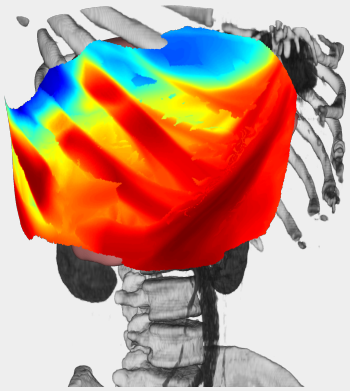
Computer Aided Medical Procedures
Prof. Dr. Nassir Navab



Dissertation

Advances in Medical Ultrasound Image Quality

Rüdiger Göbl

TUM School of Computation,
Information and Technology
Technische Universität München

Technische Universität München
TUM School of Computation, Information and Technology

Advances in Medical Ultrasound Image Quality

Rüdiger Göbl

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:	Prof. Dr.-Ing. Nils Thuerey
Prüfende der Dissertation:	1. Prof. Dr. Nassir Navab
	2. Prof. Frank Lindseth

Die Dissertation wurde am 25.10.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 02.12.2024 angenommen.

Rüdiger Göbl
Advances in Medical Ultrasound Image Quality
Dissertation, Version 1.0.1

Technische Universität München
TUM School of Computation, Information and Technology
Lehrstuhl für Informatikanwendungen in der Medizin
Boltzmannstraße 3
85748 Garching bei München

Abstract

Ultrasound is a versatile medical imaging modality that complements X-ray computed tomography and magnetic resonance imaging. It is safe, low-cost and devices can be exceptionally portable. However, artifacts from the physics governing the propagation of ultrasound waves and the image formation process can limit the image quality. This thesis investigates several advances in medical ultrasound imaging addressing this.

First, a comprehensive review of techniques for ultrasound image reconstruction, specifically for receive beamforming is presented. Conventionally, ultrasound receive beamforming is performed either mechanically (through an acoustic lens), electronically (by means of delay-lines), or in digital logic (field-programmable gate arrays) and consists of a two-step approach: Delaying of signals to create focus and summation of the aperture data to emulate a single ultrasound receiver. The review focuses on algorithms replacing the summation of aperture-data with more refined means and the developments in this field.

To facilitate research in ultrasound imaging algorithms an open framework for software-based ultrasound imaging is presented. There are aspects in an ultrasound imaging setup that explicitly require electronics, such as pulse emission and digitization of received echo signals. All other components are covered in the software framework, from ultrasound sequence pulse programming (also called transmit beamforming), over receive beamforming, envelope detection, log-compression, and scan-conversion. It is easily extensible for other techniques, *e.g.* for the addition of alternative algorithms for comparison.

Depending on the diagnostic task, the presence of speckle in the images—an intrinsic property of ultrasound imaging—can be detrimental for visualization. For this reason algorithms for speckle-removal are employed in virtually all clinical scanners. This thesis investigates synthetic data-generation and training considerations of deep-learning based speckle filters, resulting in a speckle filter that performs favorably without requiring real data for training or manual tuning.

In the presence of large tissue property differences, mainly in the vicinity of bones and air cavities, special care must be taken by the ultrasound system operator to select a so-called acoustic window that allows for sufficient transmission of ultrasonic energy to the anatomy under investigation in order to preserve image quality and visibility of the target anatomy all-together. For robotic ultrasound acquisitions, this thesis presents a technique for fully automatic acquisition planning using a target visibility model based on ultrasound-physics.

As these advances continue to improve the quality, ease of use and applicability of ultrasound as imaging modality, we can be sure to see it employed in an ever growing array of ways, leading to better patient care through better image guidance and non-invasive diagnosis.

