.24, and the use of autofocus methods is not straightforward anymore [Callow (2003)]. Extensions such as the phase curvature autofocus [Wahl et al. (1994a)] assume a narrowband synthetic aperture system to overcome the range-dependency of the blurring model by reducing the underlying blurring model to 1-D. However, this discards the 2-D character of the stripmap blurring function [Callow et al. (2003a), Wahl et al. (1994a)]. Especially for SAS applications, mostly wideband signals occur, making this approach inapplicable. Another algorithm called stripmap phase gradient autofocus algorithm was introduced in [Callow et al. (2003)] to work with fewer approximations. Rather than focusing the blurred image directly by applying an iterative phase correction, it estimates the path deviations of the imaging platform similarly to the introduced micronavigation technique. Given the estimated platform path, a synthetic aperture image is then reconstructed with an enhanced quality. This procedure may require several iterations to obtain a well-focused image, which leads to an extremely high computational load. As a consequence, stripmap PGA is inferior to micronavigation, which works on a sequential ping-to-ping basis.

In order to combine an adequate stripmap blurring model with the ability to work directly on image data, a different approach was introduced in [Bonifant (1999)]. It is named mosaic phase gradient autofocus (mPGA). The method applies the standard

PGA to overlapping cross-range segments of stripmap SAS images to handle the spatially varying nature of the PSF – see Figure 4.24a. Each segment is then processed individually by the PGA and finally reassembled to form the focused SAS image. While this approach overcomes the range-dependency of stripmap phase errors and is at the same time computationally attractive, it neglects targets at different cross-range locations – see Figure 4.24b. In doing so, the varying error information along the synthetic aperture is ignored within an individually processed image segment. This may place an emphasis on correctly focusing most prominent targets, while weaker targets become even more blurred. Additionally, the individual processing may cause a linear misalignment during the stitching process of the cross-range segments due to residual linear phase errors. The latter typically arise during the phase estimation process of the PGA algorithm. In the following, an extended version of the mPGA is proposed [Leier et al. (2013b)] to alleviate the original shortcomings. By dividing the entire image into true mosaics, *i.e.*, into 2-D sub-images, the new method considers the cross-range dependency of motion errors. Moreover, the linear shift problem is handled by a co-registration of the individual sub-images before and after each iteration of the PGA technique. The validity of the new technique is demonstrated using synthetic data simulations and a full-reference image quality method that is applied to measure and compare the focusing capability with the standard autofocus techniques. An example based on a real SAS image is then provided that contains the man-made object depicted in Figure 3.2b. Although the proposed modifications are similar to the ideas presented in [Zhu et al. (2011)], employing a co-registration process based on the cross-correlation is a more systematic approach to mitigate the linear shift problem than arbitrarily identifying a strong scatterer and comparing its position shift. This method fails in the absence of strong scatterers and cannot be compensated for by removing the linear trend in the phase gradient as it is claimed in [Zhu et al. (2011)]. Synthetic data simulations are more suitable to validate the proposed method, and the real SAS image example clearly demonstrates an improvement using the modified autofocus technique in contrast to the image comparison presented in [Zhu et al. (2011)].

4.3.1 Image degradation model

In general, autofocus techniques require a degradation model that relates the focused image to the degraded image by taking into account the characteristics of the imaging mode. Given an appropriate model, the phase error is then estimated and corrected accordingly. In contrast to spotlight operation, each target contains different information on the residual motion errors in stripmap imaging. This is illustrated in Figure 4.24, where two targets are located at different cross-range and ground-range



(a) Different ground-range positions (b) Different cross-range positions

Figure 4.24: Stripmap imaging scenario with sway motion along the synthetic aperture that is observed differently by two point targets depending on their location. The point targets in (a) are located at the same cross-range coordinate but at different ground-range positions and *vice versa* in (b).

positions. Here, X_c and X_0 denote the center of the imaging scene and half of the swath width in ground-range direction, respectively. In the sequel, the blurring model of both operation modes is introduced and discussed. To this end, the wavenumber transform [Soumekh (1999), Callow et al. (2003a)] of a single transceiver synthetic aperture system, which is characterized by

$$p\Delta^{\mathrm{A}} = y - \frac{k_y}{k_x}\big(X_c + x\big), \tag{4.61}$$

is used to relate the spatial wavenumber frequencies k_x and k_y to the aperture position $p\Delta^{\mathrm{A}}$. Note that only ground-range sway motion $\Delta^{\mathrm{s}}(p) \equiv \Delta^{\mathrm{s}}_p$ is considered in the following as it mainly affects the image degradation given common SAS geometries. For ease of notation, the physical aperture position is expressed here by $p \equiv p\Delta^{\mathrm{A}}$, dropping the APP. Furthermore, the continuous spatial domain is considered to omit the discrete sampling subscripts for the sake of clarity. Using the wavenumber transform in (4.61), a relationship between the 2-D Fourier transform of the blurred image $\tilde{F}(k_x, k_y)$ and the wavenumber representation of the ideal reconstructed image $\hat{F}(k_x, k_y)$ can then be found as follows [Callow (2003)]

$$\tilde{F}(k_x, k_y) \approx \hat{F}(k_x, k_y) e^{jk_x\Delta^{\mathrm{s}}\left(y - \frac{k_y}{k_x}(X_c + x)\right)}, \tag{4.62}$$

which represents the blurring model for stripmap operation. The blurring model highlights the spatial dependency of the PSF and is only approximate due to a small change in the stationary phase point in the presence of sway errors [Callow et al. (2003a)]. In spotlight imaging, especially for SAR, a narrowband ($k \approx k_c$) and a narrowbeam system ($k_x \approx 2k$) can often be assumed [Soumekh (1999)]. Moreover, given common SAR geometries, *i.e.*, either air-borne or space-borne, the condition of a small swath-extent compared to the standoff ground-range distance of the target scene center with $X_0, Y_0 \ll X_c$ is normally fulfilled. Consequently, the

wavenumber transform simplifies to

$$p = -\frac{X_c}{2k_c} k_y, \tag{4.63}$$

where k_c is the wavenumber at the carrier frequency [Callow et al. (2003a)]. Thus, the relation between aperture position p and cross-range wavenumber k_y simplifies to a scaling factor. Substituting (4.63) into (4.62) and combining the narrowbeam and narrowband assumption by choosing $k_x \approx 2k_c$ and setting $\varsigma = -X_c/(2k_c)$ yields

$$\tilde{F}(k_x, k_y) \approx \hat{F}(k_x, k_y)\, e^{j2k_c\Delta^{\mathrm{s}}(\varsigma k_y)}, \tag{4.64}$$

which represents the common spotlight blurring model. The latter highlights that a sway error $\Delta^{\mathrm{s}}(p)$ is invariant in range and causes blurring in cross-range direction. As a consequence, autofocus techniques such as the PGA produce reliable focusing results by exploiting range redundancy via averaging.

4.3.2 Phase gradient autofocus

As PGA [Wahl et al. (1994), Jakowatz et al. (1996)] is the main part of the proposed autofocusing technique, a brief review of its basic idea is presented in this section. PGA directly operates on the erroneous image data, $\tilde{f}(x_k, y_l)$, with $k = 1, \ldots, M_x$ and $l = 1, \ldots M_y$, and applies the blurring model of (4.64). It consists of four main processing steps referred to as circular shifting, windowing, phase gradient estimation, and iterative phase correction, which are briefly described below:

1. Circular shifting: For each range bin x_k, with $k = 1, \ldots, M_x$, this processing block selects the cross-range position of the most dominant scatterer in the blurred image $\tilde{f}(x_k, y_l)$ as follows

$$\bar{y}(x_k) = \arg\max_{y_l} \Big| f(x_k, y_l) \Big|. \tag{4.65}$$

   Thereupon, each range bin x_k is circularly shifted by the corresponding cross-range position $\bar{y}(x_k)$ of the most dominant scatterer, which is denoted by

$$\tilde{f}^{\mathrm{c}}(x_k, y_l) = \tilde{f}\big(x_k, y_l - \bar{y}(x_k)\big). \tag{4.66}$$

   This operation avoids the occurrence of a linear phase term in the wavenumber domain due to the cross-range position of a target. Simultaneously, it maintains the information about the phase error.

2. Windowing: In the second step, a rectangular window function $\Pi(y_l)$ truncates the circular shifted image $\tilde{f}^{\mathrm{c}}(x_k, y_l)$ as follows

$$\tilde{f}^{\mathrm{w}}(x_k, y_l) = \tilde{f}^{\mathrm{c}}(x_k, y_l)\, \Pi(y_l), \tag{4.67}$$

which dismisses the phase history data of weak targets. The latter typically act as noise and should not contribute to the phase error estimation. Thus, after windowing, only data with the highest SNR is considered.

3. Phase gradient estimation: The estimation of the phase gradient $\Delta\Phi(k_{y_l})$ is the core of PGA and is performed in the cross-range wavenumber domain, *i.e.*, using $\tilde{F}^{\mathrm{w}}(x_k, k_{y_l})$. The maximum likelihood estimate of the phase error gradient [Wahl et al. (1994)] is given by

$$\Delta\hat{\Phi}(k_{y_l}) = \angle \left\{ \sum_{m=1}^{M_x} \tilde{F}^{\mathrm{w}}(x_m, k_{y_l})\; \tilde{F}^{\mathrm{w}}(x_m, k_{y_{l-1}})^* \right\}, \text{ for } l = 2, \ldots, M_y, \tag{4.68}$$

assuming a circular-symmetric complex normal distribution of image pixels, which are mutually uncorrelated in ground-range direction and cross-range direction. In (4.68), $\angle\{\cdot\}$ denotes the angle operator. The unknown path deviation in the wavenumber domain is then estimated by

$$\hat{\Delta}^{\mathrm{s}}(k_{y_l}) = \frac{1}{2k_c} \sum_{l'=1}^{l} \Delta\hat{\Phi}(k_{y_{l'}}), \text{ with } \Delta\hat{\Phi}(k_{y_1}) = 0. \tag{4.69}$$

4. Iterative phase correction: After removing any residual linear phase trend, phase correction is performed by multiplying the conjugate of the estimated phase error as follows

$$\hat{F}(x_k, k_{y_l}) = \tilde{F}(x_k, k_{y_l})\, e^{-j2k_c \hat{\Delta}^{\mathrm{s}}(\varsigma k_{y_l})}, \tag{4.70}$$

with $\varsigma = -X_c/(2k_c)$. An inverse Fourier transform of $\hat{F}(x_k, k_{y_l})$ to the spatial domain yields the phase corrected image $\hat{f}(x_k, y_l)$. The processing steps 1–4 are then performed iteratively until a convergence criterion is fulfilled.

4.3.3 Mosaic phase gradient autofocus

In order to overcome the discussed limitations of the PGA algorithm for stripmap imaging, a mosaic phase gradient autofocus (mPGA) method was proposed in [Bonifant (1999)]. The mPGA algorithm applies the standard PGA algorithm to

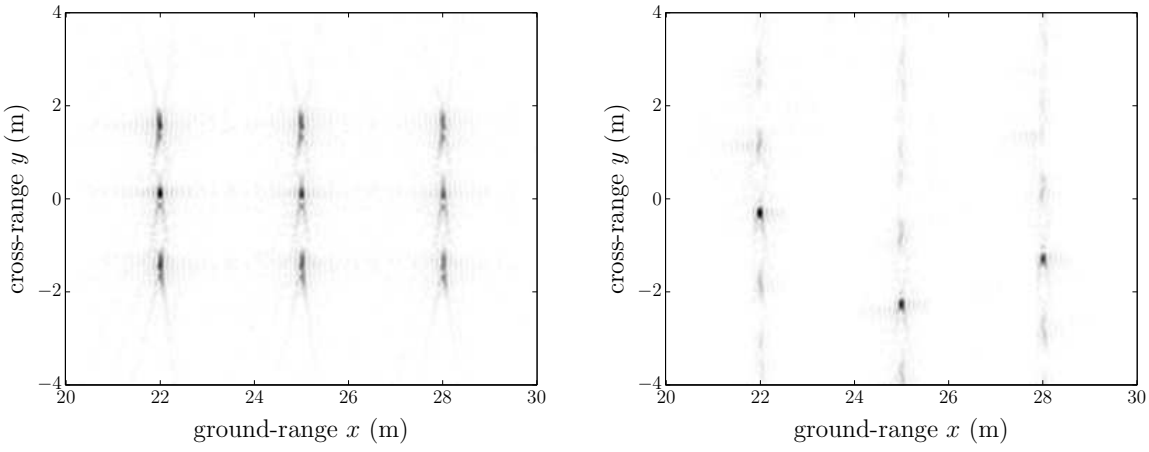
overlapping cross-range image segments, implicitly assuming a negligible range-dependency of the blurring function inside each segment. This principle is strongly related to the idea of local stationarity. Given erroneous image data $\tilde{f}(x_k, y_l)$ with $k = 1, \ldots, M_x$ and $l = 1, \ldots, M_y$, an cross-range image segment i is extracted using the expression

$$\tilde{f}_i(x_k, y_l) = \tilde{f}(x_{k+(i-1)M'_x}, y_l), \text{ for } k = 1, \ldots, M'_x. \tag{4.71}$$

Here, each segment i, with $i = 1, \ldots, K_x$, consists of M'_x ground-range bins. In total, there are K_x segments to be processed by the PGA. By choosing a small width M'_x, the phase error can be assumed constant in ground-range direction inside a segment and, as a consequence, the spotlight blurring model can be applied. However, a small segment width leads to less range redundancy, which affects the estimation accuracy of the phase error.

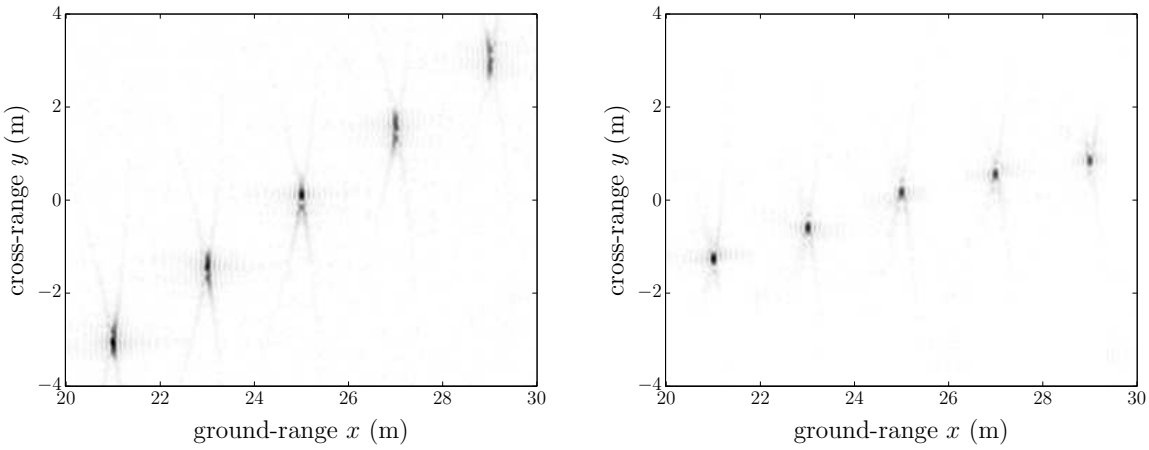
In a next step, each segment is used as an input to the standard PGA as described in Section 4.3.2, and is focused separately. Afterwards, the individual image segments are stitched together to obtain the focused SAS image. In doing so, the mPGA accounts for the range-dependency of the motion error $\Delta^s(p)$ but it still neglects that targets contain different fractions of the phase error at varying cross-range positions as depicted in Figure 4.24b. In this case, the PGA selects the brightest target, shifts it towards the center followed by a truncation. Thereby, other targets are discarded. Moreover, the estimated phase error is potentially applied to the entire cross-range segment, ignoring that other targets suffer from different phase errors. This scenario is demonstrated exemplarily in Figure 4.25 using synthetic data. Here, the targets selected by the PGA are focused accurately but others in the same segment are severely blurred and vanish. Another issue related to the mPGA is the inability of the PGA to estimate linear phase errors [Callow et al. (2003a), Wahl et al. (1994a)], which normally induces an entire image shift in cross-range direction. However, for the mPGA, this may lead to varying cross-range shifts for different image segments, which causes problems in the reassembling process. This is highlighted by the cross-range position shifts of the focused point targets in Figure 4.25b compared to Figure 4.25a.

Another example is shown in Figure 4.26. Although the individual segments are focused correctly in Figure 4.26b, the reassembled image is staggered. Thus, the mPGA is not a preferred choice as a post-processing autofocusing technique for the proposed SAS processing chain; but an alternative has been developed to deal with the aforementioned problems. It is described in the next section.



(a) Blurred image (b) Processed by mPGA

Figure 4.25: Illustration of mPGA focusing issue, which leads to wrong phase error estimates for non-prominent targets displaced in cross-range direction.



(a) Blurred image (b) Processed by mPGA

Figure 4.26: Illustration of mPGA focusing issue, which leads to a staggered image due to varying linear shifts for different image segments.

4.3.4 Modified mosaic phase gradient autofocus

In order to overcome the introduced drawbacks of the mPGA, a modified version is introduced called modified mosaic phase gradient autofocus (mmPGA) algorithm. It accounts for both the cross-range dependency of the error and varying linear shifts for different image segments. The proposed modifications are described in the following. In order to account for different cross-range positions of targets, the synthetic aperture image is additionally divided in cross-range direction to obtain proper mosaic pieces. The issue of varying linear shifts among image mosaic pieces is then addressed by

estimating the individual linear shifts via cross-correlation of sub-images before and after each iteration of the PGA. A block diagram of the mmPGA is depicted in Figure 4.27, which provides an overview of the individual processing steps. The blurred input image $\tilde{f}(x_k, y_l)$ is decomposed into a total number of $K_x K_y$ mosaic pieces $\tilde{f}_{i,j}(x_k, y_l)$ with $i = 1, \ldots, K_x$ and $j = 1, \ldots, K_y$. Each mosaic has a width of $k = 1, \ldots, M'_x$ ground-range bins and a height of $l = 1, \ldots, M'_y$ cross-range pixels. Inside such a mosaic, it is assumed that every target is exposed to the same cross-range section of the motion error along the synthetic aperture. Thus, the phase error estimate can be applied to the entire mosaic piece.

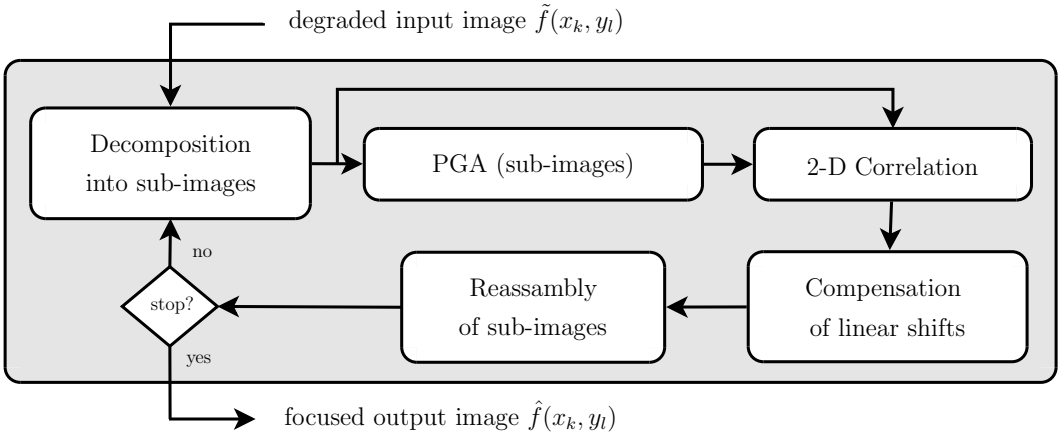


Figure 4.27: Block diagram of the proposed modified mPGA.

In order to handle the varying linear shifts among different mosaic pieces, the initial input mosaic $\tilde{f}_{i,j}(x_k, y_l)$ is registered to the output piece $\hat{f}_{i,j}(x_k, y_l)$ by applying the normalized cross-covariance function. This procedure is repeated in each iteration of the PGA to estimate the occurring linear shifts. The cross-covariance function between mosaic pieces $\hat{f}_{i,j}(x_k, y_l)$ and $\tilde{f}_{i,j}(x_k, y_l)$ for $M'_x = M'_y$ is estimated by

$$\hat{\rho}_{i,j}(\kappa_x, \kappa_y) = \frac{1}{\hat{\sigma}^{(1)}_{i,j} \hat{\sigma}^{(2)}_{i,j} {M'_x}^2} \sum_k \sum_l \bar{f}^{(1)}_{i,j}(x_k, y_l) \bar{f}^{(2)}_{i,j}(x_k + \kappa_x \Delta x, y_l + \kappa_y \Delta y), \tag{4.72}$$

where

$$\bar{f}^{(1)}_{i,j}(x_k, y_l) = \left| \hat{f}_{i,j}(x_k, y_l) \right| - \hat{\mu}^{(1)}_{i,j} \tag{4.73}$$

and

$$\bar{f}^{(2)}_{i,j}(x_k, y_l) = \left| \tilde{f}_{i,j}(x_k, y_l) \right| - \hat{\mu}^{(2)}_{i,j} \tag{4.74}$$

are the mean subtracted input and output intensities of the corresponding mosaic pieces. The sample mean and sample standard deviation of the mosaic piece $f_{i,j}(x_k, y_l)$

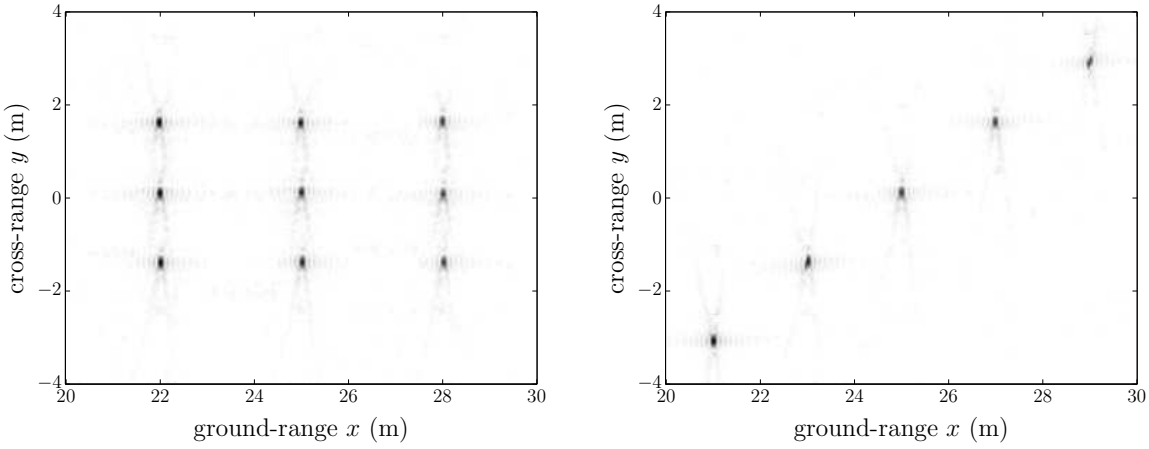
are denoted by $\hat{\mu}_{i,j}$ and $\hat{\sigma}_{i,j}$, respectively. Moreover, Δx represents the pixel size in ground-range direction and Δy in cross-range direction. The maximum peak correlation value is then related to the spatial shift between mosaic pieces as follows

$$\left[\kappa_{x_0}, \kappa_{y_0}\right] = \arg \max_{\kappa_x,\, \kappa_y} \left| \hat{\rho}_{i,j}(\kappa_x, \kappa_y) \right|. \tag{4.75}$$

Thus, the linear-shift corrected mosaic $\hat{f}^{\mathrm{sc}}_{i,j}(x_k, y_l)$ is given by

$$\hat{f}^{\mathrm{sc}}_{i,j}(x_k, y_l) = \hat{f}_{i,j}(x_k - \kappa_{x_0}\Delta x, y_l - \kappa_{y_0}\Delta y). \tag{4.76}$$

Note that the shift in ground-range κ_{x_0} is typically zero since the linear phase error only yields a shift in cross-range direction. Moreover, only shifts with pixel precision are considered in (4.76). For small correlation coefficients, the estimation of the linear shift is unreliable. Consequently, only linear shifts are considered where the peak correlation value is larger than a certain threshold $\rho_{\min} = 0.5$. Otherwise, the mosaic is not corrected by the estimated shift. Since the registration process relies on the blurred image data, performance is limited by the severity of the blurring. Furthermore, it should be remarked that a linear trend in the motion error $\Delta^{\mathrm{s}}(p)$ may still not be estimated by the algorithm, but rather a residual linear phase error in the phase estimate of the PGA. The effect of the proposed modifications is illustrated in Figure 4.28. Comparing the focusing results of Figure 4.28a with Figure 4.25, and Figure 4.28b with Figure 4.26, significant improvements are visible in the focused images, which demonstrate that the drawbacks of the mPGA algorithm are successfully eliminated by the introduced modifications.

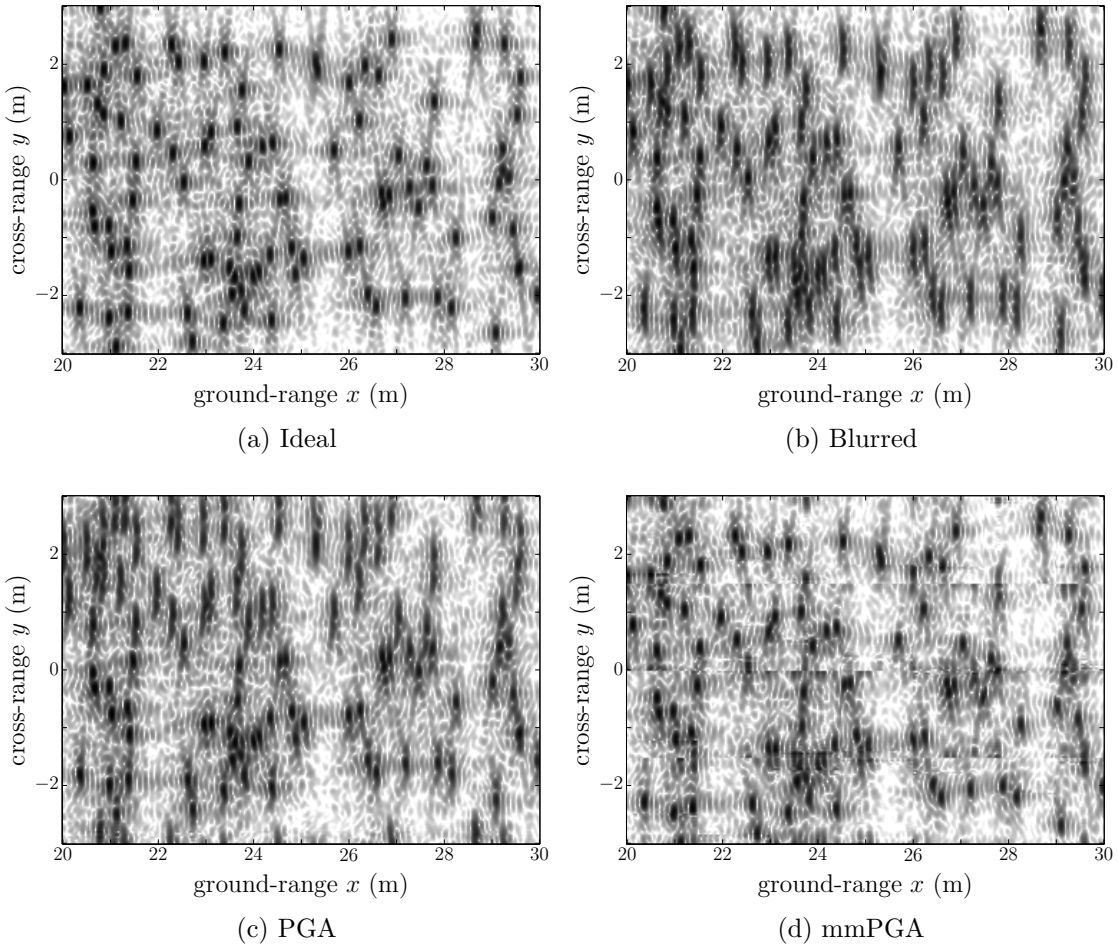


(a) Processed by mmPGA – Example 1 (b) Processed by mmPGA – Example 2

Figure 4.28: Focusing results of a synthetic image data example using the proposed mmPGA algorithm. The focused SAS images in (a) and (b) are the counterparts to the SAS images depicted in Figure 4.25b and Figure 4.26b, respectively.

4.3.5 Experiments

The mmPGA method has been tested using synthetic data generated from $D = 100$ point targets, which have been placed uniformly at random into a target area. The size of the target area is specified by $Y_0 = 3$ m, $X_0 = 5$ m and $X_c = 25$ m. A sinusoidal path deviation function has then been used to deteriorate the reconstruction process, where the motion amplitude has been set to $A^s = 0.1\lambda_c$. Here, λ_c is the wavelength at the carrier frequency. Furthermore, a cycle per synthetic aperture length frequency of $f_\mathrm{p} = 2$ has been chosen, and a total number of $N_\mathrm{MC} = 100$ experiments has been conducted. Figure 4.29a and Figure 4.29b illustrate the reconstruction results of the ideal SAS image and the blurred SAS image, respectively. In addition, the SAS images after applying the PGA and the proposed mmPGA autofocus technique are depicted in Figure 4.29c and Figure 4.29d, respectively. All SAS images have a



(a) Ideal (b) Blurred

(c) PGA (d) mmPGA

Figure 4.29: Illustration of an autofocus example that shows the ideal reconstruction of the target scene (a), the blurred image (b), and the SAS images after applying the PGA (c) and the mmPGA (d) algorithms.

dynamic range of 40 dB. The mPGA autofocusing result is not illustrated due to its poor performance given the issues addressed in Section 4.3.3. While the PGA focuses some of the scatterers in the lower region of the SAS image and, simultaneously, shifts the entire scene in positive cross-range direction, the mmPGA is capable of focusing the entire scene at the correct cross-range locations. However, it introduces some block artifacts at the borders of the mosaic pieces that are especially noticeable at the cross-range coordinate $y = 0$.

In order to evaluate the simulation results, the structural similarity (SSIM) index [Wang et al. (2004)] has been applied as a full-reference image quality measure. Here, full-reference means that the measure assesses a potentially degraded image with respect to its high-quality reference image, *i.e.*, the perfectly focused SAS image. The SSIM aims at taking the human visual perception into account by measuring and comparing the luminance, contrast, and structural similarities between the reference high-quality image and the image under test. Therefore, it provides a better image quality assessment than, for example, the mean-square error, which is typically applied to obtain a quantitative measure for the difference of two signals. The SSIM is defined as

$$\Xi(\boldsymbol{F}_1, \boldsymbol{F}_2) = L_{\mathrm{u}}(\boldsymbol{F}_1, \boldsymbol{F}_2)\, C_{\mathrm{o}}(\boldsymbol{F}_1, \boldsymbol{F}_2)\, S_{\mathrm{t}}(\boldsymbol{F}_1, \boldsymbol{F}_2), \tag{4.77}$$

where $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ are matrices containing the data of the reference image and the image under test, respectively. The SAS image matrix $\boldsymbol{F}_i$ with $i = 1, 2$ is given by

$$\boldsymbol{F}_i = \hat{f}^{(i)}(x_k, y_l) \;\; \text{with } k = 1, \ldots, M_x, \;\; \text{and } l = 1, \ldots, M_y. \tag{4.78}$$

In (4.77), the functions $L_{\mathrm{u}}(\cdot)$, $C_{\mathrm{o}}(\cdot)$, and $S_{\mathrm{t}}(\cdot)$ describe luminance, contrast, and structure measures between the images, respectively. Each SAS image has first been registered with respect to the ideal image before applying the metric. The SSIM evaluation outcome is depicted in Figure 4.30 for $N_{\mathrm{iter}} = 10$ fixed iterations of each autofocus technique. The initial iteration $q_{\mathrm{it}} = 0$ provides the SSIM value of the blurred SAS image. Therefore, it is identical in all three cases. The curves show the failure of PGA and mPGA, which both degrade the blurred image significantly during the first iteration and then rapidly converge to a low SSIM value. Contrarily, the mmPGA improves the image quality in the first iteration and then converges, which is important for finding reliable and adaptive stopping criteria. In summary, the simulation proves the validity of the proposed method using synthetic data. In the sequel, the mmPGA autofocusing technique is applied to a real stripmap SAS image.

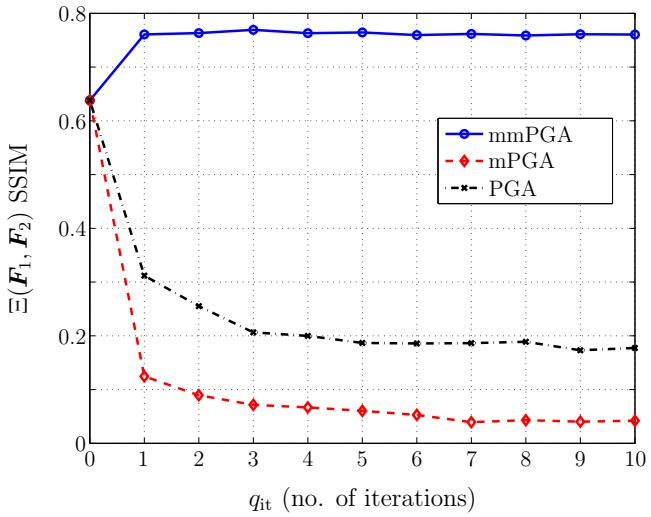
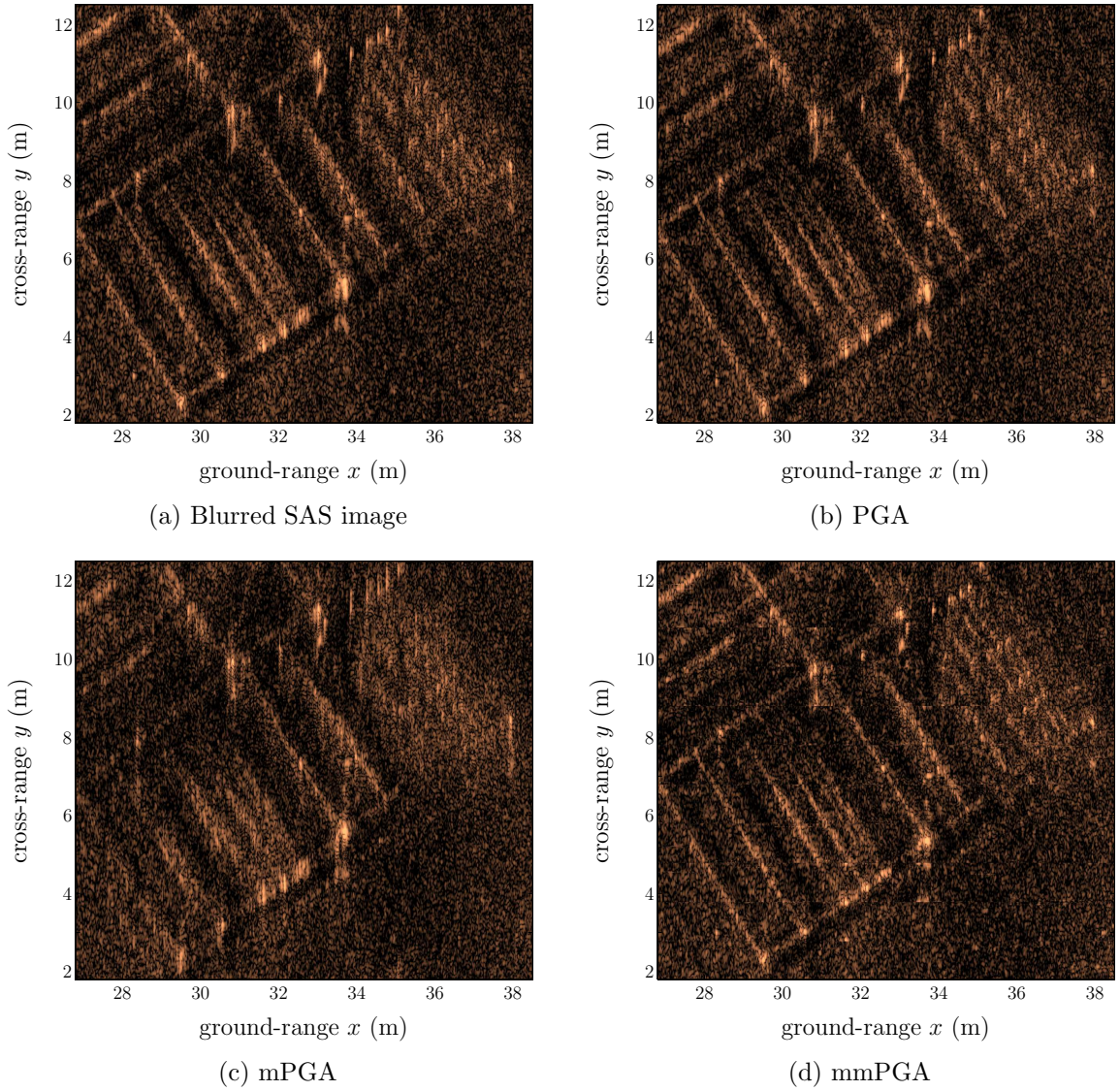


Figure 4.30: SSIM evaluation of different autofocus techniques (PGA, mPGA, and mmPGA) for $N_{\mathrm{iter}} = 10$ iterations. The initial iteration $q_{\mathrm{it}} = 0$ provides the SSIM value of the blurred SAS image.

4.3.6 Real data results

After validating the mmPGA method based on synthetic data simulations, a proof of concept on a real SAS image is shown in Figure 4.31, which illustrates the man-made target of Figure 3.2b. In order to highlight the autofocusing results, only a part of the object is shown. First of all, the proposed SAS processing chain is applied including the introduced micronavigation approach to estimate the true motion of the imaging platform based on the raw sonar measurements – see Section 4.2, which would yield a high-quality SAS image. However, an artificial sinusoidal sway motion is added to these motion estimates to blur the SAS image in a controlled manner. For the blurred SAS image depicted in Figure 4.31a, the error amplitude is set to $A^s = 2\lambda_c$ and the cycle per synthetic aperture length is chosen as $f_{\mathrm{p}} = 1$. The autofocus results for a fixed number of $N_{\mathrm{iter}} = 10$ iterations are depicted in Figure 4.31b to Figure 4.31d for the PGA, mPGA, and mmPGA algorithms, respectively. The PGA outcome neither improves nor degrades the image quality notably in contrast to the mPGA. For the mPGA, the quality degradation becomes apparent in the top-right region of the resolution target, where the strings and plastic balls are no longer identifiable. By contrast, the proposed mmPGA is capable of focusing especially all strongly reflecting target spots. This leads to a visible improvement of the overall image quality. However, a SSIM calculation yields nearly identical values for all four SAS images in Figure 4.31b to Figure 4.31d. For this reason, an evaluation of the algorithm's capabilities in a wide range of target scenes should be analyzed in future research work.



(a) Blurred SAS image

(b) PGA

(c) mPGA

(d) mmPGA

Figure 4.31: Proof of concept of the proposed autofocus technique for real SAS images.

4.4 Sound speed estimation

Precise knowledge of the sound speed is essential for high-quality SAS imaging to enable a coherent processing of multiple pings [Hansen et al. (2007)]. Otherwise, quadratic phase errors are induced along the synthetic aperture [Cook and Brown (2009)], which may cause a severe distortion of image quality – see also Section 4.1.1. Due to the inherent principle of dynamic range adjustment of the synthetic aperture length, the effect of sound speed errors becomes range-variant. Furthermore, widebeam SAS systems strongly amplify the distortion effect due to an extension of the synthetic aperture length with respect to narrowbeam SAS systems [Hansen et al. (2007)]. The

inhomogeneity of water additionally causes the sound speed to change with depth and range [Lurton (2002)]. However, measurements of the sound speed are only locally available, *e.g.*, using a sound speed sensor on-board of an AUV. Therefore, the combination of range sensitivity and measuring inaccuracy of the sound speed poses a major challenge to any long-distance SAS system. Another aspect concerns motion compensation. As the introduced micronavigation technique estimates translational ping-to-ping motion by scaling time delay differences with sound speed in order to fit a range difference model, its capability of compensating sound speed errors is limited.

In this section, the influence of sound speed variations is analyzed using both synthetic and real sonar measurements. The latter are collected by the VISION SAS system – see Section 3.1. To this end, different quality metrics are employed to measure the sound speed impact on SAS images. Based on this quality analysis, an estimation technique is proposed to determine an optimal average sound speed with respect to the behavior of quality metrics. Two processing schemes are available. While the first one varies the sound speed inside micronavigation and image reconstruction, the second processing scheme only varies the sound speed for imaging. A metric comparison between both schemes points out that micronavigation is incapable of compensating sound speed errors. The material of this section has been published in [Leier et al. (2013a)].

4.4.1 Image quality metrics

In order to asses the quality of SAS images, two groups of image quality metrics are introduced. The first group is based on the PSF of isolated scatterers. It encompasses metrics such as the 3 dB mainlobe width and the PSF peak value, which is linked to the SAS gain [Bellettini and Pinto (2002), Groen et al. (2012)]. Other common figures of merit, *e.g.*, the sidelobe level (SLL) and the integrated sidelobe level (ISLL) are not considered here. However, they also fall into this group and could be employed. All PSF based metrics have in common that they require both the presence of isolated point targets as well as an automatic selection process in practice. An automatic selection process of point targets is suggested in [Glover and Campell (2010)]. The procedure can be summarized as follows: (i) it automatically searches for the location of the strongest intensity in a SAS image, (ii) cuts out a 2-D image patch around the location, and (iii) extracts the PSF properties. As this automatic selection process may include extended targets rather than isolated point targets, it is repeated a predefined number of times, always excluding all previous image patches in the next iteration. For each iteration, the desired PSF characteristics are determined. In order to eliminate the influence of non point-like targets and to obtain a single figure of merit

for each PSF characteristic, a robust averaging, *e.g.*, the median, is performed. For the demonstration of the main idea of the proposed sound speed estimation technique, however, point target candidates are selected manually while their corresponding PSFs are extracted automatically.

Provided that an image patch is available and contains an isolated point target, a parabolic function

$$h(y) = \beta_1 + \beta_2\, y + \beta_3\, y^2 \tag{4.79}$$

is fitted to the sinc-like PSF peak in along-track direction [Moddemeijer (1991), Bonifant (1999)]. The model parameters β_j, with $j = 1, \ldots, 3$, are then estimated based on three sample points of the PSF at along-track locations y_{l-1}, y_l, and y_{l+1}, where the maximum sampled peak value is located at y_l. Thereupon, the 3 dB mainlobe width of the PSF can be expressed in terms of the estimated parameters as follows

$$\hat{\Upsilon}_{\mathrm{3dB}} = 2\sqrt{\left(\frac{\hat{\beta}_2}{2\hat{\beta}_3}\right)^2 - \frac{\hat{\beta}_1 + 3}{\hat{\beta}_3}}. \tag{4.80}$$

Next, the peak value of the PSF that is related to the SAS gain can be estimated by

$$\hat{\Upsilon}_{\max} = h(y_{\max}),\ \text{with}\ y_{\max} = -\frac{\hat{\beta}_2}{2\hat{\beta}_3}. \tag{4.81}$$

Contrarily, the second group of quality metrics can be directly applied to the entire image, which is advantageous since the metric evaluation does not rely on the presence, selection, and extraction of isolated point targets. The group consists of metrics such as contrast measures [Fortune et al. (2001)] or texture measures [Haralick et al. (1973)]. Here, a common measure for the contrast is considered that is defined by the ratio between the sample standard deviation, $\hat{\sigma}_f$, and sample mean, $\hat{\mu}_f$, of the vectorized image matrix $\hat{\boldsymbol{f}} = \mathrm{vec}\{\hat{\boldsymbol{F}}\}$, with $\hat{\boldsymbol{F}} = \hat{f}(x_k, y_l)$ for $k = 1, \ldots, M_x$ and $l = 1, \ldots, M_y$, as

$$\hat{\Upsilon}_{\mathrm{co}} = \frac{\hat{\sigma}_f}{\hat{\mu}_f}, \tag{4.82}$$

where $\mathrm{vec}\{\cdot\}$ denotes the operator to vectorize a matrix column after column. As both groups have their assets and drawbacks, it is best practice to apply all metrics for the sound speed estimation technique and average their individual outcomes since the calculation is computationally inexpensive.

4.4.2 Sound speed estimation technique

Provided that metrics have been applied to measure image quality as a function of the current sound speed value c, an optimum average sound speed $\hat{c}_0$ can be estimated

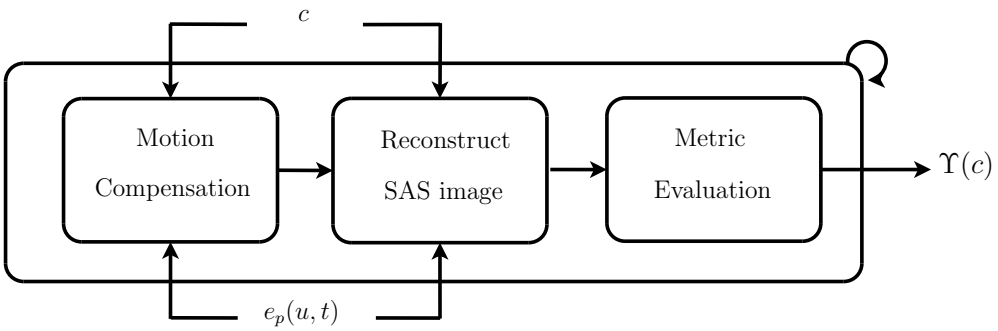


Figure 4.32: Block diagram of sound speed estimation technique.

that yields the best focused SAS image. To this end, a bounded interval of sound speed values $\mathcal{I}_c = [c_{\min}, c_{\max}]$ is chosen. Its midpoint is typically determined by the locally measured sound speed and the interval radius is given by a maximum sound speed variation of $q_\epsilon = 2\%$ [Hansen et al. (2011)]. An image patch $\hat{f}(x_k, y_l; c)$ with $k = 1, \ldots, M'_x$ and $l = 1, \ldots, M'_y$ is then reconstructed for each value of the interval $\mathcal{I}_c$, given a pre-determined step size to cover the interval length. The sound speed dependency of the SAS image patch is highlighted by the notation $\hat{f}(x_k, y_l; c)$. Thereupon, a quality metric $\Upsilon(c)$ as a function of sound speed is estimated by

$$\hat{\Upsilon}(c) = \mathcal{Q}\left\{\hat{f}(x_k, y_l; c)\right\}, \ \forall\, c \in \mathcal{I}_c, \tag{4.83}$$

where $\mathcal{Q}\{\cdot\}$ denotes an estimator or mapping operator of the corresponding quality metric. Based on the metric function $\hat{\Upsilon}(c)$, an optimal average sound speed estimate $\hat{c}_0$ is obtained as follows

$$\hat{c}_0 = \begin{cases} \arg\max\limits_c \hat{\Upsilon}(c), & \text{for } \frac{\mathrm{d}^2}{\mathrm{d}c^2}\Upsilon(c) < 0 \ \forall\, c, \\ \arg\min\limits_c \hat{\Upsilon}(c), & \text{for } \frac{\mathrm{d}^2}{\mathrm{d}c^2}\Upsilon(c) > 0 \ \forall\, c. \end{cases} \tag{4.84}$$

The procedure is summarized in the block diagram of Figure 4.32. It is highlighted that both major processing blocks, namely, motion compensation and SAS image reconstruction, require a correct sound speed leading to two options for the processing scheme: (i) a variation of sound speed inside image reconstruction and micronavigation, (ii) a variation of the sound speed only during image reconstruction. For the second option, the locally measured sound speed is used for motion compensation. Thus, a comparison of both processing options provides an indication whether micronavigation is able to compensate properly for a sound speed mismatch. First of all, a metric evaluation for a sound speed variation in the interval $\mathcal{I}_c = [1460, 1510]$ m/s is shown in Figure 4.33 for synthetic data with a true sound speed of $c_0 = 1485$ m/s. Here, the metric curves are normalized except for $\hat{\Upsilon}_{\mathrm{3dB}}(c)$ so as to display them in a single plot. While the metric curves of the contrast $\hat{\Upsilon}_{\mathrm{co}}(c)$ and SAS gain $\hat{\Upsilon}_{\max}(c)$ feature

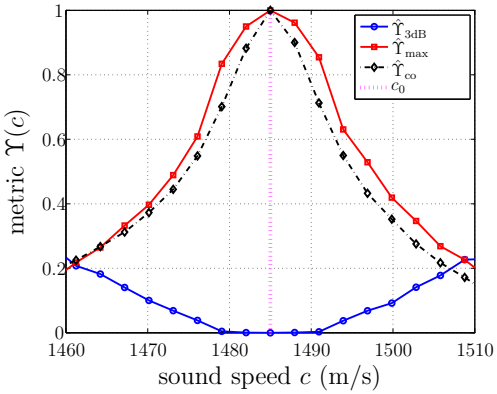
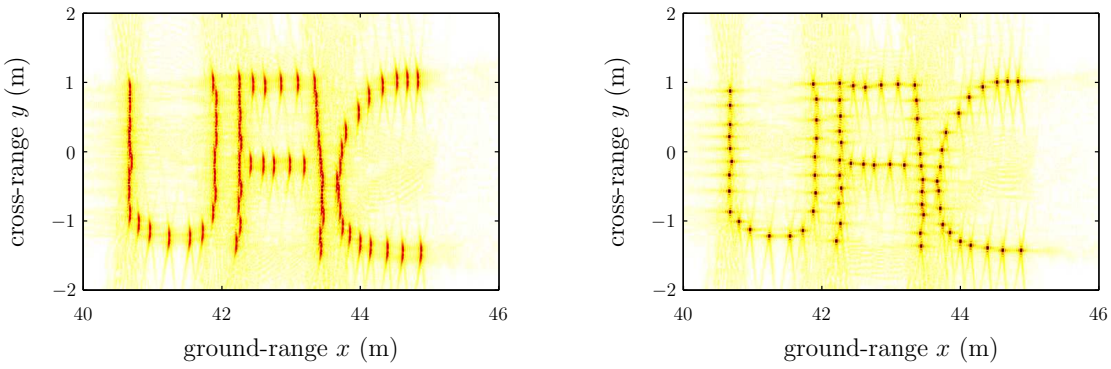


Figure 4.33: Metric evaluation for varying sound speed values inside the SAS image reconstruction process using synthetic data with a true sound speed $c_0 = 1485$ m/s.

a distinct extreme value at the true sound speed c_0 in Figure 4.33, the curve of the half-power mainlobe width $\hat{\Upsilon}_{3dB}(c)$ is rather flat in the vicinity of c_0. It is best practice to combine the introduced metrics for a reliable estimation result. In order to obtain the PSF based metrics in this example, $D = 5$ point targets are selected manually. The individual PSF curves are then averaged.

Figure 4.34b illustrates the focused SAS image of the underlying target scene, which has been used in the metric evaluation example above. The system parameters for generating the synthetic phase history data match the settings of the VISION SAS system. The dynamic range is set to 40 dB and no synthetic aperture shading is applied. While an error of $\Delta c = 17$ m/s is induced during the reconstruction process of the SAS image depicted in Figure 4.34a, the SAS image in Figure 4.34b



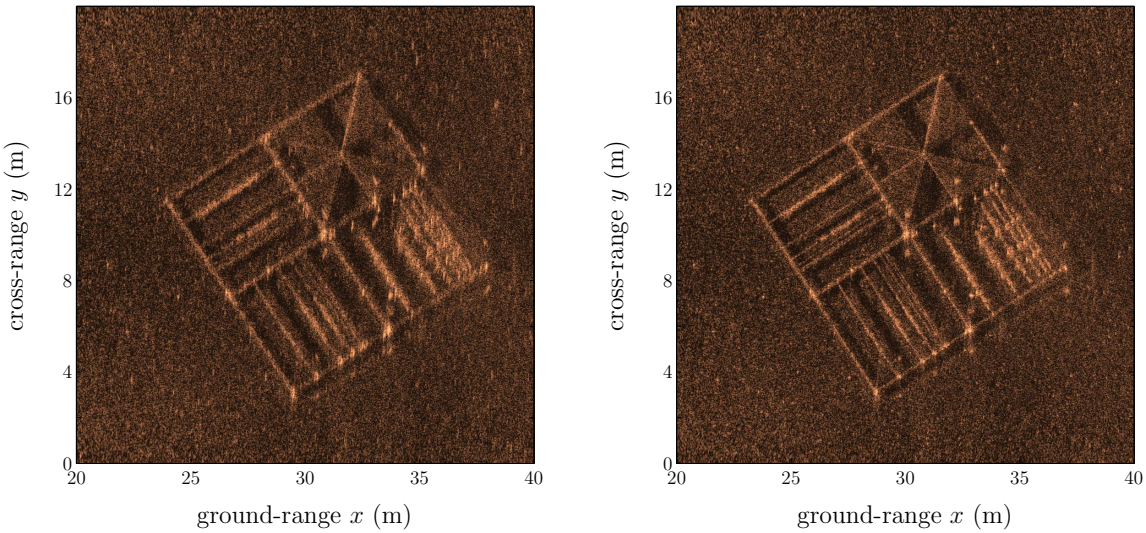
(a) $\tilde{c} = 1502$ m/s (b) $c_0 = 1485$ m/s

Figure 4.34: Illustration of sound speed influence on SAS images using synthetic data. While the image in (a) is defocused due to an incorrect sound speed measurement $\tilde{c}$, using the correct sound speed c_0 yields a focused SAS image in (b).

is reconstructed assuming the correct sound speed c_0. Clearly, the defocusing becomes apparent by elongated PSFs in cross-range direction of Figure 4.34a. Although the error is intentionally chosen to be large to illustrate the blurring effect, the value is still within the limit of feasible sound speed variations in the ocean [Hansen et al. (2011)].

4.4.3 Real SAS image results

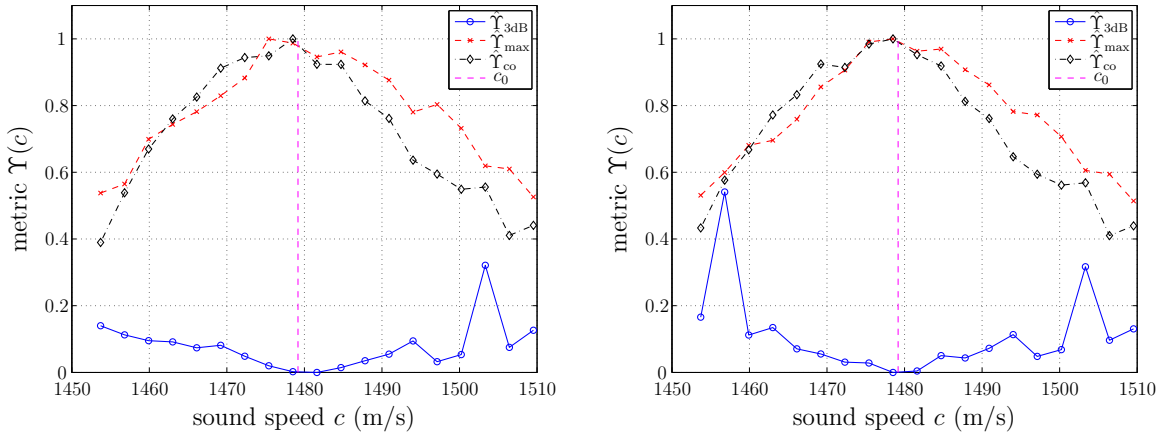
In the sequel, the introduced metrics are applied to a set of SAS images, which are reconstructed by varying the sound speed inside the backprojection algorithm. As an initial value, the measured sound speed of the on-board device is used and varied within the bounds of ±2% [Hansen et al. (2011)]. Figure 4.35a and Figure 4.35b illustrate SAS image examples for an erroneous sound speed of $\tilde{c} = 1500$ m/s and an estimated sound speed of $\hat{c}_0 = 1479$ m/s, respectively, where the latter is obtained by the proposed technique. Both SAS images have a dynamic range of 40 dB and show the resolution target of Figure 3.2b. In order to be consistent with the synthetic data example, aperture shading is not applied. The comparison of Figure 4.35a and Figure 4.35b highlights the influence of sound speed on image quality. While the erroneous sound speed $\tilde{c}$ yields a defocus, the optimum average estimate $\hat{c}_0$ clearly improves the quality of the SAS image. For example, the parallel strings with the attached plastic balls can be better distinguished in the right hand section of the object depicted in Figure 4.35b. In general, all line objects are slightly better focused in Figure 4.35b compared to



(a) Blurred SAS image $\tilde{c} = 1500$ m/s

(b) Focused SAS image $\hat{c}_0 = 1479$ m/s

Figure 4.35: Illustration of SAS image quality enhancement using the proposed sound speed estimation technique for real measurements of the VISION SAS system.



(a) Scheme I: Sound speed is varied during motion compensation and imaging.

(b) Scheme II: Sound speed is only varied for image reconstruction.

Figure 4.36: Metric comparison of both processing schemes: Subplot (a) includes the sound speed variation for motion estimation. Subplot (b) uses the measured value.

Figure 4.35a, where they appear to be more smeared out. Moreover, the erroneous sound speed leads to a spreading of the PSF of the individual scatterers, which is especially observable in the top right corner of the SAS image and for the strongly reflecting target spots on the object itself. Although the effect is marginal, an overall contrast loss occurs in the contours of shadow areas. The latter is observable in the upper part of the object, which is akin to rotor blades of a windmill.

Similar to the synthetic data, the optimum average sound speed is estimated using the proposed approach. The corresponding metric curves are shown in Figure 4.36. Although the metric values are noisier compared to the synthetic data results, all curves have their extreme value close to a sound speed value of $c_0 = 1479$ m/s, which has been measured by the on-board sound speed sensor of the AUV. It is used as a reference value to additionally assess the proposed technique. In order to judge the capability of micronavigation to compensate a sound speed error, two different evaluation schemes are considered. In the first scheme, the sound speed is varied in both processing steps, *i.e.*, motion compensation and SAS image reconstruction. In the second scheme, the sound speed is only varied during image reconstruction while keeping the measured value c_0 fixed to estimate the true trajectory of the AUV – see Section 4.2.6. The metric evaluation of the second scheme is considered as a reference to assess the micronavigation. In case micronavigation is able to compensate the variations in the sound speed, the metric curves are expected to differ significantly. Moreover, the metric curves should be constant over a large interval of different sound speed values, hereby, indicating that a varying sound speed is compensated instead of affecting the image quality by decreasing the metric value. A comparison of the metric characteristics for

both evaluation schemes is depicted in Figure 4.36a and 4.36b. While Figure 4.36a depicts the metric outcome for the first scheme, Figure 4.36b illustrates the result for varying the sound speed only during image reconstruction. A comparison reveals that both plots differ marginally but that the general trend is identical. Furthermore, the different metric functions have almost the same location in their extreme values for both schemes. This indicates that they all estimate the same optimum average sound speed $\hat{c}_0$. As the metric behavior in Figure 4.36a is not flat over an interval of sound speed values in the vicinity of the optimum value, the micronavigation technique is considered to be unable to correct sound speed mismatches.

4.5 Conclusions

In this chapter a data-driven motion compensation technique, also known as micronavigation, has been introduced, which estimates unknown platform motion to reconstruct high-resolution and well-focused SAS images. After reviewing the influence of undesirable platform motion and error-phase terms that may arise from inaccurate sound speed measurements, a thorough mathematical description of the signal processing steps for motion estimation has been presented. Amongst others, the processing steps mainly involve surge and time delay estimation as well as nonlinear least squares fitting. A compensation method based on binary image processing techniques has been proposed to correct biased time delays, which may occur due to a high ratio of carrier frequency to bandwidth. Superior performance compared to median filtering has been demonstrated on real sonar measurements.

Furthermore, it has been theoretically proven that biased ping-to-ping motion estimates occur if time delay estimation is directly performed on raw echo measurements. Although the presented nonlinear least squares approach has implicitly accounted for these biased estimates in the case of lateral ground-range sway motion, the biased time delays may still lead to an erroneous calculation of pitch and heave. In order to compensate for this effect, a near-field and widebeam beamforming approach has been proposed. The developed processing chain for motion compensation has been validated on synthetic and real sonar measurements, which have been collected by two different SAS systems in various sea trials. SAS images have been provided to highlight the quality enhancement using data-driven motion estimates of the true platform trajectory. The effect of bathymetry variation has then been studied on synthetic data. It has been demonstrated that bathymetry estimates are necessary to properly estimate ground-range sway and heave ping-to-ping motion. Additionally, height grids for space-

time ground-range imaging have to be provided along with correct motion estimates to obtain high-quality SAS imagery in environments with a challenging bathymetry.

Besides micronavigation, autofocus techniques for stripmap synthetic aperture imaging have been considered in this chapter. In this context, two modifications to enhance an existing autofocus technique have been introduced: The first extension accounts for varying motion errors along the synthetic aperture by partitioning the entire SAS image into 2-D mosaic pieces. The second modification is a co-registration process. It overcomes linear shift problems during the reassembly of the individual mosaic pieces. An improvement of the modified autofocus technique has been validated using synthetic data simulations. Further, the method has been exemplarily tested on real SAS data, where it has also enhanced image quality. However, further verification of the technique for additional scenarios is required and remains the subject of future work. Moreover, the proposed autofocus method only works for residual motion errors, *i.e.*, it has not been possible to focus SAS images that have been reconstructed with imperfect motion provided by the INS unit. A possible reason for this is the dependency on the required similarity of mosaic pieces during co-registration.

Another major issue that affects SAS image quality is inaccurate knowledge on the sound-speed profile. Thus, an estimation scheme has been suggested based on assessing the SAS image quality via appropriate metrics. The technique is able to determine an optimum average sound speed value with respect to the employed image quality metric. This has been demonstrated for synthetic data as well as real sonar measurements. For both, the estimated sound speed either coincides with the preset sound speed in the case of synthetic data or with the measured sound speed provided by an AUV on-board device. The method has proven to be a potential software alternative to hardware measurement devices that can be employed as an online calibration technique.

Chapter 5

Motion influence on target recognition

The objectives of a fully automatic detection and classification (ADAC) system for mine countermeasure applications consist of detecting mines and classifying them according to known mine types [Siantidis and Hölscher-Höbing (2009), Fandos (2012)]. A common approach to achieve this objective is to apply image processing techniques to sonar imagery of the seafloor, *e.g.*, segmentation, to detect candidate objects. In the following, each object is characterized by a set of features, and the object is classified accordingly [Doherty et al. (1989), Dobeck et al. (1997), Aridgides et al. (2001), Fawcett (2001), Ciany and Zurawski (2002), Perry and Guan (2004)]. Ideally, an autonomous underwater vehicle (AUV) would reconstruct images of the seafloor *in situ* and process them directly, *e.g.*, detecting and identifying mines types, followed by a transmission of the extracted and relevant information to an operation base to reduce the reporting time during missions. This automatic procedure assumes a reliable ADAC system, which is able to function under varying image quality conditions or difficult scenarios, *e.g.*, shallow-water, severe ocean current situations or rough topography.

State-of-the-art ADAC processing uses synthetic aperture sonar (SAS) systems [Groen et al. (2009), Coiras and Groen (2009), Midelfart et al. (2009)] as they are superior to regular side-looking sonar systems in terms of resolution and coverage rates. Consequently, these systems facilitate higher classification performance due to a gain in the richness of detail. Typically, micronavigation techniques including bathymetry estimation together with autofocus and other improvement approaches, *e.g.*, sound speed estimation, provide sufficient accuracy to reconstruct high-quality SAS images as discussed in Chapter 4. However, in difficult environmental situations, it cannot be guaranteed that residual motion errors or sound speed mismatches are still present leading to a degradation in the quality of SAS images. Consequently, the image quality may degrade and, simultaneously, the segmentation result and hence the feature extraction process of the ADAC system are affected. For example, geometrical information about the shape of a shadow associated with an object changes, and thus, the classification performance deteriorates. Therefore, it is of major interest to study the detection and classification behavior of an automatic system under the influence of residual motion and phase errors to develop highly reliable target recognition systems in the future. In order to study the influence of motion errors on automatic target recognition (ATR), the ADAC system described

in [Fandos (2012), Fandos et al. (2013)] is trained with a database of well focused SAS images. Subsequently, real raw echo measurements are used from different sea trials to build a test database of SAS images where motion errors and sound speed mismatches are artificially induced into the image reconstruction process to study the impact on segmentation, feature extraction and classification performance. Simultaneously, a relation between the image degradation and the individual tasks of the ADAC system is empirically demonstrated by assessing the image quality using a full-reference method. The obtained results illustrate a severe dependency of the ADAC system performance on SAS image quality. This highlights the importance for both, image quality assessment schemes and robust segmentation and feature selection techniques, to improve the reliability of SAS based target recognition systems under difficult conditions.

The material of this chapter has been published in [Leier et al. (2014)]. It represents joint work from a collaboration with a former colleague, Dr.-Ing. Raquel Fandos, of the Signal Processing Group, Technische Universität Darmstadt, Germany, whose research interest was mainly in ADAC system design for mine hunting applications [Fandos and Zoubir (2011), Fandos et al. (2013a), Fandos et al. (2013)]. The SAS processing chain as introduced in Chapter 3 and Chapter 4 has been employed to build a database of SAS images to test the ADAC system proposed in [Fandos (2012), Fandos et al. (2013a)] with respect to motion errors. Therefore, each database entry consists of a set of SAS images of the same scene with varying image quality, ranging from totally focused to severely blurred. The analysis and the evaluation of the results have been performed in close collaboration.

5.1 Automatic detection and classification

The ADAC system for underwater mine hunting as proposed in [Fandos (2012)] consists of three major steps as depicted in Figure 5.1. By applying a segmentation method to the image of interest, it is partitioned into three regions: the highlight of the objects, their shadow and the background. Besides man-made objects, physical features of the terrain such as rocks and sand ripples may also be segmented. A set of features that contain information for classification is then extracted for each segmented object. Finally, the classification step assigns a class to each object according to the comparison of its feature vector and those of a training set. Ideally, the physical features of the terrain are classified as clutter and the man-made objects as mines. If not, they constitute false alarms and missed detections, respectively. In the following, the different stages of the system in Figure 5.1 are briefly described.

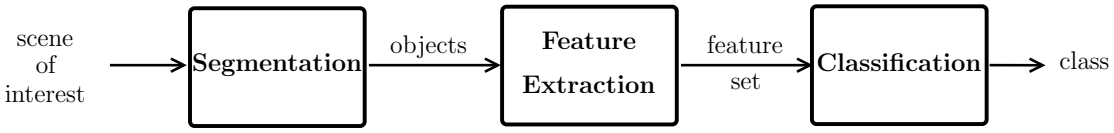


Figure 5.1: General scheme of an ADAC system for mine hunting. The image is first segmented into background, highlight and shadow regions. For each object (a shadow and its associated highlight), a set of features is then extracted. The classifier assigns a class to each object by comparison with a training data set, after [Fandos (2012)].

5.1.1 Segmentation

Segmentation is a processing technique that subdivides an image into a predefined number of regions by assigning a label to each pixel. All pixels that share the same label are then associated with the same region or segment. Typically, such a connected label representation is a first processing step for image analysis since similar regions share specific characteristics, which can be described by common features. In mine hunting and object detection applications, the segmentation process aims at identifying three regions of interest in a sonar image, namely, object, shadow, and background regions – see also [Fandos and Zoubir (2011)]. In the following, a brief description of the two employed segmentation algorithms within the ADAC processing chain is provided. For more details, the reader is referred to [Fandos (2012)].

- Iterative conditional modes (ICM)
  This method relies on a Markov random field (MRF) model of the image [Besag (1986)]. Besides the statistical properties of the pixel intensity, MRF takes into account the neighborhood relations among pixels, which are modeled by the so-called *a priori* Markovian probability. ICM for sidescan sonar image segmentation were introduced in [Collet et al. (1996), Thourel et al. (1996)] and later used in [Mignotte et al. (2000), Reed et al. (2003)], where they were combined with the iterative conditional estimates (ICE) [Salzenstein (1997)] to achieve an unsupervised implementation. The method was tested in [Fandos and Zoubir (2011)] with SAS images.

- Graph cut (GC)
  The segmentation is performed by adopting a graph representation of the image of interest [Ford and Fulkerson (1962)], where each pixel is modeled as a node and the pixel neighborhood relations are represented by edges. This algorithm also assumes a MRF model of the image. By assigning weights to the edges, both the pixel intensity and neighborhood relations are modeled.

An efficient implementation of a graph cut algorithm (min-cut/max-flow) [Boykov and Kolmogorov (2004)] is then applied to divide the graph into two groups of nodes, which correspond to the shadow and background regions. It was proposed for SAS image segmentation by [Fandos et al. (2011)].

Figure 5.2 includes two snapshots of cylindrical and spherical man-made objects and their segmentation by the ICM and GC algorithms. The highlight regions are depicted in white, while the shadow regions are delimited by a yellow line. Since the GC method divides the image into only two regions, the ICM highlight result is also assumed for the GC segmentation. The resolution of the images is 2.5 cm × 2.5 cm per pixel.

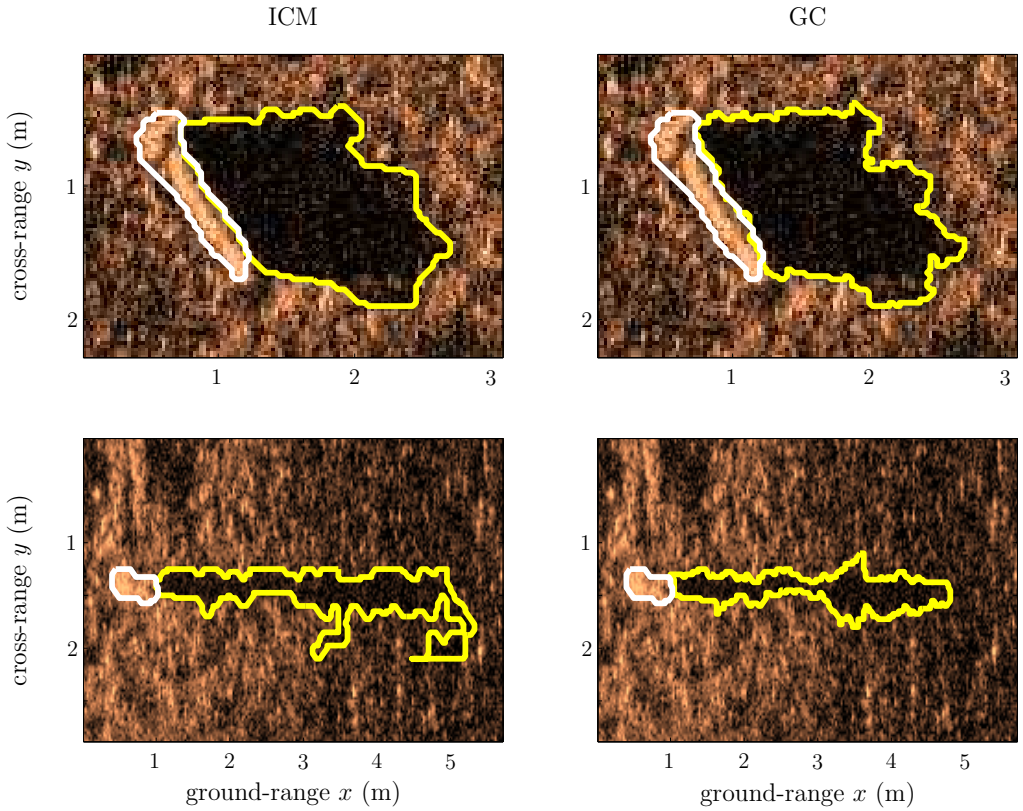


Figure 5.2: Segmentation performance of the ICM and GC algorithms for a cylindrical and spherical object. The first column shows two different snapshots of SAS images and, superimposed, the ICM segmentation results (white line for highlights and yellow line for shadows). The second column shows the GC segmentation results.

5.1.2 Feature extraction

Based on the segmentation of three image regions, different types of features or descriptors are extracted, which aim at characterizing the respective region uniquely.

Typically, traditional sidescan ADAC systems are based on descriptors of the shadow rather than the highlight. The reasons for this are twofold. On the one hand, the resolution of sidescan systems is lower, so significantly fewer pixels are available for the highlight region of sidescan images compared to SAS images. On the other hand, the orientation of the object of interest with respect to the sonar antenna might be such that the intensity of the returned echo is too weak for the object highlight to be accurately reconstructed. This aspect-dependent effect is far more remarkable in sidescan than in SAS imagery.

However, for SAS based ADAC system, features of all three regions are taken into account. The feature set as introduced in [Fandos et al. (2013)] is also adopted in this study. It consists of a combination of statistical and geometrical features for both the highlight and shadow regions. In total, the feature set $\boldsymbol{\varphi}$ consists of $N_\varphi = 204$ descriptors, which can be merged into seven groups:

- Statistical features ($\boldsymbol{\varphi}_{\text{stat}}$)
- Geometrical features of the shadow ($\boldsymbol{\varphi}_{\text{geo}}$)
- Geometrical features of the highlight ($\boldsymbol{\varphi}_{\text{obj}}$)
- 2D-Fourier coefficients ($\boldsymbol{\varphi}_{\text{Fourier}}$)
- Principal components of the segmented shadow ($\boldsymbol{\varphi}_{\text{PCA}}$) [Pearson (1901)]
- Normalized central moments ($\boldsymbol{\varphi}_{\mu}$) [Gonzalez and Woods (2002)]
- Invariant moments ($\boldsymbol{\varphi}_{\text{inv}}$) [Gonzalez and Woods (2002)]

Note that, as described in Section 5.1.3, a feature selection algorithm is applied to select the best subset $\boldsymbol{\varphi}'$ for a given training database.

5.1.3 Classification

The linear discriminant analysis (LDA) has been selected as classification system. This choice is motivated by the experiments presented in [Fandos et al. (2013)] that uses a resampling method [Fandos et al. (2013a)] to compare several classifier candidates for a database of SAS images, *e.g.*, k-Nearest Neighbor, LDA, Mahalanobis' classifier and Support Vector Machines. The LDA classifier shows better performance than the other

methods. The comparison of classification systems accomplished by the resampling method is independent of any pre-selected feature subset.

Due to the curse of dimensionality [Hughes (1968)], classification performance typically further improves by selecting a subset of features $\boldsymbol{\varphi}' \in \mathbb{R}^{N'_\varphi}$ with $N'_\varphi < N_\varphi$. In order to do that, the D_{SFS}-sequential forward selection (D_{SFS}-SFS) algorithm [Fandos et al. (2013a)] is applied with $D_{\mathrm{SFS}} = 3$. This algorithm is computationally more efficient and provides better results than the sequential forward floating selection (SFFS) method [Pudil (1994)]. For the problem at hand, the size of the optimal feature subset is $N'_\varphi = 33$ if the ICM segmentation method is applied and $N'_\varphi = 41$ for the GC algorithm. Note that these optimal feature subsets have been estimated for the set of images without motion errors. This means that the selected features optimize the classification results without taking into consideration the degradation, which potential motion errors would have. In the following section, the influence of the motion errors on segmentation, feature extraction, and classification results but not on feature selection is addressed. The selection of a set of features that is invariant to motion errors is not considered and remains as future work.

5.2 Empirical study

After motion estimation and compensation during imaging, ideally, a well focused and high-resolution SAS image is obtained. In order to study systematically how unknown platform motion and phase errors affect image segmentation, feature extraction, and classification performance, artificial errors are induced to degrade image quality in a controlled manner. In this empirical analysis, two types of errors are considered, namely, sinusoidal path deviations and quadratic phase errors due to sound speed mismatches – see Section 4.1 for details, and two databases of real SAS images are used for training and testing. The study exemplarily shows that the loss in image quality expressed in terms of the structural similarity (SSIM) [Wang et al. (2004)] is related to performance degradation of the respective ADAC tasks. Further, it is demonstrated that classification can be improved by training with motion error affected images.

5.2.1 Experimental database description

In order to analyze the effects of motion errors on an ADAC system, two databases have been considered. While the first database is used for training the classification

Table 5.1: Databases: Class 1 corresponds to clutter, while classes 2 and 3 refer to spherical and cylindrical objects, respectively.

Database	class 1	class 2	class 3
Training	3604	46	57
Test	201	16	29

system, the second database is employed as test set. Both data sets consist of clutter, cylindrical, and spherical man-made objects. The number of elements for each class is detailed in Table 5.1. The training data set has been generated with the VISION system by ATLAS UK. The second database has been built from raw echo measurements collected by ATLAS ELEKTRONIK GmbH using a SeaOtter MKII with the MCM-SLS system as specified in Section 3.1. The second database consists of in total $N_I = 32$ SAS image patches containing the object classes as specified in Table 5.1. One of the image scenes is depicted in Figure 5.3 with a dynamic range of 40 dB showing a spherical and cylindrical object. Note that the sea bottom type is almost identical for all of the $N_I = 32$ scenes. It is remarked that the subsequent analysis of motion error effects

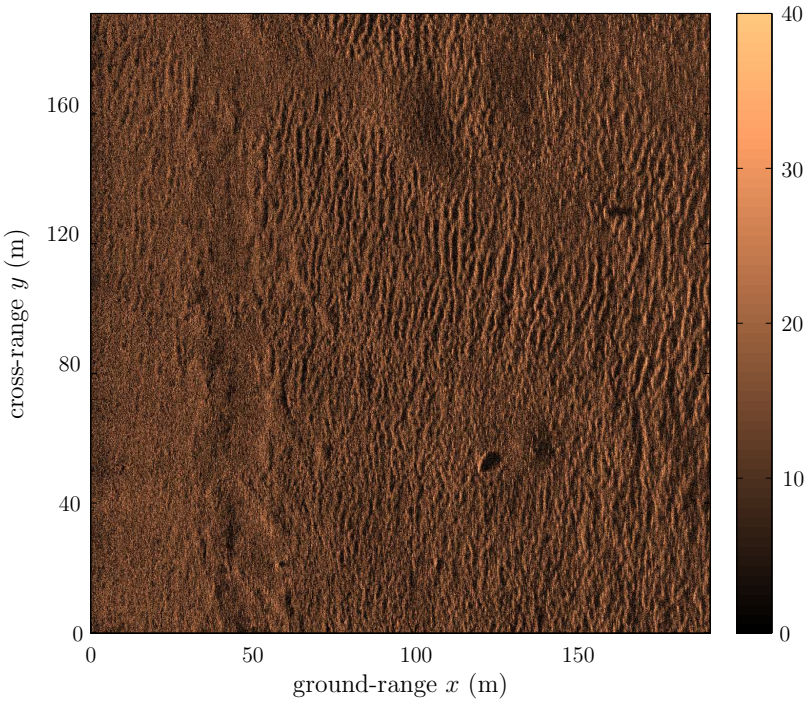


Figure 5.3: One of the $N_I = 32$ SAS image scenes used to construct the test database for the ADAC system. A dynamic range of 40 dB is displayed.

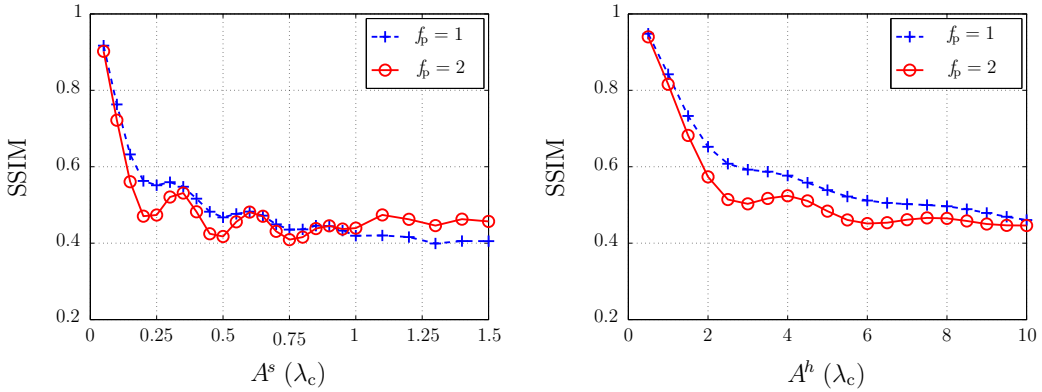
on image quality and on the individual ADAC tasks is system-dependent. Especially the angular beamwidth θ_{BW} of the sonar plays an important role since it determines the length of the synthetic aperture. In general, the relative influence of ground-range sway (called sway in the following) and heave on the image quality as well as on ADAC related tasks depends on the imaging geometry, namely, the ratio between the altitude and the imaging range of the SAS system.

5.2.2 Image quality

An assessment of the image quality of side-looking sonar and especially SAS imagery is of great importance for a reliable ADAC system. It guarantees the use of reasonable input data to the system [Williams et al. (2012), Williams (2012)]. Subsequently, an empirical relation is shown between the image quality expressed in terms of the SSIM index [Wang et al. (2004)] and segmentation performance, feature sensitivity as well as classification performance obtained by an ADAC system as depicted in Figure 5.1. Figure 5.4 illustrates an average assessment using the SSIM measure – see (4.77), and $N_I = 32$ SAS image patches for varying sinusoidal amplitudes A^{ϵ}, with $\epsilon \in \{s, h\}$ for sway and heave, in units of wavelength and for two different cycles per synthetic aperture length frequencies, *i.e.*, $f_{\mathrm{p}} = 1$ and $f_{\mathrm{p}} = 2$.

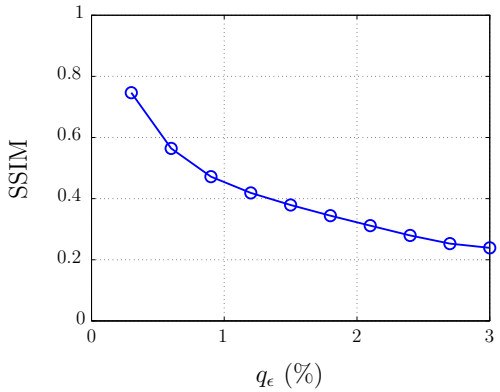
Comparing Figure 5.4a and Figure 5.4b demonstrates that sway errors degrade synthetic aperture images more severely than other type of motion errors due to the translation towards the direction of the scene of interest. While the SSIM already drops to a value of 0.7 in case of a sway error amplitude of approximately $A^s = 0.1\lambda_c$, the same drop in the SSIM value only occurs for heave error amplitudes larger than $A^h = 1.4\lambda_c$. Moreover, increasing the cycles per synthetic aperture length frequency f_{p} results into a faster SSIM drop-off, followed by a region in which the curves for both translational errors fluctuate around a constant value. The latter can be interpreted such that the image quality worsens only up to a certain error amplitude. A similar behavior is observed when investigating the impact on the segmentation and feature sensitivity results with respect to motion errors, and ultimately, for classification performance. In Figure 5.4c the average image quality assessment is depicted in the case of rising mismatches between locally measured and actual sound speed at the scene of interest. Note that a variation in the sound speed of $\Delta c = \tilde{c} - c_0 = 9$ m/s ($q_\epsilon = 0.6$ %) already yields a significant drop in the similarity index. Thus, a severe influence on subsequent ADAC tasks is to be expected for small discrepancies between actual and measured sound speed. Typically, a variation in the sound speed of up to $q_\epsilon = 2\%$ can occur along the acoustical path [Lurton (2002), Hansen et al. (2007), Hansen et al. (2011)].

In order to study the effects of sound speed variation on image quality and ADAC tasks, a variation of up to $q_\epsilon = 3\%$ is considered in the simulations.



(a) SSIM for sway error

(b) SSIM for heave error



(c) SSIM for sound speed mismatch

Figure 5.4: SSIM evalution of image quality assessment for sway (a) and heave (b) affected SAS images with increasing error amplitudes A (λ_c). Subplot (c) shows the SSIM for increasing sound speed mismatches q_ϵ.

5.2.3 Segmentation

First, based on a few example results, it is demonstrated how the segmentation of the highlight and shadow region deteriorates under the influence of motion errors as well as for a mismatch in sound speed. A quantitive study of segmentation performance is then conducted. Figure 5.5 illustrates the change in the segmentation for the ICM and GC algorithms as the amplitude of translational motion errors increases. The first row depicts the segmentation results of well focused image snapshots. The first and third columns correspond to the same snapshot as segmented by the ICM and GC algorithms,

respectively, and analogously for the second and forth columns. While the second and third rows correspond to a sway amplitude of $A^s = 0.2\lambda_c$ and $A^s = \lambda_c$, respectively, the forth and fifth rows show the segmentation results after applying a heave of $A^h = 3\lambda_c$ and $A^h = 7\lambda_c$ to the reconstruction process. For both kinds of translational errors, the deterioration of the segmentation results is gradual, and it is specially significant for the shadow segmentation results (in yellow). Compare, for instance, the ICM segmentation result (first column) for heave values of $A^h = 3\lambda_c$ and $A^h = 7\lambda_c$. Figure 5.6 shows the influence of two different sound speed mismatches, namely, for $q_\epsilon = 1.2$ % and $q_\epsilon = 2.1$ %, on the performance of both segmentation algorithms for two different

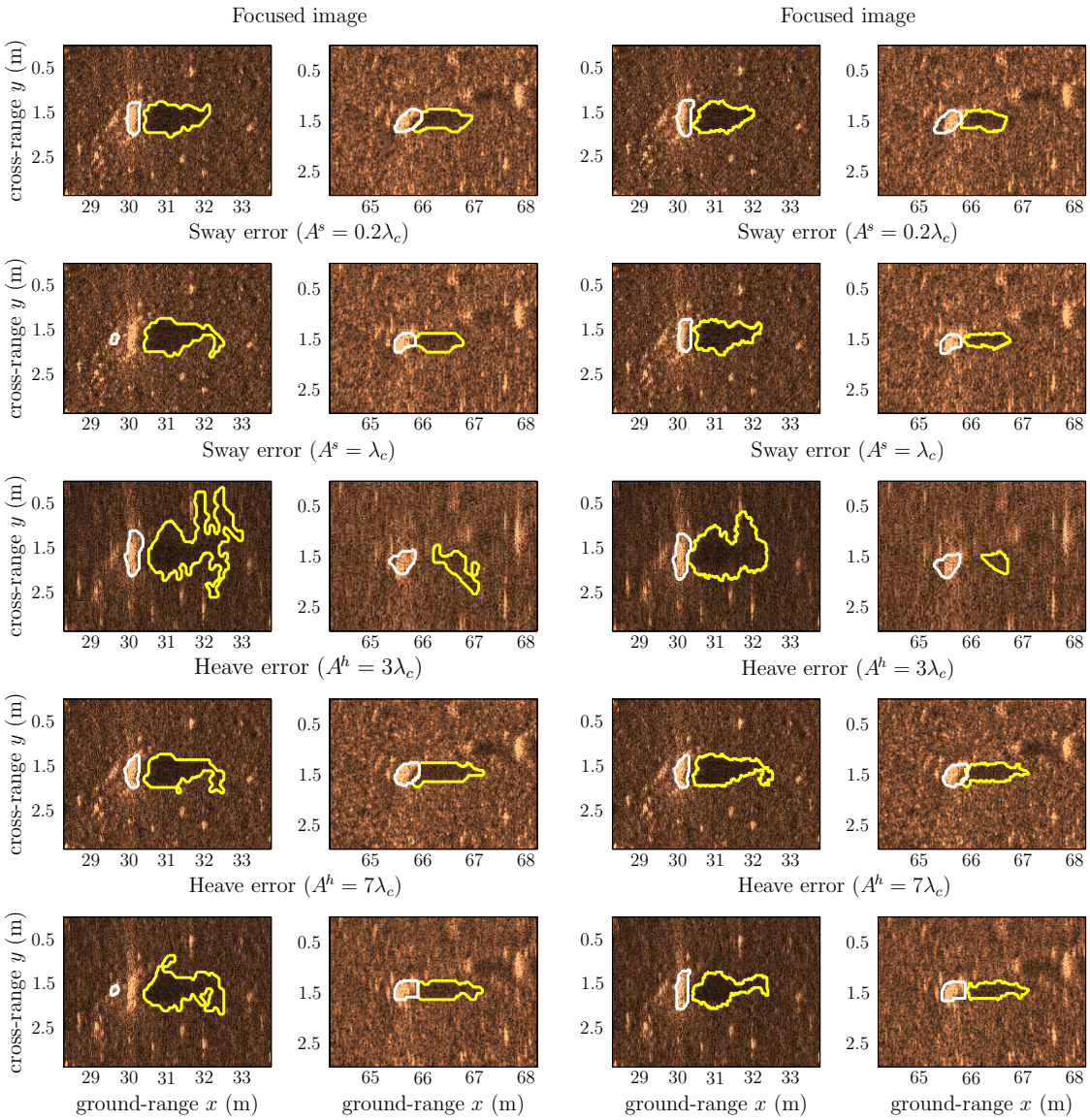


Figure 5.5: Example segmentation results for ICM (1st and 2nd column) and GC (3rd and 4th column) for two mine-like objects, respectively, and two different amplitudes of sway (2nd and 3rd row) and heave (4th and 5th row).

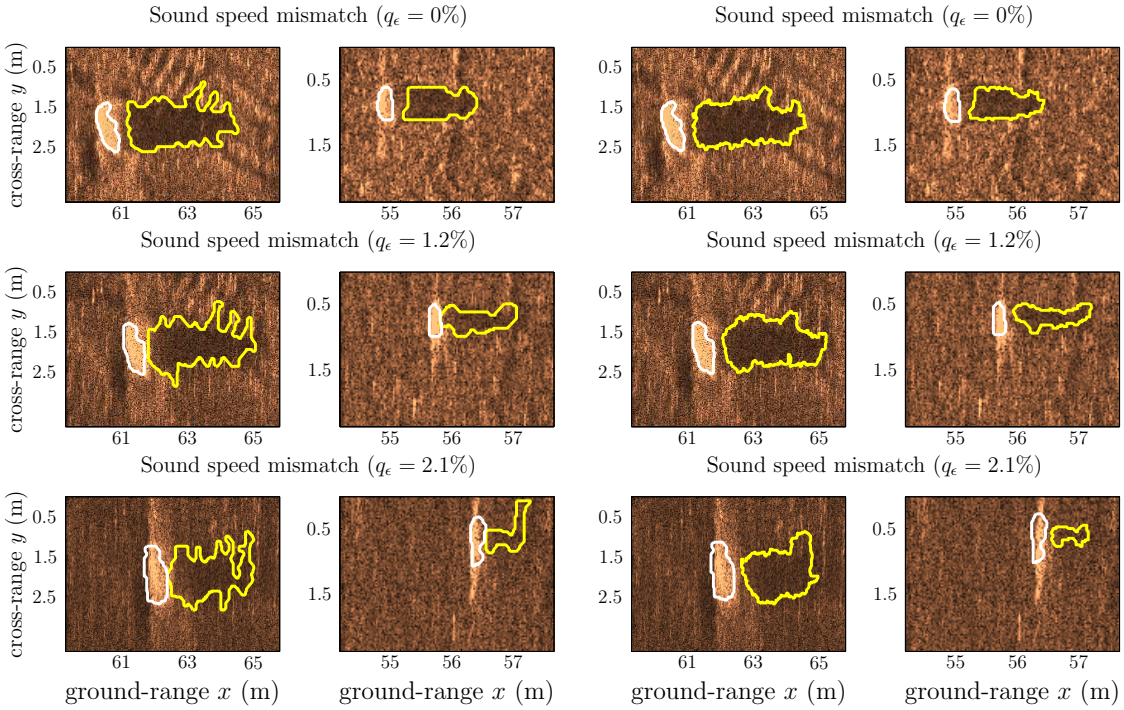


Figure 5.6: Example segmentation results for ICM (1st and 2nd column) and GC (3rd and 4thcolumn) for two mine-like objects and two different sound speed mismatch values.

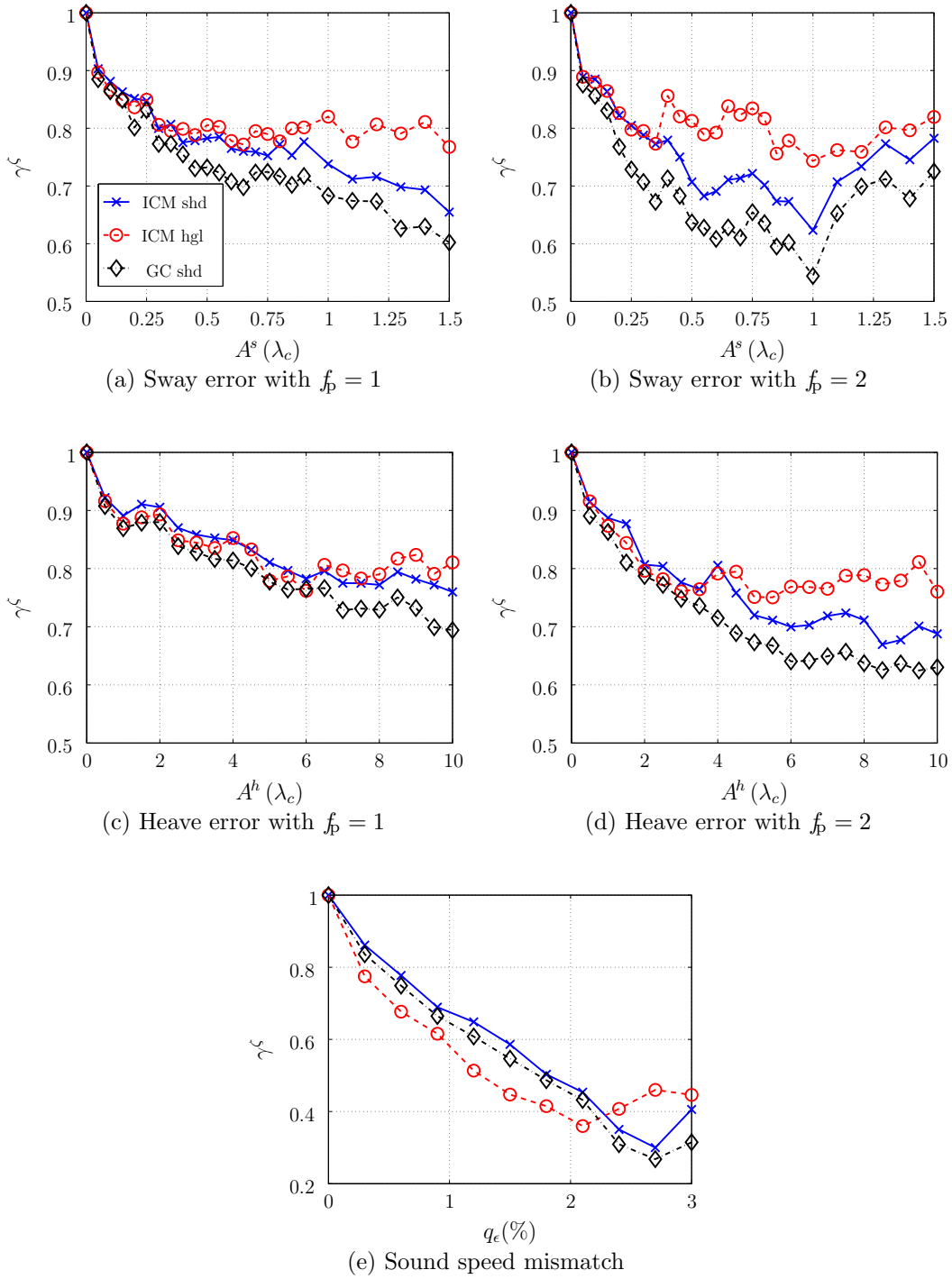
snapshots. As with motion errors, the sound speed mismatch has also a significant effect on the segmentation results. In general, the effect of the sinusoidal motion errors and sound speed mismatches are comparable for the examples of the ICM and GC segmentation methods.

In order to analyze the influence of uncompensated motion errors on the segmentation procedure, the ratio of correctly segmented pixels is averaged for all image snippets as a function of the amplitude of the respective SAS reconstruction error, *i.e.*, sway, heave and sound speed mismatch. The analysis distinguishes between shadow (shd) and highlight (hgl) regions. This is useful to assess the reliability of the features extracted from the two segmented regions. In the following, the average segmentation ratio is first described, and the corresponding results are then shown for the used database.

Consider the binary matrix $\boldsymbol{\mathcal{S}}^{\zeta}$ with $\zeta \in \{\text{shd}, \text{hgl}\}$, which is obtained after applying a segmentation technique to the SAS image. The matrix contains all segmented pixels of the corresponding region of interest ζ and can be expressed as follows:

$$\boldsymbol{\mathcal{S}}^{\zeta} = \begin{cases} \{1_{k,l}\} & \text{if } \hat{f}(x_k, y_l) \in \zeta \\ \{0_{k,l}\} & \text{otherwise} \end{cases} \tag{5.1}$$

$$\text{for } \; k = 1, \ldots, M_x \; \text{ and } \; l = 1, \ldots, M_y.$$



(a) Sway error with $f_\mathrm{p} = 1$

(b) Sway error with $f_\mathrm{p} = 2$

(c) Heave error with $f_\mathrm{p} = 1$

(d) Heave error with $f_\mathrm{p} = 2$

(e) Sound speed mismatch

Figure 5.7: Average segmentation ratio for highlight and shadow region for ICM and GC segmentation and an increasing error strength. Results are shown for a sway error with $f_\mathrm{p} = 1$ in (a) and $f_\mathrm{p} = 2$ in (b), heave error with $f_\mathrm{p} = 1$ in (c), $f_\mathrm{p} = 2$ in (d) and for a sound speed mismatch in (e).

Considering the number of objects, N_{obj}, in the database and a motion error amplitude vector $\{A_j^{\epsilon}\}$ with $\epsilon \in \{s, h\}$, a binary segmentation matrix $\boldsymbol{\mathcal{S}}_{i,j}^{\zeta}$ is obtained for each object, $i = 1, \ldots, N_{\text{obj}}$, and motion error amplitude, $j = 0, \ldots, N_{\text{err}}$. Here, N_{err} is the number of different error strengths. Thus, to measure the variation of the segmentation due to reconstruction errors, the number of overlapping pixels between the ground truth ($j = 0$) and the segmentation obtained is determined after inducing errors as

$$o_{i,j}^{\zeta} = \# \left\{ \boldsymbol{\mathcal{S}}_{i,j}^{\zeta} \odot \boldsymbol{\mathcal{S}}_{i,0}^{\zeta} \right\}, \tag{5.2}$$

where $\odot$ denotes the Hadamard product and $\# \{\cdot\}$ is an operator, which counts the number of non-zero elements of a matrix. The average ratio of correctly segmented pixels γ_j^{ζ} is then calculated depending on the error amplitude index j as follows

$$\gamma_j^{\zeta} = \frac{1}{N_{\text{obj}}} \sum_{i=1}^{N_{\text{obj}}} \Gamma_{i,j}^{\zeta} \quad \text{for} \quad j = 0, \ldots, N_{\text{err}} \tag{5.3}$$

with

$$\Gamma_{i,j}^{\zeta} = o_{i,j}^{\zeta} / o_{i,0}^{\zeta}. \tag{5.4}$$

The segmentation analysis of the ICM and GC segmentation methods, both for shadow and highlight regions and for the three different error types are depicted in Figure 5.7. The sway results are shown in Figure 5.7a-b, Figure 5.7c-d corresponds to the heave and Figure 5.7e to the mismatch in sound speed. All curves show a fast decreasing behavior already for small error amplitudes, where the degradation of the segmentation is comparable for both regions of interest, *i.e.*, shadow and highlight. As the error amplitude increases, small fluctuations in the figure of merit of segmentation performance can be observed, but the average tendency is clearly decreasing. These fluctuations are likely to be a reason for the noisy behavior of the feature sensitivity plots in Section 5.2.4 as well as of the misclassification results in Section 5.2.5. Another reason is the limited size of the database.

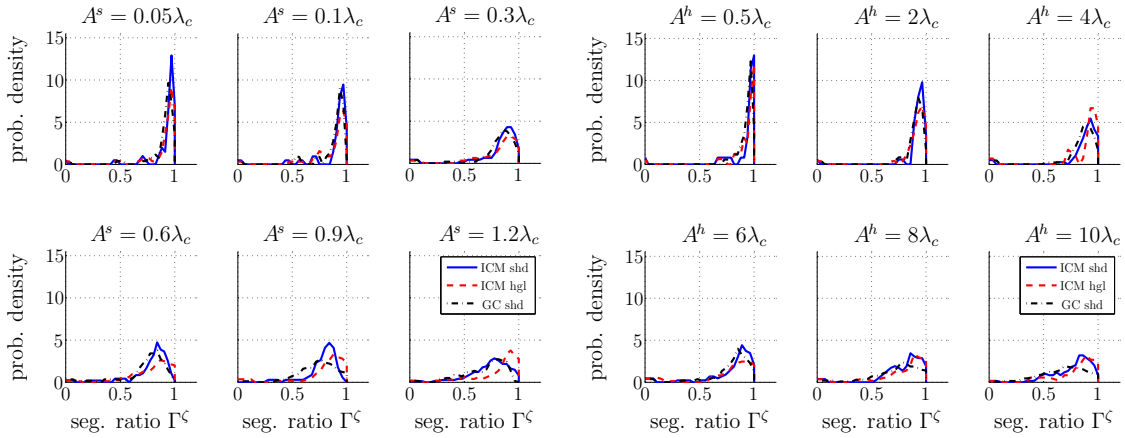
Furthermore, the degradation in the segmentation seems to be almost identical for both regions of interest, *i.e.*, shadow and highlight, in case of small error amplitudes. Only for larger amplitudes, there is a noticeable difference in the behavior of the shadow and highlight segmentation ratios. While the highlight curves approximate a constant, *e.g.*, $\gamma^{\text{hgl}} \approx 0.8$ for a sway error with $A^s > 0.25\lambda_c$, the shadow segmentation ratio continues to decrease for an increasing error amplitude – except for a sway error with $f_{\text{p}} = 2$ in Figure 5.7b. Consequently, the results of this empirical study demonstrate that uncompensated motion errors destroy shadow information rather than highlight information during the SAS reconstruction for increasing error amplitudes. For smaller error amplitudes, the average segmentation ratio is essentially identical for the shadow

and the highlight region in the case of ICM segmentation. Note that the highlight region is only segmented using the ICM method. Since the GC algorithm is only able to segment the images into two regions, it is not feasible to obtain a segmentation of both highlight and shadow using this method. The ICM segmentation of the highlight is satisfactory, and therefore, the GC algorithm is only employed to distinguish between shadow and background. However, comparing the average segmentation ratios, the ICM appears to be more resistant to motion errors than the GC algorithm in all cases. For an increasing sound speed mismatch as depicted in Figure 5.7e, average segmentation ratio performance is different than in the case of motion errors – see Figure 5.7a-d. The highlight region suffers slightly more than the shadow region. Moreover, a steeper and rather continuous decay is observable.

As expected, the average segmentation ratio is strongly correlated with the deterioration of the image quality as illustrated in Section 5.2.2. For example, for a sway error with $f_\mathrm{p} = 1$, there is a strong decay in γ^ζ as well as in the SSIM curve for error amplitudes until approximately $A^s \leq 0.25\lambda_c$. For larger values, the segmentation ratio curves start to differ for highlight and shadow region as pointed out earlier. While the highlight segmentation ratio γ^hgl approaches a constant value, the shadow segmentation ratio γ^shd continuous to decay but at a much smaller rate. An analogous behavior can be observed in the corresponding SSIM curve – see blue line in Figure 5.4a. First, the SSIM decays fast for $A^s \leq 0.25\lambda_c$, then decays slower in the interval $0.25\lambda_c \leq A^s \leq \lambda_c$ and finally approaches a constant value of approximately 0.4 for $A^s > \lambda_c$. This characteristic tendency is similar for the other motion errors and for the sound speed mismatch. For the latter, however, the decay in both SSIM curve and average segmentation ratio is much stronger than for the motion errors. As a consequence, these results illustrate that the distortion in image quality directly affects segmentation algorithms, which are not robust against motion errors.

A second analysis of the segmentation degradation has been conducted for both segmentation methods and regions of interest by considering the empirical distribution of the segmentation ratio $\Gamma^\zeta_{i,j}$ of (5.3) with respect to the object dimension i and for individual motion error amplitudes. Example results for selected error amplitudes of a sway and a heave error with $f_\mathrm{p} = 1$ are depicted in Figure 5.8a and Figure 5.8b, respectively. Note that the empirical probability density functions have been estimated using the kernel density estimation (KDE) technique [Silverman (1986)] with a rectangular kernel. As can be seen in Figure 5.8a for $A^s = 0.05\lambda_c$, the mode of the distribution is very close to one and the scale is rather small. Thus, this indicates that the segmentation ratio is still close to one for the majority of objects. However, for an increasing sway error amplitude A^s, the mode of the empirical distribution shifts towards smaller ratios. Furthermore, the scale of the distribution broadens. While the

mode shift demonstrates the general tendency of the degradation in segmentation, the increase in scale illustrates a varying impact of the motion errors on the segmentation of different objects, *e.g.*, due to varying stand-off distances between object and sonar trajectory and a random component of the segmentation algorithms. A similar characteristic is observable for the heave results in Figure 5.8b.



(a) Sway error $f_\mathrm{p} = 1$ (b) Heave error $f_\mathrm{p} = 1$

Figure 5.8: Empirical probability density function of the segmentation ratio $\Gamma_{i,j}$ with respect to object dimension i for different amplitudes of the motion error vector $\{A_j^\epsilon\}$. The density function is depicted for both segmentation algorithms, the two regions of interest for a sway error with $f_\mathrm{p} = 1$ in (a) and for heave error with $f_\mathrm{p} = 1$ in (b).

5.2.4 Feature extraction

The previous subsection has empirically demonstrated the deterioration of segmentation due to sinusoidal motion errors and sound speed mismatches as a consequence of a degrading image quality. In this section, a quantitative assessment of the effect of motion errors on feature extraction is examined, which is the next block in the ADAC processing chain – see Figure 5.1. To this end, a metric is introduced that measures the distance between a given feature value affected by motion errors and its corresponding ground truth. The latter is considered as the value obtained during feature extraction in the error-free scenario.

Let $\varphi_{m,i,j}$ represent the m^{th} feature value of object i, with $m = 1, \ldots, N_\varphi$ and $i = 1, \ldots, N_\mathrm{obj}$ of a SAS image, which has been reconstructed after inducing a motion error with the j^{th} amplitude of the amplitude error vector $\{A_j^\epsilon\}$, with $j = 0, \ldots, N_\mathrm{err}$. The feature matrix $\boldsymbol{\varphi}_m = \{\varphi_{i,j}\}_m$ is normalized for each individual feature m such that $\bar{\boldsymbol{\varphi}}_m \in [0, 1]$, where $\bar{\boldsymbol{\varphi}}_m = \{\bar{\varphi}_{i,j}\}_m$ denotes the normalized feature value matrix of the

m^{th} feature. The absolute distance of the normalized feature values with respect to the error-free feature value, *i.e.*, for $j = 0$, is then given by

$$\Delta\bar{\varphi}_{m,i,j} = \left|\bar{\varphi}_{m,i,j} - \bar{\varphi}_{m,i,0}\right| . \tag{5.5}$$

Hence, for each feature m and each motion error index j, the mean-square error is calculated as follows

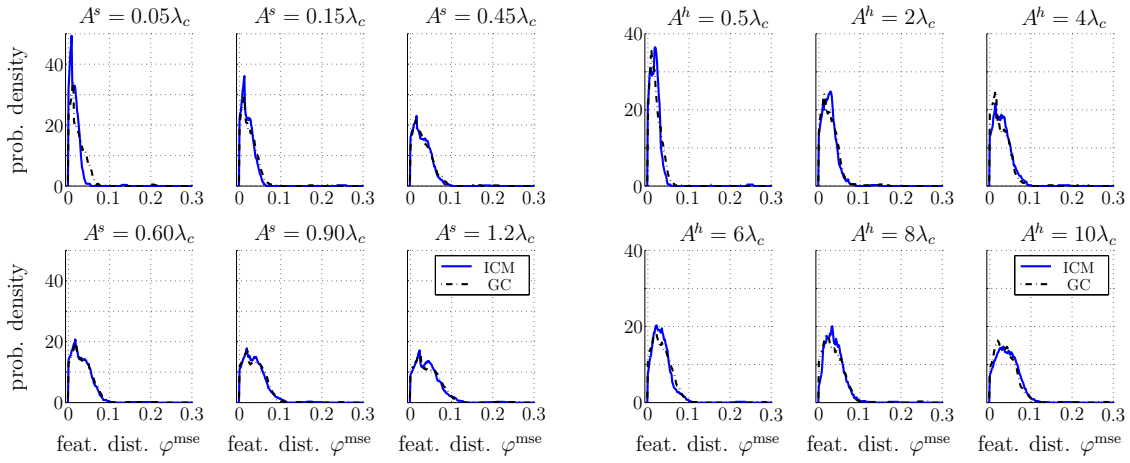
$$\varphi^{\text{mse}}_{m,j} = \frac{1}{N_{\text{obj}}} \sum_{i=1}^{N_{\text{obj}}} \Delta\bar{\varphi}^2_{m,i,j}. \tag{5.6}$$

The degradation of feature values depending on the error amplitude is illustrated in two different ways. First, the distribution of the feature values is depicted for a selection of error amplitudes, which is shown in Figure 5.9a for a sway error with $f_{\text{p}} = 1$, in Figure 5.9b for a heave error with $f_{\text{p}} = 1$, and in Figure 5.9c for the sound speed mismatch. This measure of feature degradation is analogous to the illustration of the empirical probability density functions of the segmentation ratio in Figure 5.8. For both segmentation methods, Figure 5.9a shows the density function of (5.6) with respect to the feature dimension for increasing indices of the sway error amplitude vector $\{A^{\epsilon}_j\}$ and $f_{\text{p}} = 1$. For all three motion errors and for both segmentation algorithms, it is evident that despite small fluctuations the scale of the density function of the features rises for an increasing error amplitude. Simultaneously, the mode shifts towards larger values of the mean-square feature distance. Both effects demonstrate the loss in feature quality. Moreover, these results also agree with the progressive degradation of the segmentation observed in Figure 5.8.

A different perspective of this degradation is provided by Figure 5.10 and Figure 5.11, which presents the second illustration of the feature sensitivity against motion errors, again, for the sway and the heave with $f_{\text{p}} = 1$, respectively. These figures show the average of the feature values within a feature group for all objects as a function of the error amplitude. Figure 5.10a and Figure 5.10b show the results for the GC and ICM method, respectively. Each subfigure contains seven curves showing the feature average of the seven different types of feature groups as listed in Section 5.1.2. An additional curve (thicker line) represents the average over the optimal feature subset as selected by the D_{SFS}-SFS algorithm – see Section 5.1.3. The subset contains $N'_{\varphi} = 33$ features for the ICM and $N'_{\varphi} = 41$ for the GC algorithm. These optimal features are selected to maximize performance without motion errors and contain features from all seven groups. Note that the D_{SFS}-SFS algorithm does not find the optimal feature subset regarding motion errors. It just selects the optimal subset for the training database.

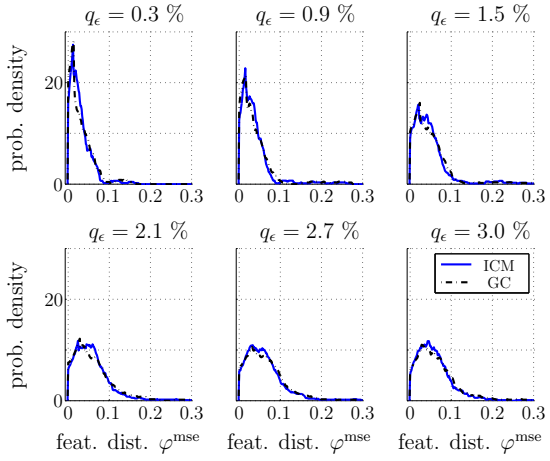
In general, the principal components ($\boldsymbol{\varphi}_{\text{PCA}}$) as well as the Fourier features ($\boldsymbol{\varphi}_{\text{Fourier}}$) are the most sensitive features on average while the normalized central moments ($\boldsymbol{\varphi}_{\mu}$)

and the invariant moments ($\boldsymbol{\varphi}_{\text{inv}}$) are the least ones. Logically, the sensitivity of the optimal feature subset $\boldsymbol{\varphi}'$ lies in between these extreme cases for all motion errors since it contains features from all seven groups. This characteristic is similar for both segmentation approaches. Given that the highlight is not segmented by the GC algorithm, the ICM highlight segmentation is assumed. For this reason, the object feature sensitivity curves are identical for both segmentation algorithms. Furthermore, the statistical feature sensitivity curves are also identical. The statistical features are based on a Weibull model of the SAS image and the estimated distribution parameters for all three regions. The difference between the two Weibull parameters and all possible combinations of segmented regions constitute the statistical features – see



(a) Sway error $f_{\text{p}} = 1$

(b) Heave error $f_{\text{p}} = 1$



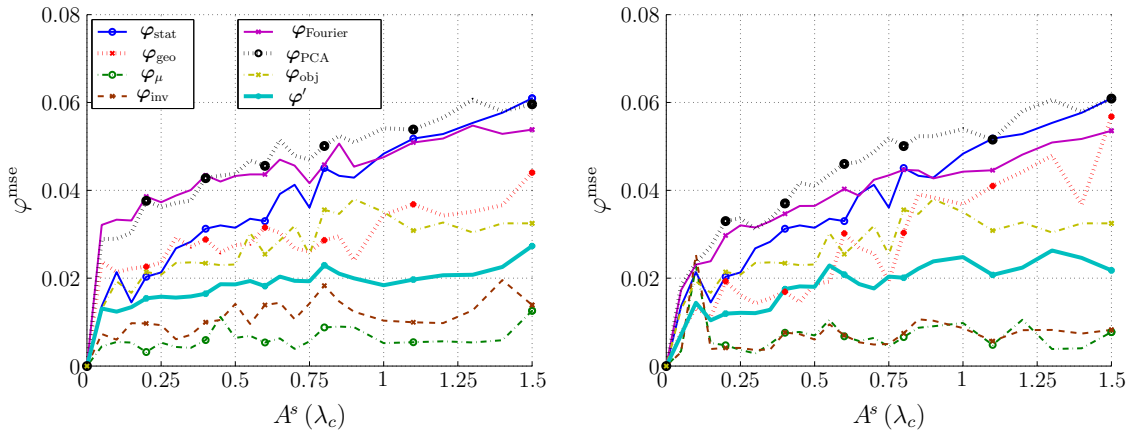
(c) Sound speed mismatch

Figure 5.9: The empirical probability density function of the feature distribution is depicted for both segmentation algorithms and for all error types with different amplitudes. The subplots show the density functions of the sway error with $f_{\text{p}} = 1$ in (a), the heave error with $f_{\text{p}} = 1$ in (b) and sound speed mismatches in (c).

[Fandos et al. (2013)] for details. The fact that the GC does not segment the highlight has two consequences. On the one hand, obviously, the Weibull parameters for the highlight are not available for the GC segmentation method. On the other hand, the highlight pixels remain part of the background region and thus, the Weibull parameter estimates of the background are inaccurate. Hence, the ICM statistical features are also used for the classification stage, which is based on the GC segmentation method.

A significant difference between both segmentation algorithms is observable for the geometrical feature group, which is as expected considering the difference in the average segmentation ratios of Figure 5.7. The remaining feature groups are comparable in their characteristics for both segmentation methods except that features obtained by GC segmentation are immediately affected strongly after inducing motion errors while the increase in sensitivity is less abrupt for ICM – compare Figure 5.10a with Figure 5.10b for $A^s = 0.05\lambda_c$. The comparison of segmentation ratio and feature sensitivity is very interesting. Although small fluctuations exist, it is obvious that the quality of the segmentation determines the quality of the features. For example, the highlight segmentation ratio stabilizes for $A^s > 0.25\lambda_c$ rather than continuing to decrease as does the shadow segmentation ratio for both methods in Figure 5.7a. A similar behavior is noticeable when comparing the object feature ($\boldsymbol{\varphi}_{\text{obj}}$) group (mainly based on highlight information) with the statistical feature ($\boldsymbol{\varphi}_{\text{stat}}$) group, principal components ($\boldsymbol{\varphi}_{\text{PCA}}$) and Fourier feature ($\boldsymbol{\varphi}_{\text{Fourier}}$) group. It is difficult to predict whether classification performance in Section 5.2.5 is better for ICM or GC segmentation based on the feature sensitivity of the used subsets.

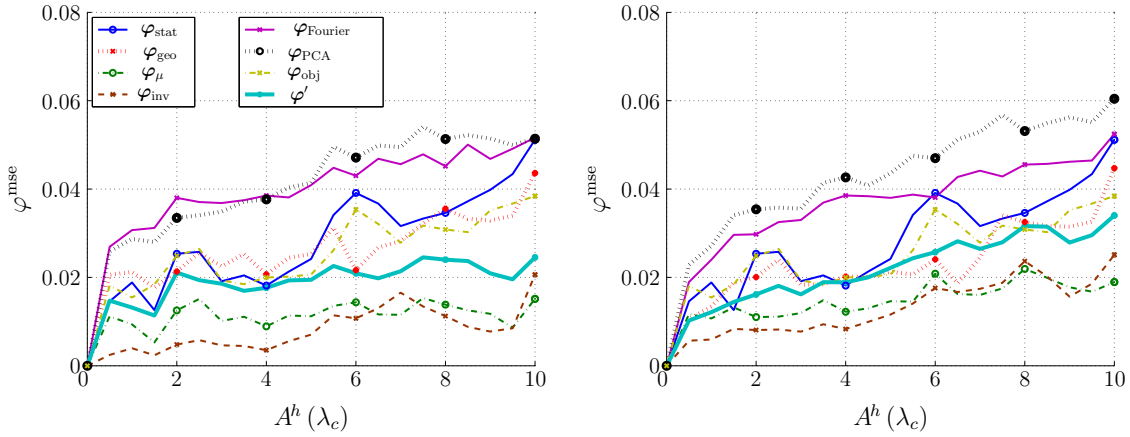
For the heave error with $f_{\text{p}} = 1$, a similar characteristic is observable. The segmentation



(a) GC segmentation

(b) ICM segmentation

Figure 5.10: Motion error sensitivity of feature groups and selected features for an increasing sway error with $f_{\text{p}} = 1$.



(a) GC segmentation

(b) ICM segmentation

Figure 5.11: Motion error sensitivity of feature groups and selected features for an increasing heave error with $f_{\text{p}} = 1$.

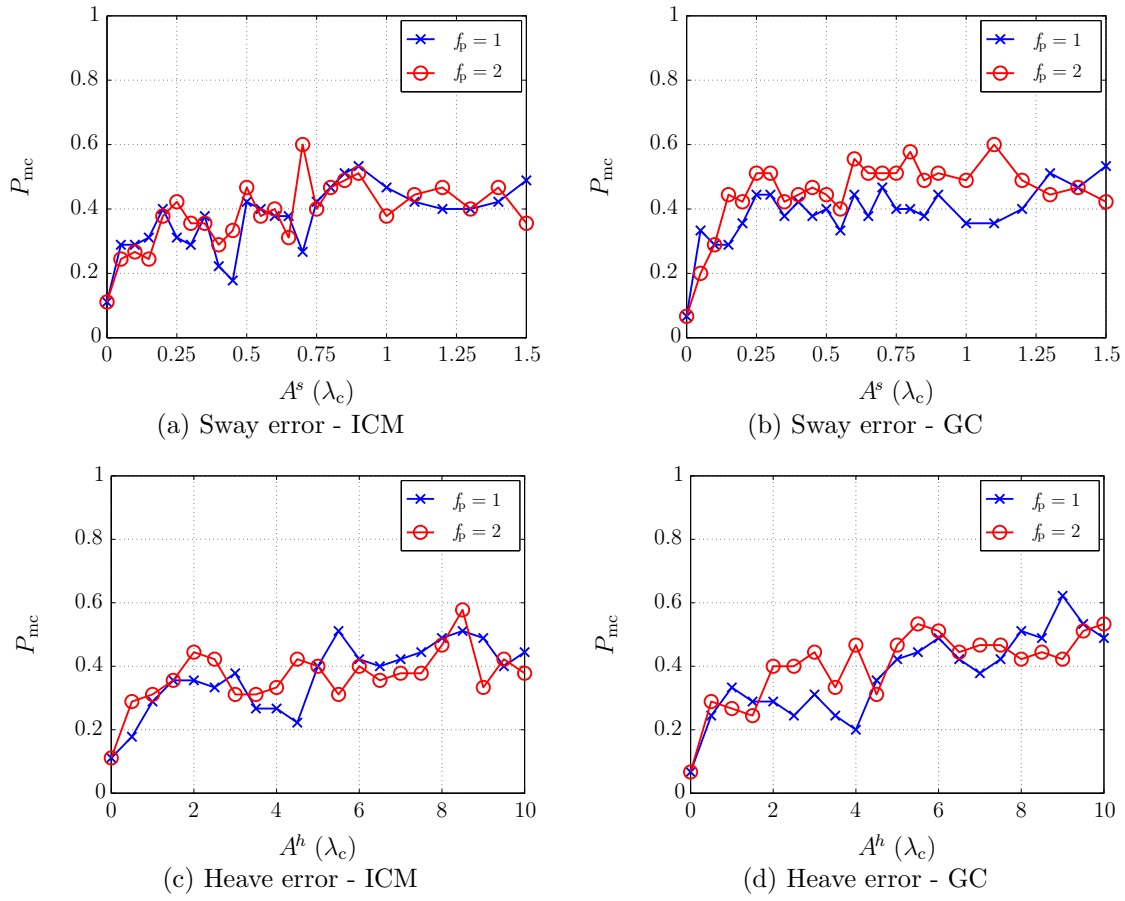
ratio curves in Figure 5.7c degrade constantly for increasing motion error amplitudes as does the feature sensitivity in Figure 5.11. Another interesting observation is the correlation between the abrupt jump in the average segmentation ratio curves of Figure 5.7 after inducing motion errors with the sudden increase in the feature distance metric of Figure 5.10 and Figure 5.11. Although the difference is small between the corresponding values of both segmentation methods, the jump of the feature distance is less on average for the ICM than for the GC – compare Figure 5.10a with Figure 5.10b and Figure 5.11a with 5.11b. The same is valid for the average segmentation ratio. Again, the effect is small. Nevertheless, it shows that the ICM is less sensitive to smaller motion error amplitudes than the GC. For larger motion errors, both segmentation algorithms generate features of equal sensitivity. Note that the sensitivity curves do not indicate the quality of features to separate object classes.

5.2.5 Classification

As a result of image quality degradation, the segmentation of highlight and shadow region worsens, which affects the extracted features. Consequently, an impact on the classification performance is to be expected, which is investigated in the following. Figure 5.12 shows the evolution of the misclassification rate P_{mc} with the amplitude of the sway (first row) and heave (second row) errors, for the ICM (first column) and GC (second column) segmentation technique and for both cycles per synthetic aperture length frequency, namely, for $f_{\text{p}} = 1$ (blue curve) and $f_{\text{p}} = 2$ (red curve). A fixed false alarm rate of $P_{\text{fa}} = 0.2$ has been used to minimize P_{mc}. For the application

at hand, it is of priority to minimize P_{mc} even at the expense of a relatively high P_{fa}. The misclassification rate P_{mc} represents the percentage of mines classified as clutter, spherical mines classified as cylindrical mines and cylindrical mines classified as spherical mines.

With no motion errors ($A^{\epsilon} = 0$ with $\epsilon \in \{s, h\}$), the misclassification rate is $P_{\text{mc}} = 0.11$ when the ICM segmentation algorithm is employed, and $P_{\text{mc}} = 0.067$ if the GC method is used, see difference for $A^{s} = 0$ in Figure 5.12a and Figure 5.12b. Although the curves show fluctuations due to the limited size of the test database, in average, P_{mc} increases as the amplitude of the motion errors increases for both segmentation methods. The shape of the P_{mc} curves are in agreement with the previously seen degradation of SSIM curves in Figure 5.4, the average segmentation ratio in Figure 5.7 and the loss in feature accuracy in Figure 5.10 and Figure 5.11. For example, in Figures 5.12a and 5.12b, referring to the sway error with $f_{\text{p}} = 1$, a steep increase in the value of P_{mc} is observed



(a) Sway error - ICM

(b) Sway error - GC

(c) Heave error - ICM

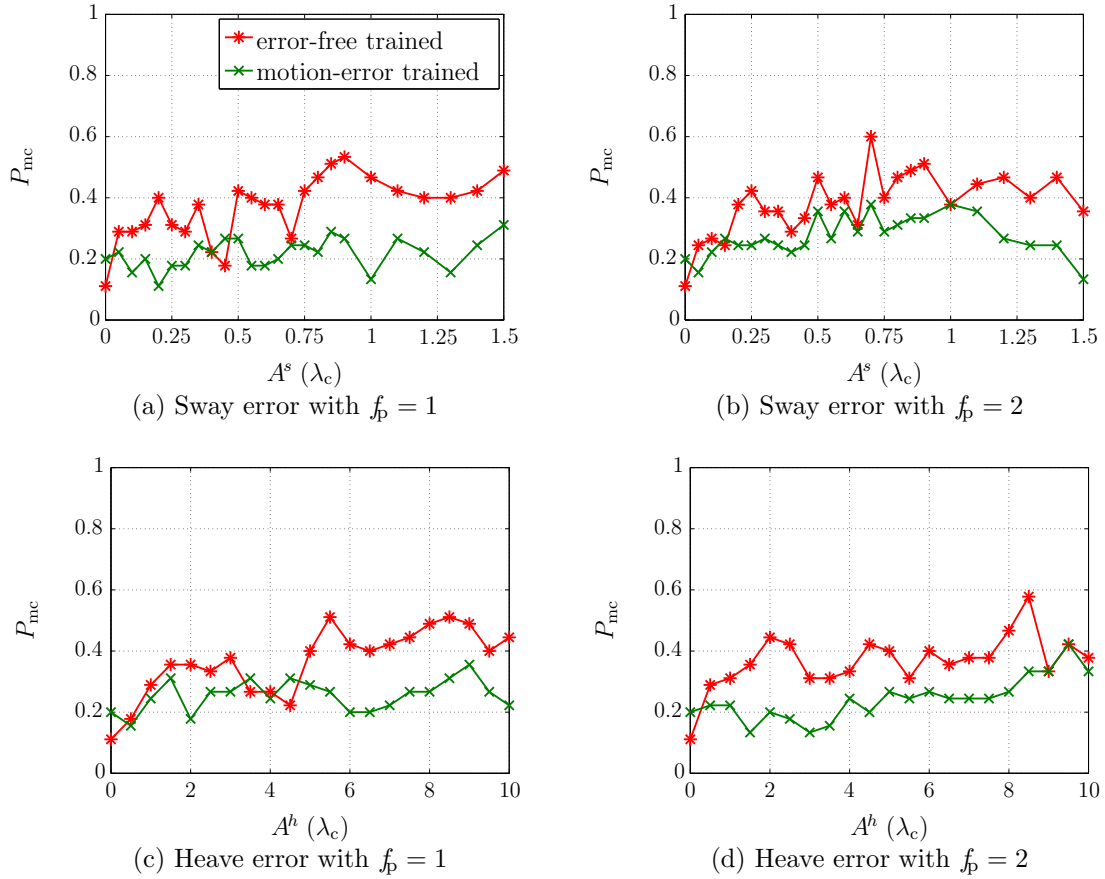
(d) Heave error - GC

Figure 5.12: Misclassification rate of sway (a)-(b) and heave (c)-(d) affected SAS images with increasing error amplitudes A^{ϵ} (λ_{c}), two different cycles per synthetic aperture length frequencies, $f_{\text{p}} = 1$ and $f_{\text{p}} = 2$, for ICM and GC segmentation.

for error amplitudes with $A^s \leq 0.2\lambda_c$. For $A^s = 0.2\lambda_c$ and $f_\mathrm{p} = 1$, the misclassification rate is approximately $P_\mathrm{mc} \leq 0.4$ for both segmentation methods. As the sway amplitude further increases until $A^s = 1.5\lambda_c$, the misclassification rate increases only slightly. This is in agreement with the fast drop of the SSIM curve for sway amplitudes with $A^s \leq 0.2\lambda_c$ and its rather stationary behavior for $A^s > 0.2\lambda_c$ in Figure 5.4a. Similar results are observed for a sway error with $f_\mathrm{p} = 2$. For the heave with $f_\mathrm{p} = 2$, a fast decay in the SSIM value is observed for $A^h \leq 2\lambda_c$ in Figure 5.4b, which is in agreement with the P_mc increase observed in the same range of error amplitudes A^h – $P_\mathrm{mc} \approx 0.4$ for both ICM and GC at $A^h = 2\lambda_c$. For larger values of the error amplitude, both the SSIM and P_mc curves show a rather stationary behavior. Although more strongly fluctuating, a comparable tendency can be noticed for a heave error with $f_\mathrm{p} = 1$.

It has been demonstrated how classification performance of the ADAC system is sensitive to motion errors given the degradation in image quality. The latter affects the segmentation and feature quality as discussed in the previous sections. Also note that the average segmentation ratio of the highlight has identical thresholds, *i.e.*, $A^s = 0.2\lambda_c$ and $A^h = 2\lambda_c$. At these error amplitudes, the curves fluctuate around a constant value. As expected, considering the typical geometry of an AUV based SAS system, especially the sway significantly degrades performance already for small error amplitudes with $A^s < 0.2\lambda_c$. Comparing the misclassification rates in the case of GC segmentation in Figures 5.12b and 5.12d with the SSIM curve for sway and heave errors in Figure 5.4a and Figure 5.4b, respectively, it becomes also apparent that the misclassification rate is worse for a cycle per synthetic aperture length frequency of $f_\mathrm{p} = 2$ than for $f_\mathrm{p} = 1$. The same is observed for the image quality, which decays slightly faster for $f_\mathrm{p} = 2$. However, this cannot be claimed for the misclassification rate using the ICM segmentation. While the ICM has shown less sensitivity to motion errors for segmentation and consequently for feature extraction, it is not feasible to state whether classification based on ICM segmentation is performing better than for GC segmentation when comparing Figure 5.12a with 5.12b and Figure 5.12c with 5.12d.

In the following, it is shown how it is possible to increase the robustness of the ADAC system against motion errors in the test database by using a training database that consists of the same image snippet set but has been reconstructed with motion errors ("motion-error trained"). Figure 5.13 compares classification performance of the system using ICM segmentation when the training database does not contain motion errors ("error-free trained" - as illustrated by the curves in Figure 5.12a for the sway and in Figure 5.12c for the heave) with its performance when the training database includes a sway error with an amplitude of $A^s = 0.2\lambda_c$ (green curves - "motion-error trained"). When motion errors are present in the test database, the misclassification rate is generally smaller if a training database with motion errors has been employed



(a) Sway error with $f_p = 1$

(b) Sway error with $f_p = 2$

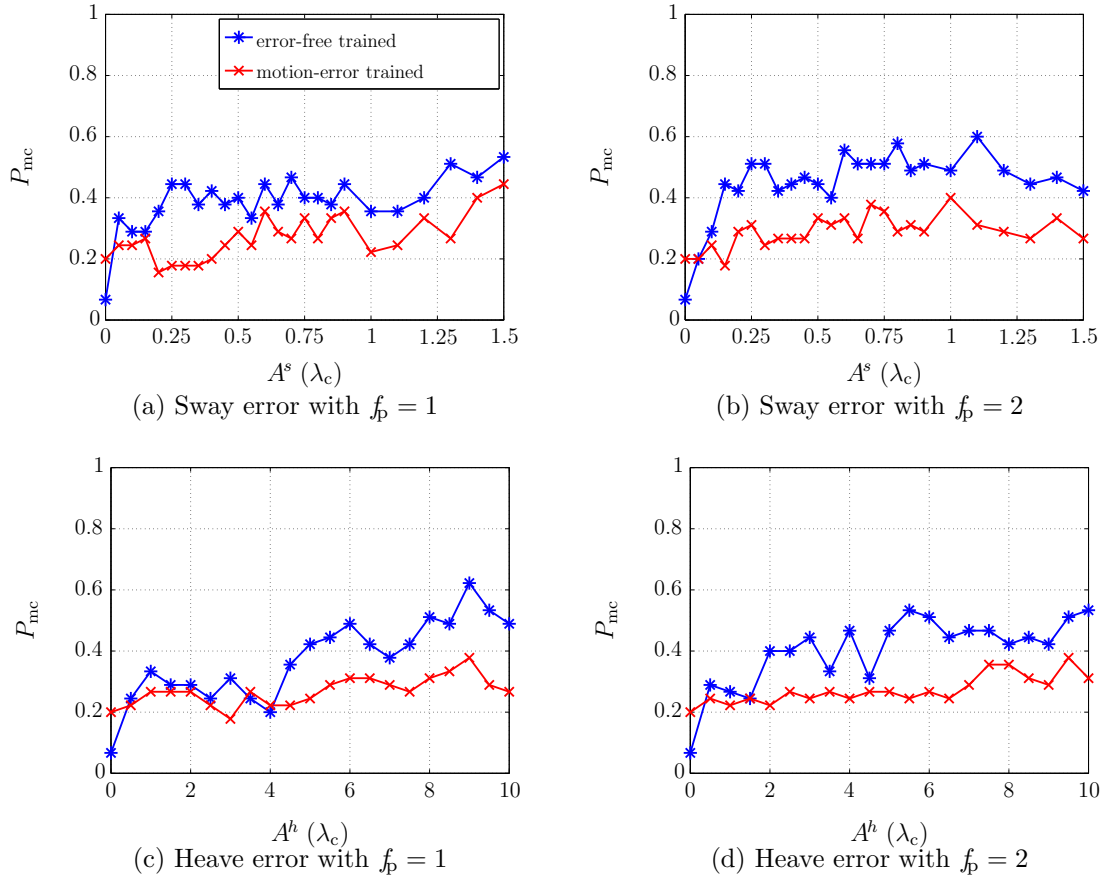
(c) Heave error with $f_p = 1$

(d) Heave error with $f_p = 2$

Figure 5.13: Misclassification rate when motion errors are included in the training set for the ICM segmentation method.

(green curve reaches lower values than red curve in Figure 5.13). As expected, when the test database contains no errors the value of P_{mc} is lower if the training database also contains no errors. Although only sway errors of a certain amplitude have been induced in the "motion-error trained" database, the ADAC system becomes more robust to not only sway but also to heave motion errors in the test database. The corresponding misclassification curves highlighting this result are depicted in Figure 5.13 for the sway and heave for $f_p = 1$ and $f_p = 2$. The improved resistance against heave errors in the case of a sway motion training of the ADAC system is because both sway and heave motion errors follow a sinusoidal path deviation. This implies that their influence on the image reconstruction process is almost identical, except for the sensitivity due to the geometry of the imaging system (here, much higher for the sway), which leads to similarly distorted images.

The reason why misclassification rate performance improves in the "motion-error" trained scenario is that the training images are more similar to the images with motion errors. Therefore, the features are more appropriate in the sense that the



(a) Sway error with $f_p = 1$

(b) Sway error with $f_p = 2$

(c) Heave error with $f_p = 1$

(d) Heave error with $f_p = 2$

Figure 5.14: Misclassification rate when motion errors are included in the training set for the GC segmentation method.

feature distances are smaller to the reference feature values of the training set. This can be explained by Figure 5.10. While there is an abrupt increase in the feature sensitivity for $A^s > 0$, the feature sensitivity only slightly increases for error amplitudes $A^s > 0.2\lambda_c$. Thus, the used features are more similar to each other for the "motion-error trained" scenario. One could argue that training the system with a database that contains motion errors is already selecting the feature subset that is less sensitive to motion errors. However, in order to do this properly, the training of the system has to consider different data sets with motion errors of different amplitude. Note that analogous results have been obtained for the GC segmentation method, as illustrated in Figure 5.14. The misclassification rate for rising mismatches in the measured and actual sound speed is depicted in Figure 5.15. Again, the shape of the P_{mc} curve is in agreement with the SSIM curve in Figure 5.4c. Both of them are rather linear, unlike the motion error effect as described above. Moreover, the SSIM values are much lower than for the motion errors for large mismatch values in sound speed. However, note that mismatch values with $q_\epsilon > 2\%$ are practically not occurring. A similar

curve characteristic can also be seen in the average segmentation ratio in Figure 5.7d. Here, the segmentation ratio is decaying more than for the motion errors. This is in agreement with the SSIM curve and the misclassification rate.

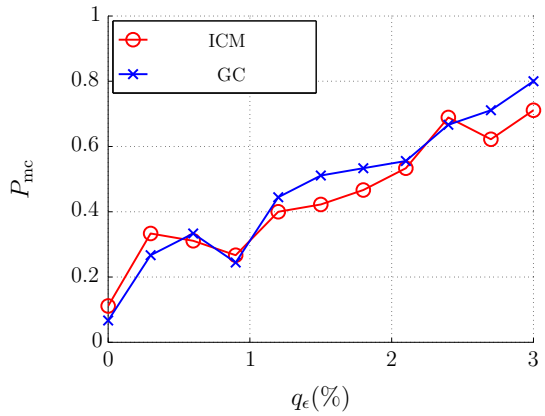


Figure 5.15: Misclassification rate for SAS images affect by increasing sound speed mismatches.

5.3 Sequential focus assessment

In the application of automatic target recognition (ATR) for naval mine hunting, high-resolution SAS images serve as input to an ADAC post-processing system, which highly relies on excellent image quality. Moreover, most existing ATR systems assume that the image quality is uniformly excellent for the entire sonar image, which is often not valid in practice [Williams and Groen (2011)]. For an operator-based system, it is feasible to intervene after the image reconstruction process to assess the image quality, and thus, guarantee a reliable input for an ADAC system. However, in a fully automatic mine hunting system, the omission of an operator necessitates an assessment scheme to still guarantee that high-quality images are used. Otherwise, a degradation in classification performance may occur as shown in Section 5.2 and published in [Leier et al. (2014)].

In this section, a strategy is proposed to assess the focusing capability during the reconstruction of a SAS image (and thereby its quality) by probing the instantaneous cross-range resolution of a synthetic sub-aperture and comparing it with its theoretical resolution. Since practical SAS systems consist of a receiver array of hydrophones – see Chapter 3, they provide the possibility of constructing low-resolution real aperture sonar (RAS) images. This can be exploited to successively evaluate the cross-range focusing of the synthetic aperture, and therefore, the quality of the resulting

images [Leier and Zoubir (2011)]. Besides assessing the reliability of the input for ADAC tasks, image quality evaluation is also important for adaptive mission planing [Williams et al. (2012)]. Since the quality of SAS images degrades especially at long ranges due to more stringent demands on the motion estimation accuracy or due to multipath effects, information about the image quality has to be considered for *in situ* track planning of the AUV. This guarantees the collection of sufficiently good image data for the entire mission area.

The proposed strategy directly operates in the image domain that is in contrast to [Williams et al. (2012), Hansen et al. (2011)] where image quality is related to the peak correlation between successive pings, and therefore, operates in the raw echo data domain. An evaluation based on the theoretical resolution becomes possible using the proposed scheme, which is not the case for the energy ratio proposed in [Debes et al. (2009), Leier and Zoubir (2011)]. Since defocusing occurs successively when coherently combining single RAS images, the proposed approach aims at sequentially assessing the focusing capability by estimating the resolution from the scene of interest for consecutive pings. The material and results presented in this section have been mainly published in [Leier et al. (2013)].

5.3.1 Sequential assessment scheme

In the following, an expression for the instantaneous cross-range resolution is introduced. As stated in Chapter 2, the angular beamwidth of a planar antenna, which is used as a single receiver element of a uniform linear array (ULA), with length D_{phy} is approximately given by $\theta_{\text{BW}} \approx \lambda / D_{\text{phy}}$. It mainly determines the maximum length of a synthetic aperture together with the focusing range, *i.e.*, $L_{\text{syn}}^{\max}(r) \equiv r\,\theta_{\text{BW}}$. As discussed in Section 2.2, the aperture length adapts with respect to the focusing range. Given the advance per ping (APP) of a single-transmitter and multi-receiver system by Δ^{INS}, the instantaneous length of the synthetic aperture can be expressed as a function of range r and ping index p as follows

$$L_{\text{syn}}(p, r) = \min\left[(p-1)\Delta^{\text{INS}} + L_{\text{phy}},\ L_{\text{syn}}^{\max}(r)\right]. \tag{5.7}$$

Here, the length of the ULA is denoted by L_{phy}. Moreover, the instantaneous synthetic aperture length $L_{\text{syn}}(p, r)$ assumes a constant surge motion and an absence of redundant phase center (RPC) pairs between consecutive pings in (5.7). Otherwise, it has to be adapted according to the surge estimation procedure outlined in Section 4.2.2. Replacing the maximum synthetic aperture length in the expression of the cross-range

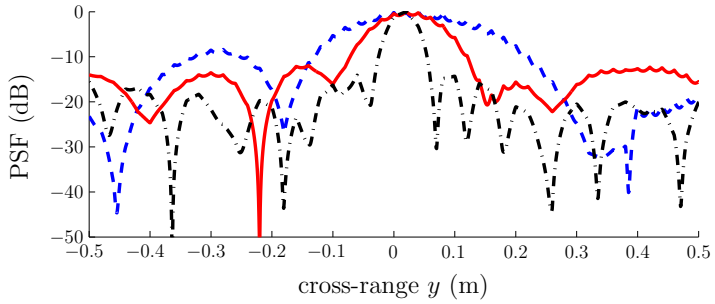


Figure 5.16: Changing mainlobe width of the point spread function (PSF) for an increasing synthetic aperture length of a point scatterer (blue dashed $M_p = 2$, red solid $M_p = 4$, black dash-dotted $M_p = 10$).

resolution with $L_{\text{syn}}(p, r)$ – see Section 2.2.1, leads to the instantaneous cross-range resolution of the synthetic aperture as a function of range and ping index as follows

$$\delta_{\text{syn}}(p, r) = \frac{r\,\lambda}{\min\left[(p-1)\Delta^{\text{INS}} + L_{\text{phy}},\ L_{\text{syn}}^{\max}(r)\right]}. \tag{5.8}$$

The range independent cross-range resolution $\delta_{\text{syn}} = D_{\text{phy}}/2$ of a synthetic aperture system as stated in (2.28) is achieved given the maximum synthetic aperture length $L_{\text{syn}}^{\max}(r)$, and assuming a physical aperture size of individual receivers of $D_{\text{phy}} \approx \lambda/\theta_{\text{BW}}$. An example of the instantaneous cross-resolution $\delta_{\text{syn}}(p, r_0)$ is depicted in Figure 5.16. It has been estimated by the changing mainlobe width of the point spread function (PSF) of an isolated point scatterer. Since the synthetic aperture length elongates with an increasing number of total pings M_p, an improved focusing capability is attained. This is clearly noticeable in Figure 5.16 by a narrowing mainlobe width.

Provided that an image reconstruction technique is employed, which formulates SAS imaging as a sequential processing of RAS images, quality assessment of the imaging process is achieved by sequentially comparing the theoretical instantaneous cross-range resolution $\delta_{\text{syn}}(p, r_0)$ to the estimated one, $\hat{\delta}_{\text{syn}}(p, r_0)$, of a point scatterer in the image scene. Clearly, this approach assumes the presence of an isolated point scatterer in the first RAS image with $p = p_0$ in order to be able to estimate its half-power mainlobe width Υ_{3dB}. The latter is then related to the instantaneous cross-range resolution as $\hat{\Upsilon}_{\text{3dB}} \equiv \hat{\delta}_{\text{syn}}(p_0, r_0)$. The subsequent RAS image is coherently added and the described process is repeated until the last ping p_1 is attained, for which the extracted point scatterer is still observed by the imaging platform. Typically, the extraction of a PSF requires some type of interpolation.

In the sequel, a model function $\delta_{\text{mod}}(p)$ is fitted to the estimated instantaneous cross-range resolution $\hat{\delta}_{\text{syn}}(p, r_0)$, with $p = p_0, \ldots, p_1$, for a fixed range bin r_0 assuming a

negligible range migration. The purpose of fitting the model function is to assess the image quality of the SAS image based on a parameter comparison that relates to the cross-range resolution over the entire synthetic aperture construction. To this end, the theoretical instantaneous cross-range resolution in (5.8) is expressed by an exponential model function in terms of ping index p and parameters β_i, with $i = 1, \ldots, 3$, as follows

$$\delta_{\mathrm{mod}}(p) = \beta_1 \exp\{-\beta_2\, p\} + \beta_3. \tag{5.9}$$

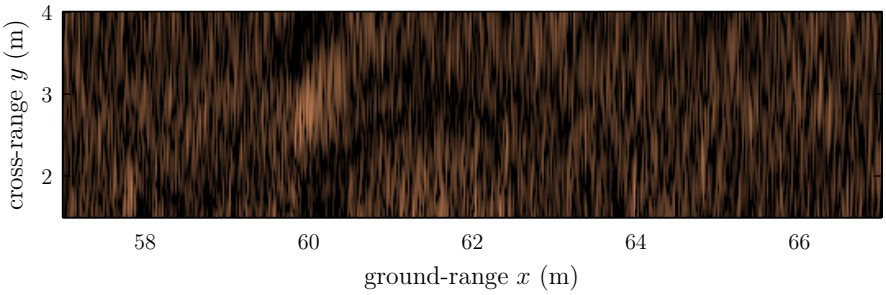
Here, β_3 describes the convergence parameter that approximates the theoretical SAS cross-range resolution δ_{syn}. In order to estimate the parameters $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]^T$, the cost function

$$C(p, \boldsymbol{\beta}) = \frac{1}{2} \left| \hat{\delta}_{\mathrm{syn}}(p, r_0) - \delta_{\mathrm{mod}}(p; \boldsymbol{\beta}) \right|^2 \tag{5.10}$$

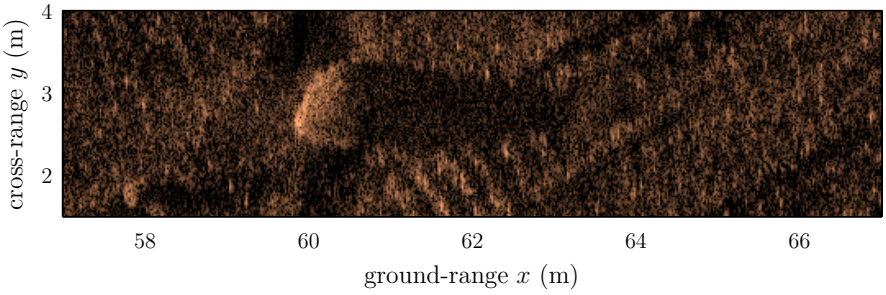
is minimized using, for example, the Newton-Raphson method as described in Section 4.2.6. The quality evaluation is then done by comparing the estimated parameter $\hat{\beta}_3$ to the theoretical convergence parameter β_3 that has been obtained by fitting $\delta_{\mathrm{mod}}(p)$ to the theoretical instantaneous cross-range resolution $\delta_{\mathrm{syn}}(p, r_0)$. Significant deviations between the parameters $\hat{\beta}_3$ and β_3 indicate the occurrence of defocus and a distortion in image quality. It should be remarked that the evaluation over the entire construction process of the synthetic aperture provides a more reliable assessment of image quality than only considering the final SAS image as it is classically done for synthetic aperture radar (SAR) images [Curlander and McDonough (1991)].

5.3.2 Real data results

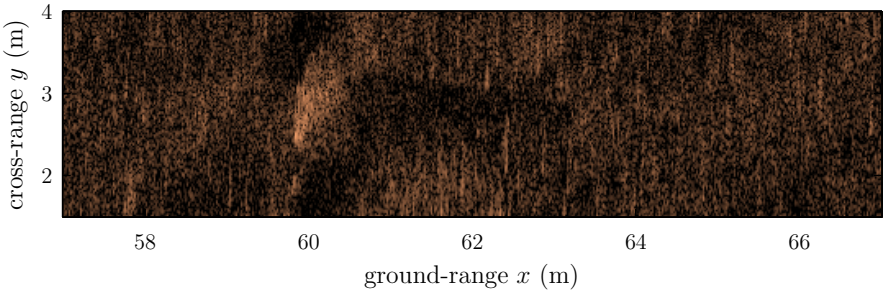
In this section, a real sonar data example based on an MCM-SLS data set is provided to assess the progression of the cross-range resolution. The underlying image scene is depicted in Figure 5.17 illustrating a RAS and a SAS image in Figure 5.17a and Figure 5.17b, respectively. Both images are displayed with a dynamic range of 40 dB. The SAS image shows a significant enhancement in the richness of detail due to an enhanced resolution, an improved shadow contour, and overall contrast. However, in the case of motion errors, the coherent summation of individual RAS images causes a blurring in the overall SAS image as demonstrated in Figure 5.17c. Here, a sinusoidal sway motion has been injected over all pings with an amplitude of $A^s = 0.3\lambda_c$ – see Section 4.1.2. The corresponding evaluation of the focus assessment scheme is depicted in Figure 5.18a and Figure 5.18b for perfect focusing and defocusing in the presence of uncompensated motion errors, respectively. A comparison between the theoretical $\delta_{\mathrm{syn}}(p, r_0)$ (blue cross) and estimated instantaneous cross-range resolution $\hat{\delta}_{\mathrm{syn}}(p, r_0)$



(a) RAS image



(b) Focused SAS image



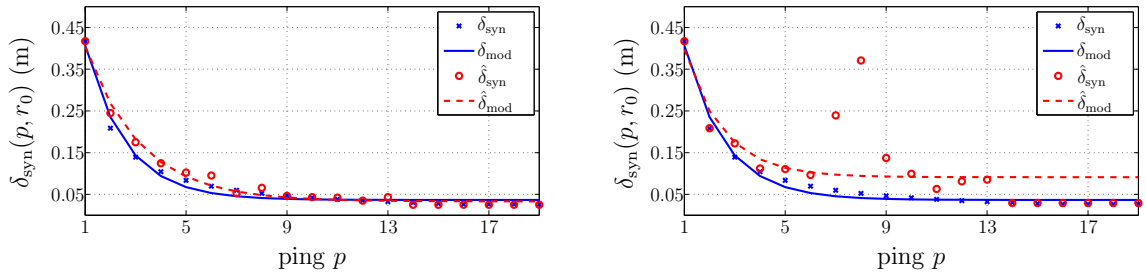
(c) Blurred SAS image

Figure 5.17: RAS image (a) and SAS image (b) comparison of a mine-like object. A significant detail gain is apparent for the SAS image. The SAS image in (c) is blurred due to uncompensated motion errors.

(red circle) is depicted in Figure 5.18a together with their fitted model functions. In Figure 5.18a, the estimated resolution converges against the theoretical SAS cross-range resolution with $\hat{\delta}_{\text{syn}} \approx \delta_{\text{syn}} \approx 0.026$ m. Additionally, the fit of both model functions shows a satisfying agreement in the case of an error-free motion scenario, with convergence parameters $\beta_3 \approx 0.036$ m and $\hat{\beta}_3 \approx 0.033$ m.

For the blurred SAS image scenario, the corresponding results of the instantaneous cross-range resolution are depicted in Figure 5.18b. Here, the estimated convergence parameter is given by $\hat{\beta}_3 \approx 0.092$ m, which is significantly larger than the convergence parameter of the theoretical cross-range resolution with $\beta_3 \approx 0.036$ m, indicating a defocus in the SAS image. The importance of fitting a model and considering the

entire cross-range resolution history along the construction of the synthetic aperture becomes apparent for the estimated cross-range resolution values starting with ping $p \geq 14$ in Figure 5.18b. Here, the deterioration of the PSF suggests a perfectly focused SAS image. Hence, the advantage of a successive evaluation is twofold: First, it allows for exploiting the focusing history to judge the quality rather than only assessing the final SAS image. The latter may suggest a wrong interpretation of the focusing as certain degradations still lead to a narrow half-power mainlobe width. Second, the number of pings, and therefore, a synthetic sub-aperture length can be determined to construct an image with possibly a lower resolution but without degradations due to motion errors. However, this would require a tracking method to sequentially estimate the resolution parameter of the model function. In the provided example the isolated point scatterers have been selected manually to estimate the PSF so as to validate the proposed idea of sequentially assessing the focusing capability of a synthetic aperture. Nevertheless, following the approach in [Glover and Campell (2010)], an automatic scheme for estimating the resolution based on the PSF can be developed.



(a) Metric evaluation for perfect focusing

(b) Metric evaluation under image defocusing

Figure 5.18: Theoretical and estimated metric evaluation over an increasing synthetic aperture length in case of perfect focusing (a) and defocusing (b) due to injected motion errors. Fitted curves of the model function are depicted as solid and dashed lines.

5.4 Conclusion

In this chapter, the impact of residual sinusoidal motion errors and sound speed mismatches on image quality, segmentation, feature extraction, and classification performance of an ADAC system for mine hunting applications have been investigated. Although all empirical studies show a noisy behavior due to the limited size of the available data set, a clearly visible trend of rising misclassification rates for increasing strength of all error types has been observed. A similar trend is noticeable for the curves of the segmentation ratio. A dependency between the degradation in segmentation

and the resulting change in the extracted feature values due to the distortion of image quality has been demonstrated. While the ICM method has shown better segmentation performance of the shadow region under the influence of motion errors, this result has not been confirmed for the misclassification rates. Additionally, using the combination of the considered feature set and the accompanying classifier of the employed ADAC system, it has been illustrated that already a small sway motion and mismatches in the measured sound speed have a severe influence on classification. Although it is difficult to generalize these observations for different target recognition systems, the importance of studying the impact of defocusing effects in SAS images is highlighted with respect to target recognition tasks such as segmentation, feature extraction, and classification to develop more reliable ADAC systems in the future. Similar effects have been observed for larger heave motion amplitudes, which is due to the common geometry of an AUV based SAS system. Such a strong heave motion is unlikely to result from an inaccurate INS, however, missing information about the seafloor topography may yield a comparable image defocusing. This demonstrates the importance of an accurate bathymetry estimation procedure for ATR purposes besides the possibility of extracting additional features about the object.

An empirical relation has been shown between the drop in image quality and the resulting impact on the subsequent ADAC tasks. While the misclassification rates have been significantly improved by training the used ADAC system with images affected by motion errors, performance degrades again using such training data for well focused images. This highlights that adapted segmentation methods, which are robust to motion errors, are of great interest for ADAC systems to guarantee reliable feature extraction. Moreover, features should be selected not only for the class separability but also for their robustness against degraded imagery. In addition to robust methods, detecting situations of occurring defocus by means of image quality assessment is of great importance in the context of automatic classification systems for mine hunting to achieve reliable classification performance in normal and difficult scenarios. Thus, a sequential focus assessment scheme has been proposed to detect the degradation in image quality by probing the instantaneous cross-range resolution. The continuous monitoring of image quality may allow for adaptive mission planning and for finding the optimum SAS aperture. This allows for the reconstruction of the highest quality SAS image under the given circumstances. However, the assessment scheme is not yet an integral part of the SAS processing chain as it still requires the implementation of an automatic selection technique to extract the PSF of isolated point targets. Further, a test procedure has to be developed to automatically detect defocusing situations.

Chapter 6

Synthetic aperture imaging based on compressive sensing

Current synthetic aperture systems produce such a large amount of data during a few hours of operation that issues with respect to data storage, data transportation and data processing [Bhattacharya et al. (2008)] are not uncommon. Additionally, the advance per ping (APP) of an imaging platform is dictated by the spatial sampling theorem [Jakowatz et al. (1996), Soumekh (1999)] – see also Section 2.3. A violation of the spatial sampling requirement yields azimuth image ambiguities also called grating lobes or ghost targets in the reconstructed image. The latter may mask important image content such as objects, which are lost beyond recall, or may destroy shadow information that is important for automatic target recognition (ATR). Moreover, the APP influences the coverage rate of the imaging system, and hence, determines the mission time. Consequently, alternative imaging techniques are of utmost interest so as to avoid the massive amount of data collection, and to suppress azimuth image ambiguities while reducing the mission time.

The emerging field of compressive sensing (CS) introduces a novel sampling framework [Donoho (2006), Candès et al. (2006)], which allows for sub-Nyquist sampling. Simultaneously, CS enables an alias-free signal reconstruction by finding a solution to an underdetermined linear system. In this chapter, a stripmap synthetic aperture imaging technique based on the CS framework is developed to increase the speed of an imaging platform while maintaining the quality of the reconstructed image [Leier and Zoubir (2014)].

The chapter is organized as follows: Section 6.1 provides a brief overview of state-of-the-art CS research for radar imaging applications. Section 6.2 introduces a vector-matrix notation of the synthetic aperture signal model, followed by a description of how conventional imaging works given such a model. The proposed CS imaging technique is then addressed and possible undersampling schemes are introduced. In Section 6.3, results on synthetic and experimental data are provided. The latter are recorded using an ultrasonic synthetic aperture system, which has been set up for validation purposes. Finally, the results are discussed in Section 6.4.

6.1 Introduction and state-of-the-art

Compressive sensing enables sampling rates that are significantly lower than the Nyquist rate if a captured signal has a sparse representation in some domain. For example, the image domain can be sparse considering a few man-made objects lying on the seafloor. In addition to this sparsity requirement, an incoherence criterion has to be fulfilled between measurement and sparsity domain. Roughly speaking, the criterion states that the measurement and sparsity domains have to be highly uncorrelated [Candès and Wakin (2008)]. If both requirements are fulfilled, CS processing can be successfully applied for reconstructing undersampled signals. Feasible applications cover diverse areas. Among others, CS has been successfully applied in the context of digital imaging [Duarte et al. (2008)], medical scanners [Lustig et al. (2008)], in various radar [Herman and Strohmer (2009), Ender (2010)] and synthetic aperture radar (SAR) applications [Baraniuk and Steeghs (2007), Stojanovic et al. (2009), Tello Alonso et al. (2010), Patel et al. (2010)], which are briefly discussed subsequently to point out their difference to the proposed imaging approach.

In [Baraniuk and Steeghs (2007)], the need for a matched-filter operation for focusing the echo signals in range direction is avoided using CS. Simultaneously, the sampling rate is reduced. While the use of a specially designed waveform is suggested in [Herman and Strohmer (2009)] to design a high-resolution CS radar, the author in [Ender (2010)] chooses a stepped-frequency signal model. CS is then applied to reduce the recording time due to the sequential transmission of numerous monochromatic signals in the application of radar pulse compression. Similarly, a CS stepped-frequency approach is suggested in [Yang et al. (2013)] in the context of spotlight SAR to decrease the recording time and data storage requirements. Contrarily, the authors in [Stojanovic et al. (2009), Tello Alonso et al. (2010), Patel et al. (2010)] use the common linear frequency modulated (LFM) pulse sequence for CS-based spotlight SAR imaging. Promising results have been achieved for both synthetic data and real radar data. However, in [Stojanovic et al. (2009), Tello Alonso et al. (2010)], narrowband and far-field assumptions are applied, and thus, range migration [Soumekh (1999), Richards (2005)], which is of major concern in stripmap sonar imaging systems, is not considered. Especially, the assumption of two separate 1-D processing steps does not hold for near-field situations and wideband systems, which are typically given for sonar applications. In [Patel et al. (2010)], CS-based synthetic aperture undersampling is motivated to reduce data storage and to obtain a wider swath width for spotlight SAR assuming a tomographic formulation [Jakowatz et al. (1996), Carrara et al. (1995)].

6.2 Compressive sensing based imaging

In this section, the synthetic aperture signal model is rewritten into a matrix-vector notation in order to represent the reconstructed image as a solution to a linear equation system. Afterwards, conventional synthetic aperture imaging and CS-based imaging is addressed. Furthermore, two different undersampling schemes are described. While the first scheme allows for randomized fast-time undersampling mainly to reduce data storage, the second scheme performs undersampling additionally in along-track direction. By regularly skipping transmission times, an increase in the speed of the imaging platform can be achieved.

6.2.1 Data model in matrix-vector notation

Given the echo data model in (2.7), an echo signal matrix representation of a transceiver based synthetic aperture system can be written as follows

$$\boldsymbol{E} = e_p^{\mathrm{LP}}(n)\Big|_{p=0,\ldots,M_p-1,\,\text{and}\,n=0,\ldots,M_n-1,} \tag{6.1}$$

where $\boldsymbol{E}$ describes the raw echo signal matrix of size $M_p \times M_n$ in the equivalent lowpass domain before pulse compression. In the sequel, it is addressed how to rewrite this matrix representation into a system of linear equations [Gunther et al. (2011)] given by

$$\underbrace{\boldsymbol{e}}_{M_p M_n \times 1} = \underbrace{\boldsymbol{S}}_{M_p M_n \times D}\,\underbrace{\boldsymbol{\sigma}}_{D\times 1} + \underbrace{\boldsymbol{v}}_{M_p M_n \times 1}, \tag{6.2}$$

such that $\boldsymbol{e} = \mathrm{vec}\{\boldsymbol{E}^T\}$, where $\mathrm{vec}\{\cdot\}$ denotes the operator to vectorize a matrix column after column. In (6.2), the vector $\boldsymbol{v}$ describes additive sensor noise and the vector $\boldsymbol{\sigma}$ contains the reflectivity of all D targets, *i.e.*, $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_D]^T$. The target reflectivity vector $\boldsymbol{\sigma}$ is multiplied with the pulse system matrix

$$\boldsymbol{S} = \left[\ \boldsymbol{S}_0,\, \boldsymbol{S}_1,\, \ldots,\, \boldsymbol{S}_p,\, \ldots,\, \boldsymbol{S}_{M_p-1}\ \right]^T, \tag{6.3}$$

which is a stacked matrix containing individual pulse matrices $\boldsymbol{S}_p$, $p = 0, \ldots, M_p - 1$. An individual pulse matrix describes the echo signals of all D targets received at the transceiver position $\boldsymbol{a}_p$ during ping p. It can be expressed as

$$\boldsymbol{S}_p = \left[\, \boldsymbol{s}_{1,p},\, \boldsymbol{s}_{2,p},\, \ldots\,,\, \boldsymbol{s}_{d,p},\, \ldots\,,\, \boldsymbol{s}_{D,p}\,\right], \tag{6.4}$$

where $\boldsymbol{s}_{d,p}$ is the received signal in terms of a delayed version of the transmitted pulse, which has been reflected by target d. The received signal in vector form is given by

$$\boldsymbol{s}_{d,p} = \left[\mathbf{0}_{1\times M_{\eta(d,p)}},\, \boldsymbol{b}_{d,p} \odot \boldsymbol{s},\, \mathbf{0}_{1\times M'_{\eta(d,p)}}\right]^T, \tag{6.5}$$

where $\boldsymbol{s} = [s(0), s(1), \ldots, s(M_\Pi - 1)]$ represents the transmitted waveform vector, here, an LFM pulse with a length of M_Π samples, $\boldsymbol{b}_{d,p}$ denotes an ideal beampattern vector, and $\odot$ is the Hadamard product. Similar to (2.5), the ideal beampattern vector $\boldsymbol{b}_{d,p}$ contains binary entries depending on whether or not target d is observed from the physical aperture during ping p. Its total size is $1 \times M_\Pi$. Moreover, in (6.5), $\mathbf{0}_{1\times M}$ denotes a zero row vector of size M. The position index of the transmitted pulse vector $\boldsymbol{s}$ within the vector $\boldsymbol{s}_{d,p}$ of the received signal depends on the number of samples of the round-trip delay. The latter can be expressed as

$$M_{\eta(d,p)} = \left\lfloor \frac{2\left\|\boldsymbol{q}_d - \boldsymbol{a}_p\right\|_2}{c\,T_s} \right\rfloor, \tag{6.6}$$

where $\lfloor\cdot\rfloor$ rounds towards the next smaller integer value, and c and T_s denote the wave propagation speed and the sampling interval, respectively. Since a total number of M_n fast-time samples is recorded, the vector $\boldsymbol{s}_{d,p}$ must be zero-padded with $M'_{\eta(d,p)} = M_n - (M_{\eta(d,p)} + M_\Pi)$ trailing zeros. Typically, the number of fast-time samples M_n depends on the maximum slant-range distance $R_{\max}$ of the synthetic aperture system [Soumekh (1999)].

6.2.2 Conventional focusing

In order to apply a time-domain correlation (TDC) imaging method [Soumekh (1999)] using vector-matrix manipulations, a focusing matrix $\boldsymbol{G}$ is required. It relates the target area to the received echo signals and is identical to the pulse matrix $\boldsymbol{S}$ in (6.3) except that it covers the entire grid $\boldsymbol{g}_{kl} = [x_k, y_l, 0]^T$ of the discretized target scene, for $k = 1, \ldots, M_x$, and $l = 1, \ldots, M_y$, rather than only target coordinates $\boldsymbol{q}_d$, with $d = 1, \ldots, D$. Thus, the focusing matrix $\boldsymbol{G} \in \mathbb{C}^{M_p M_n \times M_y M_x}$ is a stacked matrix of ping-based focusing matrices $\boldsymbol{G}_p \in \mathbb{C}^{M_n \times M_y M_x}$ with $p = 0, \ldots, M_p - 1$ that are similar to (6.4) and given by

$$\boldsymbol{G}_p = \left[\boldsymbol{s}_{p,1,1},\, \ldots,\, \boldsymbol{s}_{p,1,M_y},\, \ldots,\, \boldsymbol{s}_{p,k,l},\, \ldots,\, \boldsymbol{s}_{p,M_x,M_y}\right]. \tag{6.7}$$

The ping-based focusing matrix $\boldsymbol{G}_p$ describes the mapping between all grid points $\boldsymbol{g}_{kl}$ and the transceiver position $\boldsymbol{a}_p$. The position index of the received pulse $\boldsymbol{s}_{p,k,l}$ within each column depends on the number of samples of the round-trip delay in (6.6)

substituting the target coordinate $\boldsymbol{q}_d$ by the grid point location $\boldsymbol{g}_{kl}$. After discretizing the ideal target area $f(x, y)$ in continuous-space into matrix-form as

$$\boldsymbol{F} = \begin{bmatrix} \sigma_{x_1, y_1} & \sigma_{x_2, y_1} & \dots & \sigma_{x_{M_x}, y_1} \\ \sigma_{x_1, y_2} & \sigma_{x_2, y_2} & \dots & \sigma_{x_{M_x}, y_2} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{x_1, y_{M_y}} & \sigma_{x_2, y_{M_y}} & \dots & \sigma_{x_{M_x}, y_{M_y}} \end{bmatrix}, \tag{6.8}$$

where each element of the matrix represents the reflectivity of the corresponding grid point $\boldsymbol{g}_{kl}$, the data model of the reconstruction can be found subsequently. As in (6.2), it can be denoted in vector-matrix notation by

$$\boldsymbol{e} = \boldsymbol{G}\boldsymbol{f} + \boldsymbol{v}, \tag{6.9}$$

where $\boldsymbol{f} = \mathrm{vec}\{\boldsymbol{F}\}$. The TDC imaging technique can then be formulated using vector-matrix operations [Gunther et al. (2011)] as follows

$$\hat{\boldsymbol{f}} = \boldsymbol{G}^H \boldsymbol{e}, \tag{6.10}$$

in order to estimate the reflectivity of the target scene $\hat{\boldsymbol{f}}$, where $(\cdot)^H$ denotes the Hermitian transpose. Here, $\hat{\boldsymbol{f}}$ is a stacked vector representing the target scene, which has to be reshaped to obtain the reconstructed image, *i.e.*, $\hat{\boldsymbol{F}} = \mathrm{vec}^{-1}\{\hat{\boldsymbol{f}}\}$ with $\mathrm{vec}^{-1}\{\cdot\}$ being the inverse reshaping operation yielding a matrix given a stacked vector. While this space-time imaging method approach is not very efficient in terms of computational complexity, it does not use any approximations to solve the inverse reconstruction problem; simultaneously, it enables the use of arbitrary path deviations similar to the backprojection algorithm introduced in Section 3.3. However, unknown path deviations are not considered in the context of CS-based synthetic aperture imaging in this thesis.

In the case of a single-transmitter and multi-receiver synthetic aperture system, the focusing matrix $\boldsymbol{G}$ has to be rewritten. For each single-receiver element $u = 1, \dots, N_{\mathrm{rx}}$, the sample round-trip delay of the transceiver model in (6.6) is substituted by its equivalent delay of a transmitter-receiver pair. Thereupon, the focusing matrix $\boldsymbol{G}$ is replaced by its single-receiver counterpart $\boldsymbol{G}^u$ in (6.10), which leads to the reconstruction of a single receiver image $\hat{\boldsymbol{f}}_u$ for receiver u. The coherent sum of all single-receiver images given by

$$\hat{\boldsymbol{f}} = \sum_{u=1}^{N_{\mathrm{rx}}} \hat{\boldsymbol{f}}_u \tag{6.11}$$

leads to the reconstruction of the synthetic aperture image $\hat{\boldsymbol{f}}$. Note that azimuth image ambiguities occurring in the single-receiver images $\hat{\boldsymbol{f}}_u$ are cancelled out by the coherent summation.

6.2.3 Focusing using compressive sensing

In the sequel, the basics of CS theory are introduced in the context of synthetic aperture imaging. Furthermore, it is outlined how CS can be used for image reconstruction. CS allows for sensing a signal in a low-dimensional form. However, for this purpose, the captured signal is required to be sparse in a certain domain [Donoho (2006), Candès et al. (2006)]. Consider the signal $\boldsymbol{f} = [\sigma_1, \ldots, \sigma_{M_{xy}}]^T$ with $M_{xy} = M_x M_y$ that can be sparsely represented by means of the transformation matrix $\boldsymbol{\Psi}$, which describes an $M_{xy} \times M_{xy}$ unitary basis such that

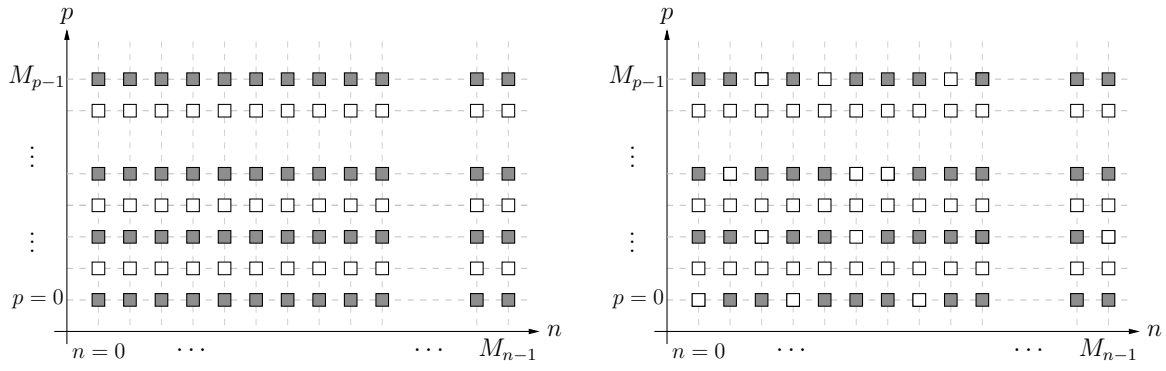
$$\boldsymbol{f} = \boldsymbol{\Psi}^H \boldsymbol{\chi}, \tag{6.12}$$

where $\boldsymbol{\chi} = \left[\chi_1, \ldots, \chi_{M_{xy}}\right]^T$ denotes the K-sparse coefficient vector. The K-sparse property states that only K coefficients are unequal to zero with $K \ll M_{xy}$ [Candès and Wakin (2008)]. For $\boldsymbol{\Psi} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, the sparse coefficient vector $\boldsymbol{\chi}$ equals the signal $\boldsymbol{f}$. This means that the signal $\boldsymbol{f}$ can be assumed to be sparse, for example, a few point-like objects lying on the seafloor. Instead of measuring the echo signal vector $\boldsymbol{e}$ of (6.9), which is of size $M = M_p M_n$, CS aims at reducing the measurements to $M' < M$. Since the reconstruction of $\boldsymbol{f}$ is of interest, an underdetermined system of linear equations has to be solved. The latter is feasible and yields an unique solution given the sparsity of $\boldsymbol{f}$. In order to undersample the received echo signals $\boldsymbol{e}$, a selection matrix $\boldsymbol{\Sigma}$ is multiplied by the reconstruction model of (6.9) leading to

$$\boldsymbol{e}^{\mathrm{cs}} = \boldsymbol{\Sigma}\boldsymbol{e} = \boldsymbol{\Sigma}(\boldsymbol{G}\boldsymbol{f} + \boldsymbol{v}) = \boldsymbol{\Sigma}\boldsymbol{G}\boldsymbol{f} + \tilde{\boldsymbol{v}}, \tag{6.13}$$

where $\boldsymbol{e}^{\mathrm{cs}}$ denotes the spatially and/or temporally undersampled vector of raw echo signals of size M'. Note that the selection matrix $\boldsymbol{\Sigma}$ is a fat matrix of dimension $M' \times M$. It resembles an identity matrix with deleted rows for spatially undersampled along-track (cross-range) positions.

In the following, two basic undersampling schemes are introduced, namely, a regular along-track sampling scheme as well as a regular along-track and random range sampling scheme similar to [Yoon and Amin (2008)]. Typically, CS shows best performance for random downsampling matrices [Baraniuk (2007), Baraniuk et al. (2008)]. By contrast, a purely random sampling in along-track direction without skipping entire spatial sampling positions does not lead to an improvement in coverage rates but only to a reduced amount of data. The two schemes are illustrated in Figure 6.1. Both subplots show a matrix of slow-time and fast-time samples with gray and white boxes, the latter meaning that corresponding samples



(a) regular in along-track direction

(b) regular in along-track, random in range

Figure 6.1: Schemes for spatially undersampling the synthetic aperture in along-track direction (a) and additionally selecting echo samples randomly in range direction (b).

have been dropped. In the case of regularly undersampling the synthetic aperture in Figure 6.1a, every other slow-time position $\boldsymbol{a}_p$ is dropped, which is denoted by $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{A}}_{\max}$. This means that the actual sampling interval is twice as long as required by the sampling theorem, and therefore, the platform speed can be increased by the same factor. In other words, the selection matrix $\boldsymbol{\Sigma}$ resembles an identity matrix of size M, where every second row is deleted. Thus, the actual dimension of the selection matrix $\boldsymbol{\Sigma}$ is given by $M' \times M$ with $M' = 0.5M$.

Figure 6.1b shows how, additionally, the range direction is randomly undersampled by dropping fast-time samples with a pre-defined ratio ϱ_{nom} (here, $\varrho_{\mathrm{nom}} = 0.25$). This scheme is an extension to the undersampling scheme of the along-track direction that leads to further storage capacity savings. However, compared to the first scheme, it requires some changes in the data acquisition hardware of the imaging system. Moreover, measurement reduction is not achieved any longer by using a matrix multiplication with the selection matrix $\boldsymbol{\Sigma}$. Instead, an element-wise reduction operation can be considered in (6.13) that depends on the binary value of the undersampling scheme as depicted in Figure 6.1b.

CS image reconstruction is addressed subsequently. Due to measurement noise, the reconstruction of the target scene is formulated as a basis pursuit denoising (BPDN) [Chen et al. (2001)] optimization problem as follows

$$\hat{\boldsymbol{f}}^{\mathrm{cs}} = \arg\min_{\boldsymbol{f}} \left\{ \left\| \boldsymbol{e}^{\mathrm{cs}} - \boldsymbol{\Sigma}\boldsymbol{G}\boldsymbol{f} \right\|_2^2 + \Lambda^{\mathrm{cs}} \left\| \boldsymbol{f} \right\|_1 \right\}, \tag{6.14}$$

which can be solved using, *e.g.*, the SpaRSA algorithm [Wright et al. (2009)] that is capable of dealing directly with complex data. Here, Λ^{cs} represents the regularization parameter of the optimization problem. Again, the reconstruction result is a stacked

vector $\hat{\boldsymbol{f}}^{\mathrm{cs}}$, which has to be reshaped to obtain an image of the target scene $\hat{\boldsymbol{F}}^{\mathrm{cs}}$. Similar to conventional imaging, the focusing matrix $\boldsymbol{G}$ can be substituted by its single-receiver counterpart $\boldsymbol{G}^u$ in (6.14) to obtain an aliased single-receiver image $\hat{\boldsymbol{f}}_u^{\mathrm{cs}}$. A coherent summation of the individual CS images $\hat{\boldsymbol{f}}_u^{\mathrm{cs}}$, with $u = 1, \ldots, N_{\mathrm{rx}}$, then leads to the synthetic aperture image $\hat{\boldsymbol{f}}^{\mathrm{cs}}$ of a single-transmitter and multi-receiver system, similar to (6.11). Alternatively, the overall focusing matrix $\boldsymbol{G}$ could be constructed by stacking the individual receiver focusing matrices $\boldsymbol{G}^u$ and solving the optimization problem of (6.14) using the complete data model. On the one hand, this may lead to better imaging results due to even a sparser content of the reconstructed scene. On the other hand, however, this approach increases the computational complexity due to a larger size of the stacked focusing matrix $\boldsymbol{G}$. Thus, computational complexity is traded off against imaging performance.

6.3 Experiments

In this section, the proposed CS imaging technique is validated on synthetic data as well as as on real acoustical measurements. Synthetic data simulations are performed to determine the amount of measurement reduction as a function of image quality distortion using a full-reference image quality metric. In the real measurement examples, it is demonstrated that the speed of the imaging platform can be doubled for the used laboratory synthetic aperture system.

6.3.1 Synthetic data results

First of all, the general ability of CS imaging to suppress azimuth image ambiguities is exemplarily demonstrated. To this end, synthetic data is generated from a scenery consisting of three homogeneous point targets. The corresponding system parameters for generating the synthetic echo signals are listed in Table 6.1, and have been chosen in accordance to the system parameters of the ultrasonic laboratory system. As the laboratory system demodulates the received echo signals in the digital domain, the sampling rate highly oversamples the lowpass echo signals. However, the echo signals are then downsampled to the Nyquist rate for the subsequent processing steps.

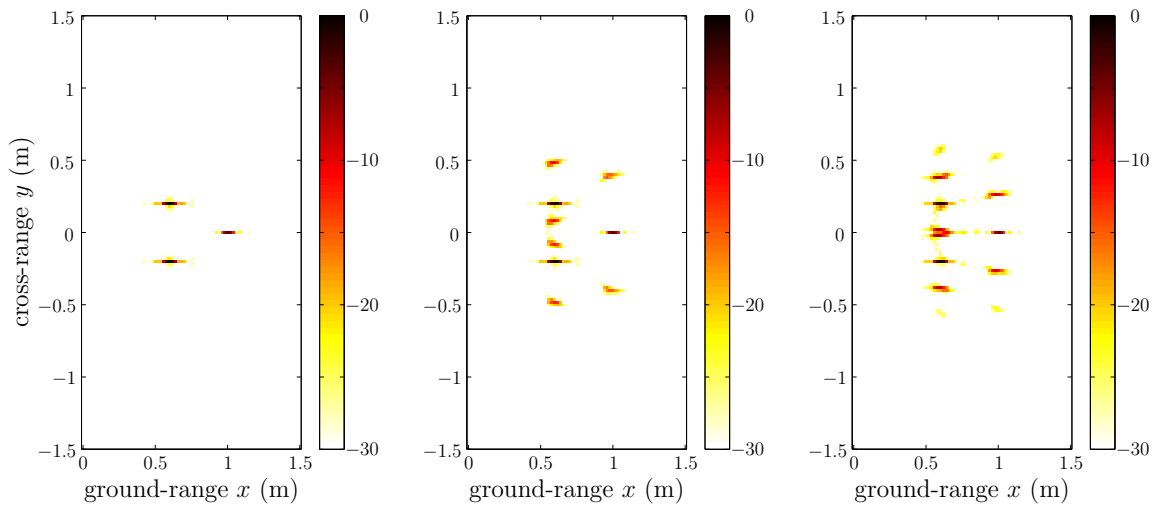
The corresponding reconstruction results of the conventional space-time imaging method and the proposed CS method are depicted in Figure 6.2 and Figure 6.3, respectively. In all examples, the dynamic range is set to 30 dB. For both methods,

Table 6.1: Synthetic aperture system parameters

Description	Variable	Value
Sampling frequency	f_s	100 kHz
Carrier frequency	f_c	40 kHz
Bandwidth	f_B	4 kHz
Pulse length	T_Π	4 ms
Beamwidth	θ_{BW}	20 deg
Advance per ping	$\Delta^{\mathrm{A}}_{\mathrm{max}}$	6 mm
Sound speed	c	340 m/s

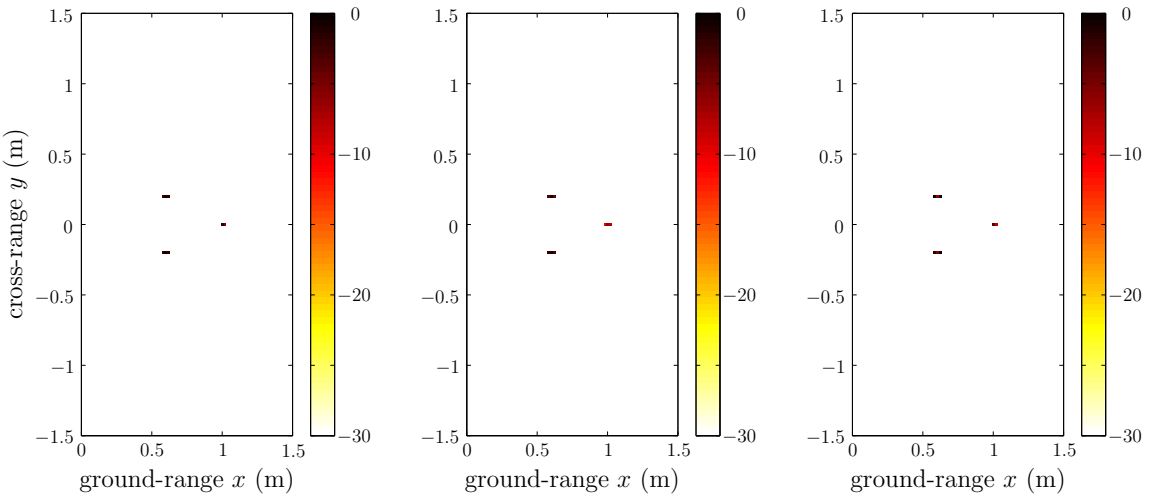
the subplots (a)-(c) illustrate the reconstructed images with an increased spatial undersampling of the synthetic aperture as highlighted by the undersampling scheme in Figure 6.1. In the case of CS reconstruction, an additional nominal drop-rate of $\varrho_{\mathrm{nom}} = 0.7$ and $\varrho_{\mathrm{nom}} = 0.8$ has been chosen in the fast-time domain in Figure 6.3b and Figure 6.3c, respectively.

The occurrence of symmetric azimuth image ambiguities due to regular undersampling is apparent for Figure 6.2b and Figure 6.2c. By contrast, the CS method is capable of suppressing the azimuth ambiguities in all three cases. Moreover, there is no noticeable difference in the quality of the reconstruction of the individual point scatterers despite an increased undersampling ratio ϱ. The regularization



(a) $\Delta^{\mathrm{A}} = \Delta^{\mathrm{A}}_{\mathrm{max}}$ (b) $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{A}}_{\mathrm{max}}$ (c) $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{A}}_{\mathrm{max}}$

Figure 6.2: Reconstruction results using the conventional TDC imaging technique and a regular along-track undersampling scheme with increasing values of the maximum advance per ping in (a)-(c).

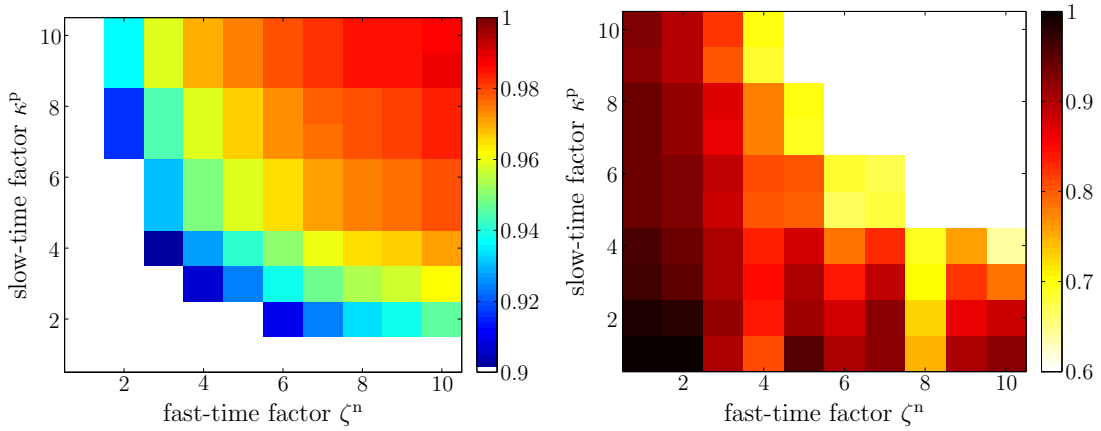


(a) $\Delta^{\mathrm{A}} = \Delta^{\mathrm{A}}_{\max}$, $\varrho_{\mathrm{nom}} = 0$ (b) $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{A}}_{\max}$, $\varrho_{\mathrm{nom}} = 0.7$ (c) $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{A}}_{\max}$, $\varrho_{\mathrm{nom}} = 0.8$

Figure 6.3: Reconstruction results using the proposed CS imaging technique and a regular along-track and random range undersampling scheme with increasing values of the maximum advance per ping in (a)-(c).

parameter, which is a trade-off measure between data fidelity and sparsity, has been empirically chosen [Figueiredo et al. (2007)] and set to $\Lambda^{\mathrm{cs}} = 0.3 \left\| (\boldsymbol{\Sigma G})^H \boldsymbol{e}^{\mathrm{cs}} \right\|_\infty$, where $\|\cdot\|_\infty$ denotes the maximum norm. The choice of the regularization parameter is a common problem in sparse reconstruction, for example, in direction-of-arrival estimation [Panahi and Viberg (2011)], and still subject of current research of imaging techniques.

In order to obtain a meaningful assessment of a maximum undersampling ratio for which the proposed CS imaging method still produces nearly identical reconstruction results, $N_{\mathrm{MC}} = 200$ Monte Carlo simulations have been conducted for different sets of undersampling ratios. Each set considers a factor κ^{p} with $\Delta^{\mathrm{A}} = \kappa^{\mathrm{p}} \Delta^{\mathrm{A}}_{\max}$ and a factor ζ^{n} with $\varrho_{\mathrm{nom}} = 1 - 1/\zeta^{\mathrm{n}}$, where ϱ_{nom} is the nominal drop-rate of M_n fast-time samples. Note that the superscripts $(\cdot)^{\mathrm{n}}$ and $(\cdot)^{\mathrm{p}}$ indicate the dimension. An average of the actual drop-rates $\hat{\varrho}(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}})$ is depicted in Figure 6.4a, where values with $\hat{\varrho}(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}}) < 0.9$ have been clipped. The actual drop-rates have been determined by thresholding the magnitude of the raw echo data, converting it to binary values, and counting the non-zero values before and after undersampling. In order to evaluate the image degradation due to azimuth image ambiguities as a consequence of undersampling, the structural similarity (SSIM) [Wang et al. (2004)] measure is applied. It compares an image under test, *i.e.*, the reconstructed CS image for undersampled echo data, with a high-quality full-reference image, where the latter is given by the CS image obtained using Nyquist sampling – see also Section 4.3.5. As for the SSIM measure Ξ, a value $\Xi = 1$ means



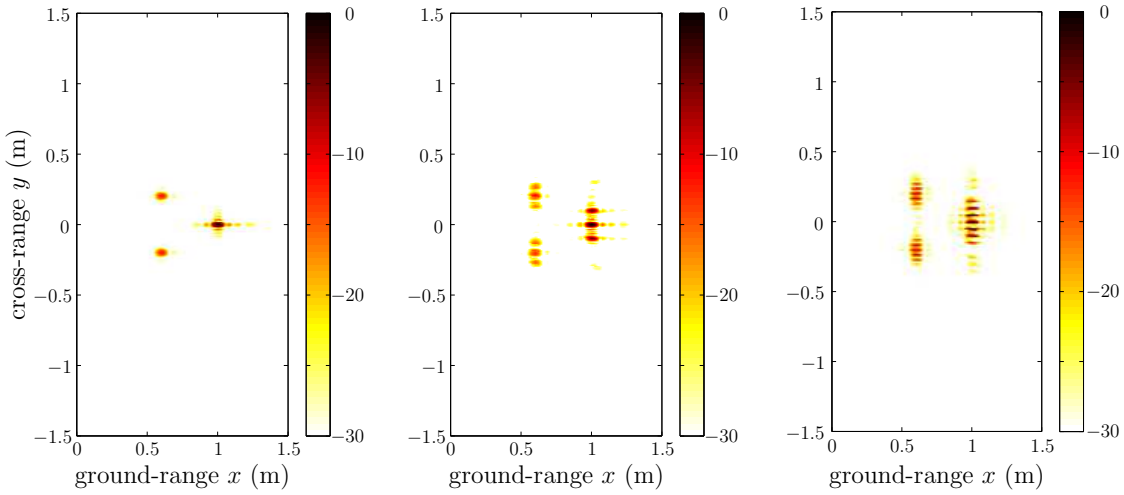
(a) Actual drop-rates $\hat{\varrho}(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}})$ (b) Structural similarity map $\Xi(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}})$

Figure 6.4: Actual drop-rates $\hat{\varrho}(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}})$ of the simulation results for CS image reconstruction are shown for different undersampling factors κ^{p} and ζ^{n} in (a) together with the structural similarity map $\Xi(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}})$ used as performance measure in (b).

that both images are identical and $\Xi = 0$ that there is no similarity. It should be noted that the homogeneous background of a sparse target scene influences the SSIM. This effect is reduced by downsizing the area under test to $x \in [0.4, 1.2]$ m in range and $y \in [-0.6, 0.6]$ m in along-track direction, where the latter boundary is determined by the occurrence of grating lobes in Figure 6.2c. The average simulation outcome of SSIM values $\Xi(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}})$, with varying undersampling factors κ^{p} and ζ^{n}, is illustrated in Figure 6.4b where values $\Xi(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}}) < 0.6$ are clipped. Moreover, images with a SSIM value of $\Xi(\kappa^{\mathrm{p}}, \zeta^{\mathrm{n}}) < 0.7$ have empirically been found to be affected by undersampled raw echo data. Consequently, relating the amount of discarded data in Figure 6.4a to the SSIM as a performance measure for successful CS reconstruction in Figure 6.4b, the simulation has revealed a data reduction of up to 95 %.

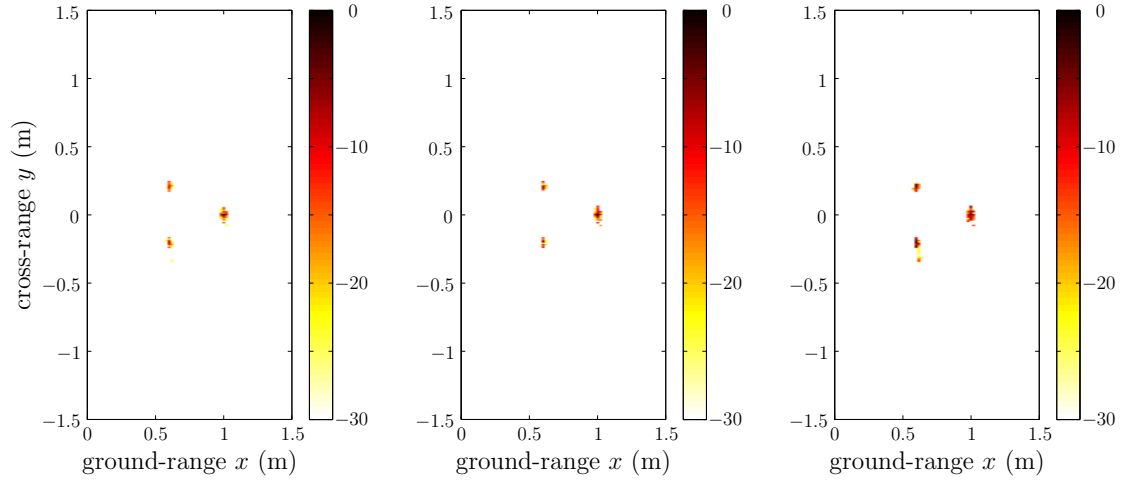
Before addressing real data examples, an extension of the proposed CS imaging technique is applied to a synthetic aperture system consisting of a uniform linear array (ULA) with $N_{\mathrm{rx}} = 4$ receiving elements. First, the spatial sampling interval Δ^{A} is set to the Nyquist limit $\Delta^{\mathrm{ULA}}_{\mathrm{max}}$, as stated in (2.35). The corresponding image in Figure 6.5a shows the three point targets of the same target scene as in Figure 6.2a. In contrast to Figure 6.5a, azimuth image ambiguities are noticeable in Figure 6.5b and Figure 6.5c. Here, the advance per ping has been set to $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{ULA}}_{\mathrm{max}}$ and $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{ULA}}_{\mathrm{max}}$, respectively. In the following, the proposed CS imaging technique is applied to each receiver element u to obtain a single-receiver image $\hat{\boldsymbol{f}}^{\mathrm{cs}}_{u}$. The coherent combination of these individual reconstruction results leads to the images in Figure 6.6a to Figure 6.6c. While Figure 6.6b with $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{ULA}}_{\mathrm{max}}$ shows a reconstruction result almost identical to Figure 6.6a, increasing the undersampling factor to three causes a small spread of

the point spread function (PSF) of the target at along-track position $y = -0.2$ m as depicted in Figure 6.6c. However, azimuth image ambiguities are also successfully suppressed for the multi-receiver configuration.



(a) $\Delta^{\mathrm{A}} = \Delta^{\mathrm{ULA}}_{\mathrm{max}}$ (b) $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{ULA}}_{\mathrm{max}}$ (c) $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{ULA}}_{\mathrm{max}}$

Figure 6.5: Reconstructed images of the conventional TDC imaging technique for a single-transmitter and multi-receiver system and a regular along-track undersampling scheme with increasing values (a)-(c) of the maximum advance per ping.

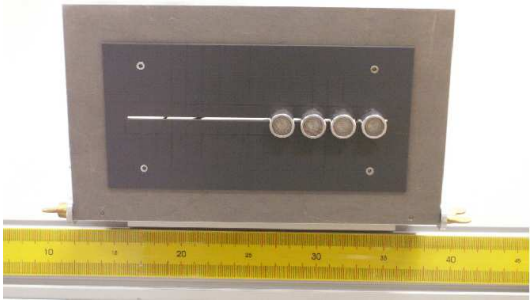


(a) $\Delta^{\mathrm{A}} = \Delta^{\mathrm{ULA}}_{\mathrm{max}}$ (b) $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{ULA}}_{\mathrm{max}}$ (c) $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{ULA}}_{\mathrm{max}}$

Figure 6.6: Reconstructed images of the proposed CS imaging method and a regular along-track scheme with increasing values (a)-(c) of the maximum advance per ping.

6.3.2 Experimental laboratory system

In order to record real acoustical measurements, a laboratory synthetic aperture system based on ultrasound sensors has been set up. Before discussing the corresponding imaging results, the laboratory system is briefly introduced. It is based on a single-transmitter and multi-receiver configuration, which is operated as a stripmap synthetic aperture system using ultrasound, similar to the system described in [Vincent et al. (2007)]. However, due to a non-calibrated array, the system is only employed as a bi-static transmitter-receiver system. Photographs of the laboratory setup are shown in Figure 6.7, where the transmitter is the rightmost sensor element of the imaging platform in Figure 6.7a. It sends LFM pulses corresponding to the signal parameters specified in Table 6.1. To the left side of the transmitter, three equally spaced receivers are mounted onto the mobile platform. The platform moves, driven by a motor with an approximately constant speed v, along a rail as shown in Figure 6.7b.



(a) Ultrasonic transmitter (right) and three receivers (left) mounted onto the imaging platform

(b) Motor, rail and imaging platform

(c) Real target scenario

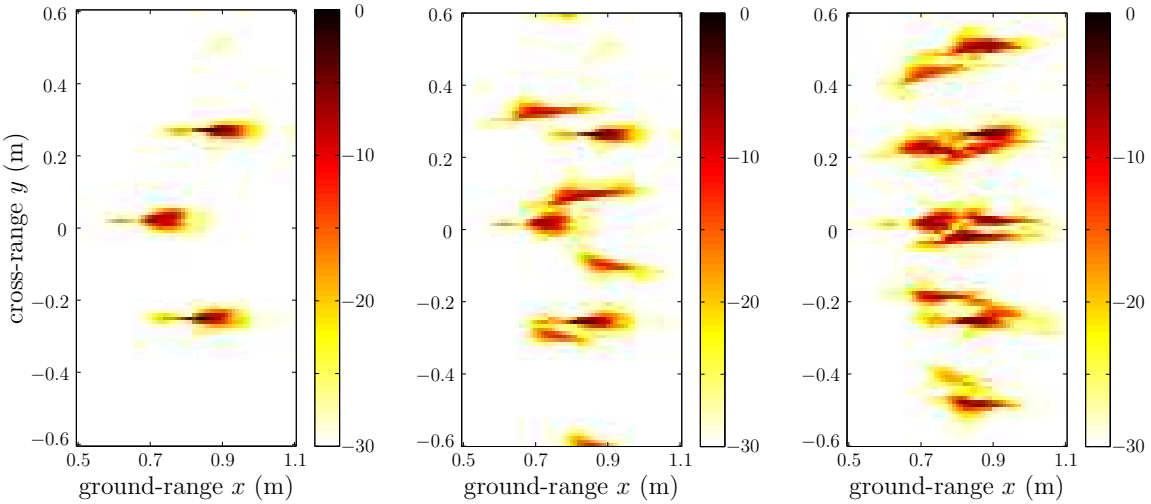
Figure 6.7: Ultrasonic synthetic aperture laboratory system (a)-(b) used to record spatially undersampled acoustical data of the target scene (c) in order to verify the proposed CS reconstruction technique.

The received signals are recorded using a National Instruments data acquisition card. The system parameters of Table 6.1 are used for the laboratory system, in accordance with the simulations. Furthermore, a speed of $v = 0.05$ m/s and a pulse repetition time of $T_{\mathrm{PRI}} = 0.12$ s are set to meet the spatial sampling requirements as discussed in Section 2.3. The employed high oversampling rate is due to discrete-time demodulation of the received echo signals. However, the discrete-time signals are then downsampled to meet the temporal Nyquist rate of the transmitted pulse. The imaging scene in Figure 6.7c consists of three ping-pong balls placed similarly to the point targets of the synthetic data examples.

6.3.3 Experimental data results

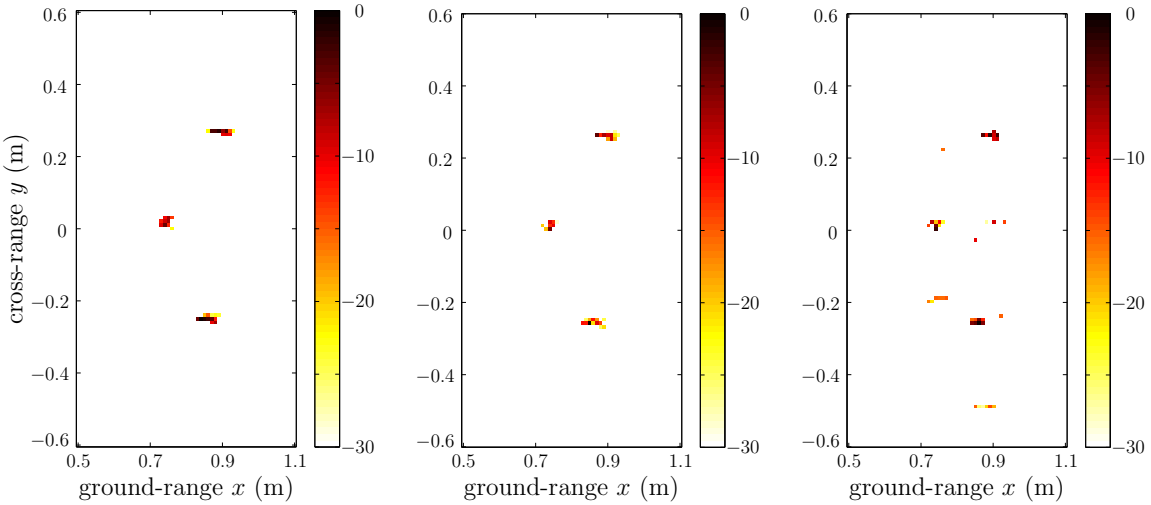
After collecting the raw measurements, both imaging methods are applied to the experimental acoustical data. In the case of aperture undersampling, the platform speed is increased to $v = 0.1$ m/s and $v = 0.15$ m/s, which is equivalent to an undersampling factor of $\kappa^{\mathrm{p}} = 2$ and $\kappa^{\mathrm{p}} = 3$, respectively. The regular along-track undersampling scheme is used as depicted in Figure 6.1a and the CS regularization parameter is set as for the synthetic data examples. The corresponding imaging results are depicted for the conventional TDC and CS-based technique in Figure 6.8 and Figure 6.9, respectively. The dynamic range of all images is 30 dB. In Figure 6.8a, a clean image reconstruction of the three ping-pong balls can be seen, where the cross-range resolution is better than the ground-range resolution due to the relatively small bandwidth of the ultrasound sensors. Increasing the platform speed for the conventional imaging method yields azimuth ambiguities of varying strength on both sides of the true target location. This effect is due to a non-straight alignment of the ultrasound sensors used in the laboratory system.

Considering the imaging results of the proposed CS method, it becomes apparent that azimuth image ambiguities can be successfully suppressed for real data measurements yielding images with almost identical quality – comparing Figure 6.9a to Figure 6.9b. Hence, the platform speed can be doubled without any loss in image quality. For a higher speed of $v = 0.15$ m/s, CS reconstruction starts to suffer from ambiguities as depicted in Figure 6.9c. It should be remarked that a target scenario with closely spaced targets has been chosen in order to keep the target strength variability small. Otherwise, weaker targets might be suppressed by enforcing the sparsity of the scene. Moreover, the undersampling ratio is significantly smaller compared to the synthetic data results but still twice as large as required for Nyquist sampling.



(a) $\Delta^{\mathrm{A}} = \Delta^{\mathrm{A}}_{\mathrm{max}}$, $v = 0.05$ m/s (b) $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{A}}_{\mathrm{max}}$, $v = 0.1$ m/s (c) $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{A}}_{\mathrm{max}}$, $v = 0.15$ m/s

Figure 6.8: Imaging results of real acoustical measurements using the conventional TDC technique for increasing values of the platform speed in (a)-(c).



(a) $\Delta^{\mathrm{A}} = \Delta^{\mathrm{A}}_{\mathrm{max}}$, $v = 0.05$ m/s (b) $\Delta^{\mathrm{A}} = 2\Delta^{\mathrm{A}}_{\mathrm{max}}$, $v = 0.1$ m/s (c) $\Delta^{\mathrm{A}} = 3\Delta^{\mathrm{A}}_{\mathrm{max}}$, $v = 0.15$ m/s

Figure 6.9: Imaging results of real acoustical measurements using the proposed CS technique for increasing values of the platform speed in (a)-(c).

6.4 Conclusions

In this chapter, a CS imaging technique for synthetic aperture systems operating in stripmap mode has been proposed. It is able to handle both transceiver as well as single-transmitter and multi-receiver synthetic aperture systems. The technique is based on the conventional TDC method but uses a BPDN solver to find the solution to an underdetermined linear equation system. Its capability to suppress azimuth image ambiguities has been demonstrated in the case of synthetic data and real acoustical measurements. Especially for synthetic data, measurements could be largely reduced given a perfect match between the original data model and the CS reconstruction model. By contrast, significantly less undersampling has been shown to be feasible regarding the laboratory system. This is most likely due to model mismatches between the used target scene and the assumption of point targets. Nevertheless, the speed of the imaging platform could be doubled while maintaining image quality.

Chapter 7
Conclusions and future work

Synthetic aperture sonar (SAS) ground-range imaging has been addressed in this thesis. A processing chain consisting of a data-driven motion compensation technique and a space-time imaging method has been developed for real sonar data. For ground-range imaging, the importance of knowing the seafloor topography at least roughly has been demonstrated. Additionally, an autofocus extension for stripmap SAS images and a technique for sound speed estimation has been proposed. Based on the SAS processing chain, an empirical sensitivity study of unknown motion influence on an automatic detection and classification (ADAC) system has been conducted, and its relation to image quality has been shown. As a consequence, an image quality assessment scheme based on the instantaneous cross-range resolution of a SAS system has been developed. Finally, a technique for synthetic aperture imaging based on compressive sensing (CS) has been addressed to overcome spatial sampling limitations while maintaing image quality. A summary of the main conclusions of this thesis is given in Section 7.1, and recommendations for possible future research work are addressed in Section 7.2.

7.1 Conclusions

Single-transmitter and multi-receiver SAS ground-range imaging requires a common grid of the target area to enable coherent processing of individual real aperture sonar (RAS) images. Thus, a sequential framework has been proposed, which also allows for the use of topography information estimated by a broadside bathymetry technique. For real sonar measurements, the impact of a continuous roll movement of an autonomous underwater vehicle (AUV) has been investigated and a compensation technique based on interpolating available roll information at consecutive transmission times has been proposed to overcome stop-and-hop limitations. For synthetic data, a known depth of focus (DOF) measure has been employed for ground-range imaging to predict the occurrence of image blurring if topography knowledge is missing. Furthermore, synthetic aperture shading has been discussed as a post-processing technique to trade off resolution with image contrast by controlling the mainlobe and sidelobes of point scatterers via apodization by a window function.

A processing chain of an echo-data-driven motion compensation technique known as micronavigation has been developed to estimate unknown platform motion, which is

essential to reconstruct high-resolution and well-focused SAS images. The processing chain has been validated on synthetic data as well as on real sonar measurements. SAS image examples have then been provided to highlight the image quality enhancement using micronavigation. A detailed description of the required signal processing steps, mainly consisting of surge estimation, time delay estimation, and ground-range sway and heave estimation based on nonlinear least squares fitting, has been presented. Due to operational constraints, subsample time delay estimates suffer from a high ratio of carrier frequency to bandwidth, which may affect ground-range sway and heave estimation. A compensation method based on binary image processing has been proposed to correct biased time delay estimates. Superior performance of the proposed approach compared to median filtering has been demonstrated on real sonar measurements. Another source of biased ping-to-ping motion estimates has been theoretically shown to stem from performing time delay estimation directly on raw echo measurements, which occurs especially for large physical arrays. As a consequence, a pitch motion may be erroneously assumed, which leads to biased estimation of heave motion. In order to compensate for these near-field and widebeam effects, a beamforming approach in broadside direction has been proposed. Finally, the impact of unknown topography information on ping-to-ping motion estimation has been exemplified. Based on synthetic data, it has been clearly demonstrated that height grids for space-time ground-range imaging along with unbiased motion estimates are required to obtain high-quality SAS imagery in environments with a challenging seafloor topography.

In addition to micronavigation, autofocus techniques for stripmap synthetic aperture imaging have been considered in this thesis. Two modifications have been introduced to enhance the mosaic phase gradient autofocus (mPGA) method. Its extended version has shown superior performance in synthetic data simulations. Exemplary testing on real SAS images has also improved image quality. Besides unknown platform motion, sound speed variations are another major source of image defocusing. A sound speed estimation technique that iteratively reconstructs small image patches and assesses their image quality based on different metrics has been developed and successfully applied to real sonar measurements. This technique has demonstrated its potential to be a software alternative to hardware measurement devices and, possibly, applicable as an online calibration routine.

An empirical sensitivity study has been conducted to investigate the influence of residual or unknown motion on ADAC performance. The study has revealed that already a small ground-range sway motion has a significant impact on conventional segmentation methods, feature extraction, and classification performance for small fractions of a wavelength given the common geometry of an AUV based SAS system.

The same impairment has been observed for heave motion amplitudes with multiples of the carrier wavelength, highlighting the importance of topography knowledge for SAS imaging and target recognition tasks. An empirical relation between a loss in image quality and a deterioration in ADAC performance has been demonstrated. However, training the ADAC system with motion-affected SAS images has noticeably improved classification performance. These results highlight the importance of developing image quality assessment schemes, robust segmentation methods, and feature selection techniques, to improve the reliability of SAS based ADAC systems in difficult scenarios. To this end, a strategy, which interprets SAS imaging as a sequential and coherent processing of RAS images, has been proposed to assess image quality by means of the instantaneous cross-range resolution. Its applicability has been shown on real sonar data, however, it requires the presence of isolated point scatterers to estimate the cross-range resolution.

The framework of CS has been employed for stripmap imaging using both a transceiver as well as a single-transmitter and multi-receiver system. The imaging technique allows for synthetic aperture undersampling while suppressing the occurrence of azimuth ambiguities. Synthetic data simulations have been conducted to highlight the potential for reducing the amount of collected data (up to 95%) with an unchanging image quality. Based on an ultrasonic laboratory system, which has been set up for validation purposes, a possible increase of the platform speed by a factor of two has been demonstrated. Consequently, higher coverage rates can be achieved, which is a first step towards reducing the mission time of future systems.

7.2 Future work

In the current SAS processing chain, the seafloor topography is estimated based on ping-wise single beam signals in broadside direction. However, the physical footprint of the real aperture covers a larger sector of the seafloor. Hence, the introduced broadside estimation technique has to be employed as a pre-processing method *prior* to SAS processing. A rough height grid of the topography is then obtained using interpolation. In order to be directly available in the SAS processing chain, sequential bathymetry estimation techniques with low computational complexity should be developed in future work. Furthermore, the DOF criterion should be adapted, taking into account the compensating effects of the micronavigation method, to produce a reliable defocusing prediction for real sonar data. This would allow the design of a decision rule on whether or not additional means in terms of bathymetry estimation are necessary to reconstruct high-quality SAS images.

In addition to height grid construction, using height information inside the data-driven motion compensation process is required to obtain unbiased ground-range sway and heave estimates. However, an initial height estimation technique has to solely rely on position data provided by the inertial navigation system (INS). Consequently, this leads to an inaccurate tracking of altitude positions that directly affects the accuracy of the height estimates. Given the low computational complexity of both broadside bathymetry estimation and nonlinear least squares fitting for displacement estimation, an iterative scheme should be investigated that repeats the process of estimating the seafloor topography as well as the sway and heave motion; possibly using the model fit accuracy as a termination criterion. Irrespective of such a scheme, the fit accuracy should be used to develop a test criterion that automatically decides whether the flat bottom assumption or the estimated height is of preferred choice. Further, a relationship between unknown seafloor topography and biased motion estimates should be derived theoretically by considering a model mismatch in the nonlinear least squares fitting process.

Irrespective of attainable high-quality SAS imagery, the development of an inherently robust processing chain for reliable automatic detection and classification (ADAC) is of utmost importance. For this purpose, each single processing step inside the ADAC system should be robustified. First of all, modified segmentation techniques should be developed to guarantee reasonable feature quality in case of image degradation. Another aspect that should be solved in the future is the design of a feature selection algorithm with the ability to determine the best feature subset in terms of class separability and motion error insensitivity. A study with larger data sets from different SAS and ADAC systems should then be used to more accurately generalize the obtained classification results. In addition to robust methods, detecting situations of image defocus by means of quality assessment is of great importance for an ADAC system in order to judge the reliability of classification performance in normal and difficult scenarios. This requires research in the field of blind image quality assessment for speckle affected imagery. Ultimately, a future ADAC system should always be tested concerning its performance under non-perfect image quality scenarios.

CS synthetic aperture imaging has achieved huge data reduction, especially for synthetic data where a perfect match between the echo signal model and reconstruction model has been given. By contrast, significantly less undersampling has been achieved for the laboratory system, which is most likely due to model mismatches between the real target scene and the assumption of point targets. Currently, there are still open challenges that have to be solved before CS can be employed in a real non-laboratory system. In particular, this involves the handling of target scenes consisting of heterogeneous target reflectivities and the consideration of extended targets rather

than point targets. While heterogenous target scenes may be addressed by improved echo data modeling, the challenge of extended targets may be handled by choosing an appropriate sparsity transform. Further, the echo data model should be adapted to mitigate the stop-and-hop assumption. Moreover, a procedure for automatically selecting the regularization parameter is required, and a jittered pulsing scheme should be applied to randomize the undersampling so as to weaken azimuth image ambiguities.

Appendix

A.1 Phase center range difference

In order to derive an expression for the relation between the ground-range sway $\Delta^{\mathrm{s}}_{p+1}$ and the range difference $\Delta r^{\mathrm{pc}}_{d,p}(u_{ij})$ of redundant phase center (RPC) pairs $u_{ij} = (u_i, u_j)$, two phase center arrays at ping p and $p+1$ are considered as follows:

- Phase center array at ping p: $\boldsymbol{a}^{\mathrm{pc}}_p(u) = \left[0, -u\Delta^{\mathrm{pc}}\right]^T,\ u = 1, \ldots, N_{\mathrm{rx}}$
- Phase center array at ping $p+1$: $\boldsymbol{a}^{\mathrm{pc}}_{p+1}(u) = \left[\Delta^{\mathrm{s}}_{p+1}, \Delta^{\mathrm{INS}} - u\Delta^{\mathrm{pc}}\right]^T,\ u = 1, \ldots, N_{\mathrm{rx}}$.

Applying a first-order Maclaurin series expansion and assuming a narrowbeam condition yields the range difference $r^{\mathrm{pc}}_{d,p}(u)$ between the phase center array at ping p and a point target located at $\boldsymbol{q_d} = [x_d, y_d]^T$ as

$$\begin{aligned} r^{\mathrm{pc}}_{d,p}(u) &= \left\|\boldsymbol{a}^{\mathrm{pc}}_p(u) - \boldsymbol{q}_d\right\|_2 = \left(x_d^2 + (y_d + u\Delta^{\mathrm{pc}})^2\right)^{1/2} \\ &= x_d \left(1 + \frac{(y_d + u\Delta^{\mathrm{pc}})^2}{x_d^2}\right) \approx x_d + \frac{\left(y_d + u\Delta^{\mathrm{pc}}\right)^2}{2x_d}. \end{aligned} \tag{A.1}$$

For ping $p+1$, it follows

$$\begin{aligned} r^{\mathrm{pc}}_{d,p+1}(u) &= \left\|\boldsymbol{a}^{\mathrm{pc}}_{p+1}(u) - \boldsymbol{q}_d\right\|_2 = \left((x_d - \Delta^{\mathrm{s}}_{p+1})^2 + (y_d + u\Delta^{\mathrm{pc}} - \Delta^{\mathrm{INS}})^2\right)^{1/2} \\ &= (x_d - \Delta^{\mathrm{s}}_{p+1}) \left(1 + \frac{(y_d + u\Delta^{\mathrm{pc}} - \Delta^{\mathrm{INS}})^2}{(x_d - \Delta^{\mathrm{s}}_{p+1})^2}\right) \\ &\approx (x_d - \Delta^{\mathrm{s}}_{p+1}) + \frac{\left(y_d + u\Delta^{\mathrm{pc}} - \Delta^{\mathrm{INS}}\right)^2}{2x_d}. \end{aligned} \tag{A.2}$$

Taking the range difference $\Delta r^{\mathrm{pc}}_{d,p}(u_{ij}) = r^{\mathrm{pc}}_{d,p}(u_i) - r^{\mathrm{pc}}_{d,p+1}(u_j)$ with $u_j = u_i + \Delta^{\mathrm{INS}}/\Delta^{\mathrm{pc}}$ and assuming that the advance per ping Δ^{INS} is an integer multiple of the phase center interelement spacing Δ^{pc} leads to

$$\begin{aligned} \Delta r^{\mathrm{pc}}_{d,p}(u_{ij}) &= 2\left(r^{\mathrm{pc}}_{d,p}(u_i) - r^{\mathrm{pc}}_{d,p+1}(u_j)\right) \\ &\approx 2\left(x_d + \frac{\left(y_d + u_i\Delta^{\mathrm{pc}}\right)^2}{2x_d} - (x_d - \Delta^{\mathrm{s}}_{p+1}) - \frac{\left(y_d + u_j\Delta^{\mathrm{pc}} - \Delta^{\mathrm{INS}}\right)^2}{2x_d}\right) \\ &= 2\Delta^{\mathrm{s}}_{p+1} + \frac{\left(y_d + u_i\Delta^{\mathrm{pc}}\right)^2 - \left(y_d + (u_i + \Delta^{\mathrm{INS}}/\Delta^{\mathrm{pc}})\Delta^{\mathrm{pc}} - \Delta^{\mathrm{INS}}\right)^2}{x_d} \\ &= 2\Delta^{\mathrm{s}}_{p+1}. \end{aligned} \tag{A.3}$$

A.2 Single-transmitter and multi-receiver range difference

In order to derive an expression for the relation between the ground-range sway $\Delta^{\mathrm{s}}_{p+1}$ and the range difference $\Delta r^{\mathrm{rx}}_{d,p}(u_{ij})$ of the corresponding receiver elements of redundant phase center (RPC) pairs $u_{ij} = (u_i, u_j)$, the transmitter and receiver array locations of ping p and $p+1$ are considered as follows:

- Transmitter and receiver at ping p:
  $\boldsymbol{a}^{\mathrm{tx}}_p = \left[0, 0\right]$ and $\boldsymbol{a}^{\mathrm{rx}}_p(u) = \left[0, -u\Delta^{\mathrm{u}}\right]^T,\ u = 1, \ldots, N_{\mathrm{rx}}$
- Transmitter and receiver at ping $p+1$:
  $\boldsymbol{a}^{\mathrm{tx}}_{p+1} = \left[\Delta^{\mathrm{s}}_{p+1}, \Delta^{\mathrm{INS}}\right]$ and $\boldsymbol{a}^{\mathrm{rx}}_{p+1}(u) = \left[\Delta^{\mathrm{s}}_{p+1}, \Delta^{\mathrm{INS}} - u\Delta^{\mathrm{u}}\right]^T,\ u = 1, \ldots, N_{\mathrm{rx}}.$

Applying a first-order Maclaurin series expansion and assuming a narrowbeam condition yields the range difference $r_{d,p}(u)$ between the multi-static system at ping p and a point target located at $\boldsymbol{q_d} = [x_d, y_d]^T$ as

$$\begin{aligned} r_{d,p}(u) &= \left\|\boldsymbol{a}^{\mathrm{tx}}_p - \boldsymbol{q}_d\right\|_2 + \left\|\boldsymbol{a}^{\mathrm{rx}}_p(u) - \boldsymbol{q}_d\right\|_2 = \left(x_d^2 + y_d^2\right)^{1/2} + \left(x_d^2 + (y_d + u\Delta^{\mathrm{u}})^2\right)^{1/2} \\ &= x_d \left(1 + \frac{y_d^2}{x_d^2}\right)^{1/2} + x_d \left(1 + \frac{(y_d + u\Delta^{\mathrm{u}})^2}{x_d^2}\right)^{1/2} \\ &\approx 2x_d + \frac{y_d^2 + \left(y_d + u\Delta^{\mathrm{u}}\right)^2}{2x_d}. \end{aligned} \tag{A.4}$$

For ping $p+1$, it follows

$$\begin{aligned} r_{d,p+1}(u) &= \left\|\boldsymbol{a}^{\mathrm{tx}}_{p+1} - \boldsymbol{q}_d\right\|_2 + \left\|\boldsymbol{a}^{\mathrm{rx}}_{p+1}(u) - \boldsymbol{q}_d\right\|_2 \\ &= \sqrt{x_d - \Delta^{\mathrm{s}}_{p+1})^2 + (y_d - \Delta^{\mathrm{INS}})^2} + \sqrt{(x_d - \Delta^{\mathrm{s}}_{p+1})^2 + (y_d + u\Delta^{\mathrm{u}} - \Delta^{\mathrm{INS}})^2} \\ &= (x_d - \Delta^{\mathrm{s}}_{p+1}) \left[\left(1 + \frac{(y_d - \Delta^{\mathrm{INS}})^2}{(x_d - \Delta^{\mathrm{s}}_{p+1})^2}\right)^{1/2} + \left(1 + \frac{(y_d + u\Delta^{\mathrm{u}} - \Delta^{\mathrm{INS}})^2}{(x_d - \Delta^{\mathrm{s}}_{p+1})^2}\right)^{1/2}\right] \\ &\approx 2(x_d - \Delta^{\mathrm{s}}_{p+1}) + \frac{(y_d - \Delta^{\mathrm{INS}})^2 + (y_d + u\Delta^{\mathrm{u}} - \Delta^{\mathrm{INS}})^2}{2(x_d - \Delta^{\mathrm{s}}_{p+1})} \\ &\approx 2(x_d - \Delta^{\mathrm{s}}_{p+1}) + \frac{(y_d - \Delta^{\mathrm{INS}})^2 + (y_d + u\Delta^{\mathrm{u}} - \Delta^{\mathrm{INS}})^2}{2x_d}. \end{aligned} \tag{A.5}$$

Taking the range difference $\Delta r^{\mathrm{rx}}_{d,p}(u_{ij}) = r_{d,p}(u_i) - r_{d,p+1}(u_j)$ with $u_j = u_i + \Delta^{\mathrm{INS}}/\Delta^{\mathrm{pc}}$ and assuming that the advance per ping Δ^{INS} is an integer multiple of the phase center interelement spacing Δ^{pc} leads to

$$\Delta r^{\mathrm{rx}}_{d,p}(u_{ij}) = 2\Delta^{\mathrm{s}}_{p+1} - \frac{\Delta^{\mathrm{INS}}(\Delta^{\mathrm{INS}} + u_i\Delta^{\mathrm{u}})}{x_d},\ u_i = 1, \ldots, N_{\mathrm{pc}}. \tag{A.6}$$

List of Acronyms

ADAC	Automatic detection and classifcation
ADC	Analog-to-digital converter
ATR	Automatic target recognition
AUV	Autonomous underwater vehicle
BPDN	Basis pursuit denoising
CS	Compressive sensing
DOF	Depth of focus
DPCA	Displaced phase center antenna
DTFT	Discrete-time Fourier transform
FFT	Fast Fourier transform
GC	Graph cut
ICE	Iterative conditional estimates
ICM	Iterative conditional modes
INS	Inertial navigation system
ISLL	Integrated sidelobe level
KDE	Kernel density estimation
LDA	Linear discriminant analysis
LFM	Linear frequency modulation
MRF	Markov random fields
mPGA	Mosaic phase gradient autofocus
mmPGA	Modified mosaic phase gradient autofocus
PGA	Phase gradient autofocus
PRI	Pulse repetition interval
PSF	Point spread function

RAS	Real aperture sonar
RPC	Redundant phase center
SAR	Synthetic aperture radar
SAS	Synthetic aperture sonar
SFFS	Sequential forward floating selection
SFS	Sequential forward selection
SLL	Sidelobe level
SNR	Signal-to-noise ratio
SSIM	Structural similarity measure
TDC	Time-domain correlation
TDE	Time delay estimation
ULA	Uniform linear array
WSS	Wide-sense stationary

List of Symbols

The following list contains symbols in alphabetic order, which occur more than once in this thesis. The remaining symbols are directly introduced where they are used.

Latin symbols

$\boldsymbol{a}_p, \boldsymbol{a}_p^{\mathrm{pc}}(u)$	Vector of Cartesian coordinates of a transceiver and of element u of a phase center array at ping p
$\boldsymbol{a}_p^{\mathrm{ref}}$	Vector of beamforming reference point in Cartesian coordinates at ping p
$\boldsymbol{a}_p^{\mathrm{rx}}(u), \boldsymbol{a}_p^{\mathrm{tx}}$	Vector of Cartesian coordinates of receiver u and of a transmitter in a single-transmitter and multi-receiver system at ping p
$A(y)$	Current distribution along the physical aperture dimension y
A_{cr}	Area coverage rate of an imaging system
A^s, A^h	Amplitude of unknown sway motion and heave motion
$\mathrm{AF}_{\mathrm{syn}}(\theta), \mathrm{AF}_{\mathrm{phy}}(\theta)$	Array factor as a function of azimuth direction θ of a synthetic array and a physical array
$b_{\mathrm{phy}}(\theta)$	Beampattern or element factor of a physical aperture
B, B_x, B_z	Interferometric baseline between a pair of receiving elements, and baseline components in Cartesian x and z directions
B_{k_x}, B_{k_y}	Spatial frequency support in 2-D wavenumber domain in ground-range and along-track dimensions
c	Speed of propagating wave in the medium, *e.g.*, sound in water or air, or speed of light in air
$c_0, \hat{c}_0, \tilde{c}$	Measured, estimated, and erroneous sound speed value
$c_{\min}, c_{\max}$	Minimum and maximum sound speed values of the interval $\mathcal{I}_c$
D	Number of point scatterers in a target scene
D_{phy}	Diameter of a physical receiver element in along-track direction
D_{SFS}	Number of branches for feature selection algorithm
$e_p(n), e_p(u, n)$	Discrete and pulse-compressed lowpass echo signals at ping p for a transceiver system, and for receiver u of a multi-receiver system
$\bar{e}_p(u, n)$	Mean removed echo signals

$e_p^{\mathrm{B}}(\theta, n)$	Beamformed echo signals in discrete-time as a function of azimuth direction θ at ping p
$e_p^{\mathrm{BP}}(u,t), e_p^{\mathrm{LP}}(u,t)$	Continuous-time and pulse-compressed bandpass and lowpass echo signals of a multi-receiver synthetic aperture system
$\boldsymbol{e}, \boldsymbol{e}_{\mathrm{cs}}$	Vector representation of discrete-time lowpass echo signals, and vector of compressive sensing model
$\boldsymbol{E}$	Matrix representation of discrete-time lowpass echo signals along the entire synthetic aperture
$E_A(\theta)$	Amplitude pattern of a planar antenna of azimuth direction θ
f_B, f_c	Bandwidth and carrier frequency of transmitted signal $s^{\mathrm{BP}}(t)$
$f_{\min}, f_{\max}$	Minimum and maximum frequency of transmitted signal $s^{\mathrm{BP}}(t)$
f_{p}	Cycle per synthetic aperture length frequency for a sinusoidal path deviation function
f_s	Sampling frequency
$f(x,y)$	Ideal reflectivity function of the target area with continuous coordinates x in ground-range and y in along-track direction
$\hat{f}(x_k, y_l), \tilde{f}(x_k, y_l)$	Reconstructed and degraded synthetic aperture image of the target area with discrete-spatial coordinates x_k and y_l in ground-range and along-track direction, respectively
$\bar{f}(x_k, y_l)$	Image with removed sample mean intensity
$\hat{f}_p(x_k, y_l)$	Reconstructed real aperture sonar image of the discretized target area with coordinates (x_k, y_l) at ping p
$\hat{f}_{i,j}(x_k, y_l)$	Discrete spatial mosaic piece (2-D sub-image) with index pair (i,j) after autofocus iteration
$\tilde{f}_{i,j}(x_k, y_l)$	Discrete spatial mosaic piece (2-D sub-image) with index pair (i,j) before autofocus iteration
$\hat{F}(k_x, k_y)$	Continuous representation of the 2-D wavenumber domain of the reconstructed image $\hat{f}(x_k, y_l)$
$\tilde{F}(k_x, k_y)$	Continuous representation of the 2-D wavenumber domain of the blurred and defocused image $\tilde{f}(x_k, y_l)$
$\hat{\boldsymbol{f}}$	Stacked vector of the reconstructed synthetic aperture image
$\hat{\boldsymbol{F}}$	Matrix notation of the reconstructed synthetic aperture image
$\boldsymbol{g}^{\mathrm{fp}}(\theta, r)$	Slant-range focusing vector
$\boldsymbol{g}_{kl}$	Grid point vector of target area with $\boldsymbol{g}_{kl} = [x_k, y_l, 0_{kl}]^T$ for a flat bottom grid or $\boldsymbol{g}_{kl} = [x_k, y_l, z_{kl}]^T$ for an *a priori* height grid
$\boldsymbol{G}$	Focusing matrix used in vector/matrix based imaging

h_p^{og}	Altitude component of a synthetic aperture at ping p
$\boldsymbol{H}$	Hessian matrix used in nonlinear least squares estimation
$\boldsymbol{I}_B$	Binary image matrix used in phase wrap error compensation
$\boldsymbol{I}_C$	Corrected binary image after performing morphological operation
$\boldsymbol{I}^{\mathrm{PCR}}$	Matrix with labels of detected phase cycle regions
$\mathcal{I}_c$	Sound speed interval with $\mathcal{I}_c = [c_{\min}, c_{\max}]$
$\boldsymbol{J}$	Jacobian matrix used in nonlinear least squares estimation
k	Grid point index in ground-range dimension
k_{s}	Short-time (slant-range) sliding window index
k_c, k_r	Wavenumber at the carrier frequency, and wavenumber
k_x, k_y	Wavenumber in ground-range and along-track direction
K	Number of sliding short-time windows in slant-range direction
K_x, K_y	Number of mosaic pieces in ground-range and along-track direction
l	Grid point index in along-track dimension
L_k	Length of short-range (short-time) window
$L_{\mathrm{phy}}, L_{\mathrm{syn}}$	Length of physical array and length of synthetic aperture
$L_{\mathrm{syn}}^{\max}(r), L_{\mathrm{syn}}(p, r)$	Maximum synthetic aperture length and instantaneous synthetic aperture length as a function of ping index p and focusing range r
m_f	Ambiguous number of full phase cycles in time delay estimation
M, M'	Number of measurements and number of reduced measurements
M_k	Number of samples of a short-time window
M_n, M_p	Number of fast-time samples, and number of pings (slow-time)
M_x, M_y	Number of grid points in ground-range and along-track direction
M'_x, M'_y	Number of pixels in ground-range and along-track dimension of a 2-D window
M_η	Number of delay samples in vector notation of signals
M_Π	Number of samples per pulse length
n	Discrete fast-time index variable
N_{iter}	Number of iterations, *e.g.*, for autofocus algorithm
N_{MC}	Number of experiments

$N^{y}_{\max}$	Number of phase centers of a maximum expected surge
N_{obj}	Number of man-made objects in SAS image database
$N_{\text{pc}}, N'_{\text{pc}}$	Number of redundant phase centers, and number of redundant phase centers inside a sliding window used for beamforming
N_{pcr}	Number of phase cycle regions
N_{rx}	Number of receiving elements of a uniform linear array
$N_{\varphi}, N'_{\varphi}$	Number of feature descriptors and size of optimal feature set
p	Ping index variable (slow-time variable)
p_0, p_1	First and last ping for which a certain target area is illuminated
$P_{\text{fa}}, P_{\text{mc}}$	Probability of false alarm and probability of misclassification in automatic target recognition
q_{it}	Iteration index variable in simulation of autofocus techniques
q_{ϵ}	Mismatch value in percent to model sound speed errors
$\boldsymbol{q}_d$	Position vector of point target d with $\boldsymbol{q}_d = [x_d, y_d, z_d]^T$
r	Slant-range variable
$r_{p,d}$	Slant-range distance between transceiver position at ping p and point target d
$r^{\text{rx}}_{p,d}(u), r^{\text{tx}}_{p,d}$	Slant-range distance between location of receiver u at ping p and point target d as well as slant-range distance for the transmitter
R_{far}	Fraunhofer distance
$R_{\min}, R_{\max}$	Minimum and maximum slant-range values
$s^{\text{BP}}(t), s^{\text{LP}}(t)$	Transmitted bandpass signal, and equivalent lowpass signal
S_{E}	Structuring element for morphological operations
$\boldsymbol{s}$	Vector notation of transmitted signal
$\boldsymbol{S}$	Matrix of transmitted signals along the synthetic aperture
$\boldsymbol{S}^{\zeta}$	Matrix of segmented pixels for a region of interest ζ
t	Continuous-time variable
t_p	Time instants of pulse transmission at ping p
T_s	Sampling interval
T_{PRI}	Pulse repetition time interval between two pings
T_{Π}	Pulse duration of the transmitted signal

u	Index variable of a receiving element of a uniform linear array
u_i, u_j	Phase center index variable, $u_{ij} = (u_i, u_j)$ describes a redundant phase center pair
v	Speed of the imaging platform
$v_p(n), v_p(u, n)$	Thermal noise process for a transceiver system and for receiving element u of a single-transmitter and multi-receiver system
$\boldsymbol{v}$	Vector representation of thermal noise process
$w^{\mathrm{tx/rx}}(\cdot)$	Aspect depending weighting function for transmitter and receiver
$W(k_x, k_y)$	2-D window defined in the wavenumber domain
x	Ground-range spatial variable
x_d	Ground-range coordinate of target d
X_0, X_c	Half swath width of target area and center of target area in ground-range direction
$X_{\min}, X_{\max}$	Minimum and maximum boundaries of ground-range swath with $X_{\min} = X_c - X_0$ and $X_{\max} = X_c + X_0$
y	Along-track spatial variable
y_d	Along-track coordinate of target d
Y_0, Y_c	Half swath width of target area and center of target area in along-track direction
$Y_{\min}, Y_{\max}$	Minimum and maximum boundaries of along-track extent with $Y_{\min} = Y_c - Y_0$ and $Y_{\max} = Y_c + Y_0$
$Y_{\min}^{\mathrm{RAS}}, Y_{\max}^{\mathrm{RAS}}$	Minimum and maximum boundaries of along-track extent of a real aperture sonar image
z	Height variable
z_d	Height coordinate of target d

Greek symbols

α_{c}	Linear chirp rate
α	Phase value of complex signals
β	Model parameter

γ^{ζ}	Average ratio of correctly segmented pixels for region of interest ζ
δ_r	Slant-range resolution
$\delta_{\mathrm{phy}}, \delta_{\mathrm{syn}}$	Along-track resolution of a physical array and of a synthetic aperture
$\delta_{\mathrm{syn}}(p, r)$	Instantaneous along-track resolution of a synthetic aperture system as a function of ping index p and focusing slant-range r
$\delta_{\mathrm{mod}}(p, \boldsymbol{\beta})$	Model function of instantaneous along-track resolution for a fixed range as a function of ping p and parameters $\boldsymbol{\beta}$
Δ^{A}	Advance per ping of the imaging platform (transceiver system)
$\Delta^{\mathrm{A}}_{\mathrm{max}}$	Spatial sampling constraint for a transceiver synthetic aperture
Δ^{INS}	Advance per ping measured by inertial navigation system
Δ^{pc}	Interelement spacing of phase center array
Δ^{u}	Interelement spacing of a uniform linear array
$\Delta^{\mathrm{ULA}}_{\mathrm{max}}$	Spatial sampling constraint for a single-transmitter and multi-receiver synthetic aperture system
$\Delta^{\mathrm{h}}_p, \Delta^{\mathrm{s}}_p, \Delta^{\mathrm{y}}_p$	Heave, sway, and surge motion at ping p
$\boldsymbol{\Delta}_p$	Ping-to-ping motion vector at ping p
$\Delta\boldsymbol{a}_p$	Absolute displacement vector at ping p
Δc	Sound speed mismatch value
$\Delta t^{\mathrm{B}}(\theta, r)$	Focusing delay for beamforming as a function of azimuth direction θ and slant-range r
$\Delta x, \Delta y$	Pixel size in ground-range and along-track directions
$\Delta\eta$	Scaled time delay difference with $\Delta\eta = \Delta\tau / T_s$
$\Delta\tau$	Time delay difference in continuous-time
$\Delta\Phi$	Phase gradient for phase estimation in autofocus
ζ	Region of interest (shadow or highlight) of segmentation procedure in automatic target recognition
ζ^{n}	Undersampling factor in fast-time domain
$\eta, \eta_{d,p}$	Scaled round-trip delay with $\eta = \tau / T_s$ and round-trip delay between target d and transceiver position at ping p
$\eta_{kl,p}$	Scaled focus delay between grid point $\boldsymbol{g}_{kl}$ and transceiver position at ping p of a synthetic aperture system

$\eta_{kl,p}(u)$	Scaled focus delay between grid point $\boldsymbol{g}_{kl}$ and u^{th} receiver position of a single-transmitter and multi-receiver system at ping p
θ, θ_0	Azimuth direction and azimuth broadside direction
$\theta_{\text{phy}}, \theta_{\text{syn}}$	Angular beamwidth of a physical array and of a synthetic aperture
$\theta_{d,p}$	Aspect angle between target d and transceiver position at ping p
$\theta^{\text{rx}}_{d,p}(u), \theta^{\text{tx}}_{d,p}$	Aspect angle between target d and receiver u, and between target and the transmitter of a synthetic aperture system at ping p
$\theta^{\text{rx}}_{kl,p}(u), \theta^{\text{tx}}_{kl,p}$	Aspect angle between grid point $\boldsymbol{g}_{kl}$ and the position of receiver u, and between grid point and transmitter position at ping p
θ_{BW}	Angular beamwidth of a physical aperture
θ_{3dB}	Half-power angular beamwidth of a physical aperture
κ^{p}	Undersampling factor in slow-time domain
λ, λ_c	Wavelength and wavelength at carrier frequency
$\lambda_{\min}, \lambda_{\max}$	Minimum and maximum wavelength of transmitted pulse signal
Λ_{cs}	Regularization parameter for compressive sensing
$\mu, \hat{\mu}$	Expected value of a quantity, and sample mean of a quantity
$\nu(x_{k_{\text{s}}}, \boldsymbol{\Delta})$	Parameter dependent slant-range residuals for nonlinear least squares estimation as a function of the ground-range index k_{s}
ρ	Correlation coefficient or peak correlation value
$\rho_{\min}$	Minimum correlation threshold
ϱ	Drop-rate of fast-time samples for compressive sensing
σ_d	Aspect- and frequency-independent reflectivity of point target d
$\boldsymbol{\sigma}$	Vector of target reflectivities of all point targets in a target scene
$\boldsymbol{\Sigma}$	Measurement reduction matrix for compressive sensing
τ	Round-trip delay in continuous-time
$\tau_{kl,p}(u)$	Focusing delay for image reconstruction between transmitter, receiver u and grid point $\boldsymbol{g}_{kl}$ at ping p
$\tau_{d,p}(u)$	Round-trip delay in echo signals between transmitter, receiver u and target d at ping p
$\hat{\tau}_c, \hat{\tau}_f$	Coarse and fine time delay estimate

$\hat{\tau}_\alpha$	Subsample time delay estimate
$\tau_{\tilde{m}_f}$	Delay error due to wrong selection of ambiguous number
$\boldsymbol{\upsilon}_\mathrm{A}$	Unit vector pointing in along-track direction
Υ_co	Contrast metric to assess image quality
Υ_max	Maximum image intensity or peak value of a point spread function
Υ_3dB	Half-power mainlobe width of a point spread function
$\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}}, \boldsymbol{\varphi}'$	Feature set, normalized feature set, and optimal feature subset
ϕ_tol	Maximum phase error tolerance
$\phi_\mathrm{err}(p)$	Phase error term as a function of ping p
ϕ_roll	Roll angle of the imaging platform
ϕ_B	Baseline angle for interferometry systems
$\boldsymbol{\chi}$	Sparse coefficient vector in compressive sensing
$\boldsymbol{\Psi}$	Sparse basis transformation matrix
ω_c	Discrete-time angular carrier frequency
Ω_c	Continuous-time angular carrier frequency

Functions and operators

$\mathcal{D}\{\cdot\}$	Motion compensation estimator
$E\{\cdot\}$	Ensemble average
$\mathcal{M}\{\cdot\}$	Mode operator selecting most frequent value in a set
$\mathcal{Q}\{\cdot\}$	Mapping operator of an image quality metric
$\mathrm{vec}\{\cdot\}$	Vectorize operation
$\mathrm{vec}^{-1}\{\cdot\}$	Reshape operation to obtain a matrix from a stacked vector
$\angle\{\cdot\}$	Angle operation
$\#\{\cdot\}$	Operator counting non-zero elements
$c_{\mathcal{X}\mathcal{Y}}(\kappa)$	Cross-covariance function
$\mathrm{r}_{ss}(\kappa)$	Auto-correlation function
$C(\cdot)$	Cost function
$\delta(\cdot)$	Dirac function
$\Lambda(\cdot)$	Triangular function
$\Pi(\cdot)$	Rectangular function

$\Xi(\cdot,\cdot)$	SSIM measure
$\oplus$	Morphological dilation operation
$\ominus$	Morphological erosion operation
$\vee$	Element-wise logical or-operation
$\odot$	Hadamard product
$*_t$	Convolution in fast-time
$*_p$	Convolution in slow-time
$(\cdot)^{\mathrm{T}}$	Transpose of a vector or matrix
$(\cdot)^{\mathrm{H}}$	Conjugate transpose of a vector or matrix
$(\cdot)^*$	Conjugate of a scalar, vector, or matrix
$(\cdot)^{-1}$	Inverse of a square matrix
$\hat{(\cdot)}$	Estimator or estimate of a quantity
$\tilde{(\cdot)}$	An erroneous or an unknown motion-affected quantity
$\lvert\cdot\rvert$	Absolute value of a scalar
$\lfloor\cdot\rceil$	Round towards nearest integer
$\lfloor\cdot\rfloor$	Round towards smallest integer
$\lceil\cdot\rceil$	Round towards largest integer
$\lVert\cdot\rVert_1$	L^1-norm of a vector
$\lVert\cdot\rVert_2$	Euclidean norm or L^2-norm of a vector
$\lVert\cdot\rVert_\infty$	Maximum norm of a vector
$\boldsymbol{I}_N$	Identity matrix of size N
$\mathbf{0}_{1\times N}$	Zero row vector with N elements
$\mathbf{0}_{N\times 1}$	Zero column vector with N elements
$\mathbb{C}$	Set of all complex numbers
$\mathbb{R}$	Set of all real numbers
$\mathbb{Z}$	Set of all integer numbers

Bibliography

S. Adak, "Time-dependent spectral analysis of nonstationary time series," *J. Amer. Statist. Assoc.*, vol. 93, no. 444, pp. 1488–1501, 1998.

T. Aridgides, M. F. Fernandez, and G. J. Dobeck, "Side-scan sonar imagery fusion for sea mine detection and classification in very shallow water," in *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 4394, no. 1, 2001, pp. 1123–1134.

R. Bamler, "A systematic comparison of SAR focusing algorithms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Espoo, Finland, Jun. 1991, pp. 1005–1009.

R. Bamler, "A comparison of range-Doppler and wavenumber domain SAR focusing algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 4, pp. 706–713, 1992.

R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, Boston, MA, Apr. 2007, pp. 128–133.

R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, 2007.

R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

S. Basu and Y. Bresler, "$O(N^2 \log_2 N)$ filtered backprojection reconstruction algorithm for tomography," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1760–1773, 2000.

S. Bhattacharya, T. Blumensath, B. Mulgrew, and M. Davies, "Synthetic aperture radar raw data encoding using compressed sensing," in *Proc. IEEE Radar Conf.*, Rome, Italy, May 2008, pp. 1–5.

J. Besag, "On the statistical analysis of dirty images," *J. R. Stat. Soc.*, vol. B-48, pp. 259–302, 1986.

L. Bjørnø, "Developments in sonar technologies and their applications," in *Proc. IEEE Int. Underwater Tech. Symp.*, Tokyo, Japan, Mar. 2013, pp. 1–8.

J. F. Böhme, *Stochastische Signale.* Berlin, Germany: Springer, 1998.

W. Bonifant, "Interferometic synthetic aperture sonar processing," Master's thesis, Dept. Electr. Eng., Georgia Institute of Technology, 1999.

W. Bonifant, M. Richards, and J. McClellan, "Interferometric height estimation of the seafloor via synthetic aperture sonar in the presence of motion errors," *IEE Proc. Radar Sonar Navig.*, vol. 147, no. 6, pp. 322–330, 2000.

Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, 2004.

A. Bellettini and M. A. Pinto, "Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna," *IEEE J. Ocean. Eng.*, vol. 27, no. 4, pp. 780–789, Oct. 2002.

A. Bellettini and M. A. Pinto, "Design and experimental results of a 300-kHz synthetic aperture sonar optimized for shallow-water operations," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 285–293, 2009.

C. Cafforio, C. Prati, and F. Rocca, "SAR data focusing using seismic migration techniques," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 2, pp. 194–207, 1991.

H. J. Callow, "Signal processing for synthetic aperture sonar image enhancement," Ph.D. dissertation, Dept. Electr. Eng., University of Canterbury, Christchurch, New Zealand, 2003.

H. J. Callow, M. P. Hayes, and P. T. Gough, "Autofocus of stripmap SAS data using the range-variant SPGA algorithm," in *Proc. MTS/IEEE OCEANS Conf.*, vol. 5, San Diego, CA, Sep. 2003, pp. 2422–2426.

H. J. Callow, M. P. Hayes, and P. T. Gough, "Stripmap phase gradient autofocus," in *Proc. MTS/IEEE OCEANS Conf.*, vol. 5, San Diego, CA, Sep. 2003, pp. 2414–2421.

H. J. Callow, M. P. Hayes, and P. T. Gough, "Motion-compensation improvement for widebeam, multiple-receiver SAS systems," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 262–268, Jul. 2009.

E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*. Boston, MA, USA: Artech House, 1995.

G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, no. 2, pp. 236–255, 1987.

S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

C. M. Ciany and W. Zurawski, "Performance of fusion algorithms for computer aided detection and classification of bottom mines in the shallow water environment," in *Proc. MTS/IEEE OCEANS Conf.*, vol. 4, Biloxi, MS, Oct. 2002, pp. 2164–2167.

W. S. Cleveland and E. Parzen, "The estimation of coherence, frequency response, and envelope delay," *Technometrics*, vol. 17, no. 2, pp. 167–172, 1975.

W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Amer. Statist. Assoc.*, vol. 74, no. 368, pp. 829–836, 1979.

L. Cohen, *Time-frequency signal analysis.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

E. Coiras and J. Groen, "3D target shape from SAS images based on a deformable mesh," in *Proc. Underwater Acoust. Meas.*, Nafplion, Greece, Jun. 2009, pp. 303–310.

C. Collet, P. Thourel, P. Pérez, and P. Bouthemy, "Hierarchical MRF modeling for sonar picture segmentation," in *Proc. Int. Conf. Image Process.*, vol. 3, 1996, pp. 979–982.

D. A. Cook, D. C. Brown, and J. E. Fernandez, "Synthetic aperture sonar motion estimation using nonlinear least squares," in *Proc. Inst. Acoust. Int. Conf. SAS/SAR*, vol. 28, no. 5, Lerici, Italy, Sep. 2006, pp. 176–182.

D. A. Cook, "Synthetic aperture sonar motion estimation and compensation," Master's thesis, Dept. Electr. Eng., Georgia Institute of Technology, 2007.

D. A. Cook and D. C. Brown, "Analysis of phase error effects on stripmap SAS," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 250–261, 2009.

I. Cumming, F. Wong, and K. Raney, "A SAR processing algorithm with no interpolation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Houston, TX, May 1992, pp. 376–379.

J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing.* New York, NY, USA: Wiley & Sons, 1991.

L. J. Cutrona, "Comparison of sonar system performance achievable using synthetic-aperture techniques with the performance achievable by more conventional means," *J. Acoust. Soc. Am.*, vol. 58, no. 2, pp. 336–348, 1975.

C. Debes, R. Engel, A. M. Zoubir, and A. Kraft, "Quality assessment of synthetic aperture sonar images," in *Proc. IEEE OCEANS Europe Conf.*, Bremen, Germany, May 2009, pp. 1–4.

G. J. Dobeck, J. C. Hyland, and L. Smedley, "Automated detection and classification of sea mines in sonar imagery," in *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 3079, no. 1, 1997, pp. 90–110.

M. F. Doherty, J. G. Landowski, P. F. Maynard, G. T. Uber, D. W. Fries, and F. H. Maltz, "Side scan sonar object classification algorithms," in *Proc. IEEE Int. Symp. Unmanned Untethered Submersible Technol.*, Durham, NH, Jun. 1989, pp. 417–424.

D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

J. H. G. Ender, "On compressive sensing applied to radar," *Signal Processing*, vol. 90, no. 5, pp. 1402–1414, 2010.

R. Fandos, L. Sadamori, and A. M. Zoubir, "High quality segmentation of synthetic aperture sonar images using the min-cut/max-flow algorithm," in *Proc. Eur. Signal Process. Conf.*, vol. 1, Barcelona, Spain, 2011, pp. 51–55.

R. Fandos and A. M. Zoubir, "Optimal feature set for automatic detection and classification of underwater objects in SAS images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 454–468, 2011.

R. Fandos, "ADAC system design and its application to mine hunting using SAS imagery," Ph.D. dissertation, Dept. Electr. Eng., Technische Universität Darmstadt, 2012.

R. Fandos, A. M. Zoubir, and K. Siantidis, "Unified design of a feature-based ADAC system for mine hunting using synthetic aperture sonar," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2413–2426, 2014.

R. Fandos, C. Debes, and A. M. Zoubir, "Resampling methods for quality assessment of classifier performance and optimal number of features," *Signal Processing*, 2013.

J. Fawcett, "Image-based classification of sidescan sonar detections," in *Proc. CAC/CAD Conf.*, Halifax, Canada, Nov. 2001.

M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, 2007.

K. G. F. Finckenstein, J. Lehn, H. Schellhaas, and H. Wegmann, *Arbeitsbuch Mathematik für Ingenieure, Band II: Differentialgleichungen, Funktionentheorie, Numerik und Statistik.* Berlin, Germany: Springer, 2006.

L. R. Ford and D. R. Fulkerson, *Flows in Networks*, Princeton, NJ, USA: Princeton Univ. Press, 1962.

G. Fornaro, "Trajectory deviations in airborne SAR: analysis and compensation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 3, pp. 997–1009, 1999.

S. A. Fortune, P. T. Gough, and M. P. Hayes, "Statistical autofocus of synthetic aperture sonar images using image contrast optimisation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, Sydney, Australia, 2001, pp. 1509–1511.

N. P. Glover and I. Campell, "Simultaneous low and high frequency high resolution SAS and a statistical method of quantifying the resolutions obtained," in *Proc. Inst. Acoust. Int. Conf. SAS/SAR*, vol. 32, no. 4, Lerici, Italy, Sep. 2010.

R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

P. T. Gough and D. W. Hawkins, "Imaging algorithms for a strip-map synthetic aperture sonar: minimizing the effects of aperture errors and aperture undersampling," *IEEE J. Ocean. Eng.*, vol. 22, no. 1, pp. 27–39, 1997.

H. D. Griffiths, T. A. Rafik, Z. Meng, C. F. N. Cowan, H. Shafeeu, and D. K. Anthony, "Interferometric synthetic aperture sonar for high resolution 3-D mapping of the seabed," *IEE Proc. Radar Sonar Navig.*, vol. 144, no. 2, pp. 96–103, Apr. 1997.

J. Groen, "Adaptive motion compensation in sonar array processing," Ph.D. dissertation, Faculty Appl. Phys., Technische Universiteit Delft, 2006.

J. Groen, E. Coiras, and D. P. Williams, "Detection rate statistics in synthetic aperture sonar images," in *Proc. Underwater Acoust. Meas.*, Nafplion, Greece, Jun. 2009, pp. 367–374.

J. Groen, E. Coiras, J. Del Rio Vera, and B. Evans, "Model-based sea mine classification with synthetic aperture sonar," *IET Radar, Sonar and Navig.*, vol. 4, no. 1, pp. 62–73, 2010.

J. Groen, M. Couillard, and W. L. Fox, "Synthetic aperture sonar array gain measured at sea," in *Proc. 9th Eur. Conf. Synthetic Aperture Radar*, Nuremberg, Germany, Apr. 2012, pp. 74–77.

J. Gunther, R. West, N. Crookston, and T. Moon, "Maximum likelihood synthetic aperture radar image formation for highly nonlinear flight tracks," in *Proc. IEEE Digital Signal Process. Workshop*, Sedona, AZ, Jan. 2011, pp. 449–454.

P. E. Hagen and R. E. Hansen, "Area coverage rate of synthetic aperture sonars," in *Proc. IEEE OCEANS Europe Conf.*, Aberdeen, Scotland, Jun. 2007, pp. 1–5.

R. E. Hansen, H. J. Callow, and T. O. Sæbø, "The effect of sound velocity variations on synthetic aperture sonar," in *Proc. Underwater Acoust. Meas.*, Crete, Greece, Jun. 2007.

R. E. Hansen, T. O. Sæbø, H. J. Callow, and P. E. Hagen, "Interferometric synthetic aperture sonar in pipeline inspection," in *Proc. IEEE OCEANS Conf.*, Sydney, Australia, May 2010, pp. 1–10.

R. E. Hansen, H. J. Callow, T. O. Sæbø, and S. A. Synnes, "Challenges in seafloor imaging and mapping with synthetic aperture sonar," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3677–3687, 2011.

R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.,Syst.*, no. 6, pp. 610–621, 1973.

R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Boston, MA, USA: Addison-Wesley, 1992, vol. I.

F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

M. H. Hayes, *Statistical digital signal processing and modeling.* New York, NY, USA: Wiley & Sons, 1996.

M. P. Hayes and P. T. Gough, "Synthetic aperture sonar: A review of current status," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 207–224, 2009.

A. Hein, *Processing of SAR Data: Fundamentals, Signal Processing, Interferometry.* Berlin, Germany: Springer-Verlag, 2003.

M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, 2009.

G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, 1968.

J. C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. H. Thompson, *Spotlight Mode Synthetic Aperture Radar: A Signal Processing Approach.* Boston, MA, USA: Kluwer, 1996.

D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

M. Kronig, "Bathymetry processing for synthetic aperture sonar systems," Master's thesis, Dept. Electr. Eng., Technische Universität Darmstadt, 2014.

S. Leier and A. M. Zoubir, "Quality assessment of synthetic aperture sonar images based on a single ping reference," in *Proc. IEEE OCEANS Europe Conf.*, Santander, Spain, Jun. 2011, pp. 1–4.

S. Leier and A. M. Zoubir, "Phasewrap error correction for micronavigation in synthetic aperture systems," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 5915–5918.

S. Leier and A. M. Zoubir, "Time delay estimation for motion compensation and bathymetry of SAS systems," in *Proc. Eur. Signal Process. Conf.*, Bucharest, Romania, Aug. 2012, pp. 2293–2297.

S. Leier, A. M. Zoubir, and J. Groen, "Sequential focus evaluation of synthetic aperture sonar images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 5969–5973.

S. Leier, J. Groen, A. M. Zoubir, U. Hoelscher, and I. Campbell, "The influence of sound speed on synthetic aperture sonar imagery," in *Proc. 1st Int. Conf. Underwater Acoust.*, Corfu, Greece, Jun. 2013.

S. Leier, M. Kronig, and A. M. Zoubir, "A modified version of the mosaic phase gradient autofocus," in *Proc. Eur. Signal Process. Conf.*, Marrakech, Morocco, Sep. 2013.

S. Leier, R. Fandos, and A. M. Zoubir, "Motion error influence on segmentation and classification performance in SAS based automatic mine countermeasures," *IEEE J. Ocean. Eng.*, Mar. 2014, early access.

S. Leier and A. M. Zoubir, "Aperture undersampling using compressive sensing for synthetic aperture stripmap imaging," in *EURASIP J. Advances Signal Process.*, May 2014, under review.

X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications*, 1st ed. New York, NY, USA: Springer-Verlag, 2002.

M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

J. E. McRea Jr, H. G. Greene, V. M. O'Connell, and W. W. Wakefield, "Mapping marine habitats with high resolution sidescan sonar," *Oceanologica Acta*, vol. 22, no. 6, pp. 679–686, 1999.

H. Midelfart, J. Groen, and O. Midtgaard, "Template matching methods for object classification in synthetic aperture sonar images," in *Proc. Underwater Acoust. Meas.*, Nafplion, Greece, Jun. 2009.

M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, 2000.

R. Moddemeijer, "On the determination of the position of extrema of sampled correlators," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 216–219, 1991.

J. W. Oeschger, "Estimating along-track displacement using redundant phase centers," in *Proc. Inst. Acoust. Int. Conf. SAS/SAR*, vol. 28, Lerici, Italy, Sep. 2006, p. 5.

A. V. Oppenheim, R. W. Schafer, J. R. Buck *et al.*, *Discrete-time Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989, vol. 2.

A. Panahi and M. Viberg, "Maximum a posteriori based regularization parameter selection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Prague, Czech Republic, May 2011, pp. 2452–2455.

V. M. Patel, G. R. Easley, D. Healy, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 244–254, 2010.

K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901.

S. W. Perry and L. Guan, "Pulse-length-tolerant features and detectors for sector-scan sonar imagery," *IEEE J. Ocean. Eng.*, vol. 29, no. 1, pp. 138–156, 2004.

P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, pp. 1119–1125, 1994.

R. K. Raney, H. Runge, R. Bamler, I. G. Cumming, and F. H. Wong, "Precision SAR processing using chirp scaling," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 786–799, 1994.

S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 90–105, 2003.

M. A. Richards, *Fundamentals of Radar Signal Processing.* New York, NY, USA: McGraw-Hill, 2005.

L. Rosenberg and D. Gray, "Multichannel SAR imaging with backprojection," in *Proc. IEEE Intell. Sensors, Sensor Networks, Inform. Process. Conf.*, Melbourne, Australia, Dec. 2004, pp. 265–270.

F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov Random Fields and image segmentation," *Graphical Models Image Process.*, vol. 59, no. 4, pp. 205–220, 1997.

T. O. Sæbø, R. E. Hansen, and A. Hanssen, "Relative height estimation by cross-correlating ground-range synthetic aperture sonar images," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 971–982, 2007.

T. O. Sæbø, "Seafloor depth estimation by means of interferometric synthetic aperture sonar," Ph.D. dissertation, Dept. Phys. Technol., Universitetet i Tromsø, 2010.

H. Schmaljohann and J. Groen, "Motion estimation for synthetic aperture sonars," in *Proc. 9th Eur. Conf. Synthetic Aperture Radar*, Nuremberg, Germany, Apr. 2012, pp. 78–81.

K. Siantidis and U. Hölscher-Höbing, "A system for automatic detection and classification for a mine countermeasure AUV," in *Proc. Underwater Acoust. Meas.*, Nafplion, Greece, Jun. 2009.

B. W. Silverman, *Density estimation for statistics and data analysis.* London, U.K.: Chapman Hall, 1986, vol. 26.

M. I. Skolnik, *Introduction to radar systems.* New York, NY, USA: McGraw-Hill, 2001.

M. Soumekh, *Synthetic Aperture Radar Signal Processing: with MATLAB Algorithms.* New York, NY, USA: Wiley & Sons, 1999.

H. C. Stankwitz, R. J. Dallaire, and J. R. Fienup, "Spatially variant apodization for sidelobe control in SAR imagery," in *Rec. IEEE Nat. Radar Conf.*, Atlanta, GA, 1994, pp. 132–137.

I. Stojanovic, W. Karl, and M. Cetin, "Compressed sensing of mono-static and multi-static SAR," in *SPIE Defense and Security Symp., Alg. for SAR Imagery XVI*, Orlando, FL, Apr. 2009.

R. Stolt, "Migration by Fourier transform," *Geophysics*, vol. 43, no. 1, pp. 23–48, 1978.

S. A. Synnes, R. E. Hansen, and T. O. Sæbø, "Assessment of shallow water performance using interferometric sonar coherence," in *Proc. Underwater Acoust. Meas.*, Nafplion, Greece, Jun. 2009.

M. Tello Alonso, P. Lopez-Dekker, and J. J. Mallorqui, "A novel strategy for radar imaging based on compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4285–4295, 2010.

P. Thourel, C. Collet, P. Bouthemy, and P. Pérez, "Multiresolution analysis and MRF modeling applied to the segmentation of shadows in sonar pictures," in *Proc. Asian Conf. Comp. Vision*, vol. 2, Singapore, Dec. 1996, pp. 81–85.

L. M. H. Ulander, H. Hellsten, and G. Stenstrom, "Synthetic-aperture radar processing using fast factorized back-projection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 760–776, 2003.

H. L. Van Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV).* New York, NY, USA: Wiley & Sons, 2002.

F. Vincent, B. Mouton, E. Chaumette, C. Nouals, and O. Besson, "Synthetic aperture radar demonstration kit for signal processing education," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3, Honolulu, HI, Apr. 2007, pp. 709–712.

D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and J. C. V. Jakowatz "Phase gradient autofocus-a robust tool for high resolution SAR phase correction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 827–835, 1994.

D. E. Wahl, J. C. V. Jakowatz, P. H. Thompson, and D. C. Ghiglia, "New approach to strip-map SAR autofocus," in *Proc. IEEE Digital Signal Process. Workshop*, Yosemite National Park, CA, Oct. 1994, pp. 53–56.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

D. P. Williams and J. Groen, "A fast physics-based, environmentally adaptive underwater object detection algorithm," in *Proc. IEEE OCEANS Europe Conf.*, Santander, Spain, Jun. 2011, pp. 1–7.

D. P. Williams, "AUV-enabled adaptive underwater surveying for optimal data collection," *Intell. Service Robot.*, vol. 5, no. 1, pp. 33–54, 2012.

D. P. Williams, A. Vermeij, F. Baralli, J. Groen, and W. Fox, "In situ AUV survey adaptation using through-the-sensor sonar data," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Kyoto, Japan, Mar. 2012.

S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.

J. Yang, J. Thompson, X. Huang, T. Jin, and Z. Zhou, "Random-frequency SAR imaging based on compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 983–994, 2013.

Y.-S. Yoon and M. G. Amin, "Compressed sensing technique for high-resolution radar imaging," in *Proc. SPIE Signal Process., Sensor Fusion, Target Recognition XVII*, vol. 6958, Orlando, FL, Mar. 2008.

D. Youn, N. Ahmed, and G. Carter, "On using the LMS algorithm for time delay estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 5, pp. 798–801, 1982.

B. Zerr, J. Fawcett, and D. Hopkin, "Adaptive algorithm for sea mine classification," in *Proc. Underwater Acoust. Meas.*, Nafplion, Greece, Jun. 2009, pp. 319–326.

S. Zhu, J. Yue, and W. Jiang, "SAS Autofocus Based on Phase Gradient Autofocus," in *Proc. Int. Chaos-Fractals Theories Applicat.*, Hangzhou, China, Oct. 2011, pp. 298–301.

Curriculum vitae

Name:	Stefan Leier
Date of birth:	17.07.1981
Place of birth:	Lich
Family status:	single

Education

04/2006 - 05/2009	Technische Universität Darmstadt, Germany Information- and Communication Engineering Master of Science
02/2007 - 01/2008	Curtin University of Technology, Perth, Australia Graduate scholarship, German Academic Exchange Service (DAAD)
10/2002 - 03/2006	Technische Universität Darmstadt, Germany Information- and Communication Engineering Bachelor of Science
07/1998 - 05/2001	Wolfgang-Ernst Gymnasium, Büdingen, Germany High school degree (Abitur)

Work experience

since 08/2009	Research associate at Signal Processing Group Technische Universität Darmstadt, Germany
05/2012 - 08/2012	Visiting research fellow NATO Undersea Research Centre (NURC) La Spezia, Italy
11/2008 - 05/2009	Master thesis Roche Diagnostics GmbH, Mannheim, Germany
04/2008 - 08/2008	Internship Diehl Aerospace GmbH, Frankfurt am Main, Germany
07/2002 - 09/2002	Internship Siemens AG PTD M, Frankfurt am Main, Germany

Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 17.06.2014