Zusammenfassung

Ultraschall ist ein vielseitiges medizinisches Bildgebungsverfahren, das die Röntgen-Computertomographie und die Magnetresonanztomographie ergänzt. Er ist sicher, kostengünstig und Geräte können außergewöhnlich tragbar sein. Allerdings kann die Bildqualität durch Artefakte beeinträchtigt werden, die sich aus den physikalischen Bedingungen für die Ausbreitung der Wellen und dem Bildgebungsprozess ergeben. In dieser Arbeit werden mehrere Fortschritte in der medizinischen Ultraschallbildgebung untersucht, die dieses Problem adressieren.

Zunächst wird ein umfassender Überblick über Techniken zur Rekonstruktion von Ultraschallbildern, insbesondere zum Empfangsbeamforming, gegeben. Konventionell besteht Ultraschall-Empfangsbeamforming aus einem zweistufigen Ansatz: Verzögerung von Signalen, um Fokus zu erzeugen, und Summierung der Aperturdaten, um einen einzelnen Ultraschallempfänger zu emulieren. Die Übersicht konzentriert sich auf Algorithmen, die die Summierung von Aperturdaten durch differenziertere Mittel ersetzen, und die Entwicklungen hierzu.

Zur Unterstützung der Forschung auf dem Gebiet der Ultraschallbildgebungsalgorithmen wird ein freies Framework für softwarebasierte Ultraschallbildgebung vorgestellt. Es gibt Aspekte in einem Ultraschallsystem, die explizit Elektronik erfordern, die Impulserzeugung und die Digitalisierung der Echosignale. Alle anderen Komponenten werden durch das Software-Framework abgedeckt, von der Programmierung von Ultraschallimpulssequenzen über Empfangsbeamforming, Hüllkurvendetektion, Log-Kompression und Scan-Konvertierung. Es ist leicht erweiterbar für andere Techniken, so dass alternative Algorithmen hinzugefügt werden können.

Je nach diagnostischer Aufgabe kann das Vorhandensein von Speckle in den Bildern für die Visualisierung nachteilig sein. Aus diesem Grund werden Algorithmen zur Speckle-Entfernung in praktisch allen klinischen Scannern eingesetzt. In dieser Arbeit werden die Generierung synthetischer Daten und Überlegungen zum Training von Deep-Learning-basierten Speckle-Filtern untersucht. Das Ergebnis ist ein Speckle-Filter, der gute Resultate liefert, ohne dass reale Daten für das Training oder manuelle Einstellungen erforderlich sind.

Bei großen Unterschieden in den Gewebeeigenschaften muss der Ultraschallanwender darauf achten, ein akustisches Fenster zu wählen, das eine ausreichende Übertragung der Ultraschallenergie auf die Anatomie ermöglicht, um die Bildqualität und die Sichtbarkeit zu erhalten. Für robotergestützte Ultraschallaufnahmen wird in dieser Arbeit eine Technik zur automatischen Aufnahmeplanung mithilfe eines physikbasierten Modells der Sichtbarkeit vorgestellt.

Da diese Fortschritte die Qualität, Benutzerfreundlichkeit und Anwendbarkeit des Ultraschalls als bildgebendes Verfahren weiter verbessern, können wir sicher sein, dass er auf immer vielfältigere Weise eingesetzt wird, was zu einer besseren Patientenversorgung durch bessere Navigation und nicht-invasive Diagnose führt.

Acknowledgments

I would like to express my gratitude to my colleagues from the chair for computer aided medical procedures. Most importantly to Prof. Nassir Navab, for the opportunity to pursue my research, his supervision, guidance and encouragement. I am also especially grateful to Dr. Maximilian Baust and Dr. Christoph Hennersperger, for much the same reasons, for introducing me to the field as well as their friendship.

Moreover, I would like to thank my colleagues and friends from around the lab (in randomized order) Marco Esposito, Arianne Tran, Oliver Zettinig, Walter Simson, Beatrice Demiray, Seyed-Ahmad Ahmadi, Javier Esteban, Benjamin Busam, Julia Rackerseder, Fernanda Levy Langsch, Thomas Wendler, Benjamin Frisch, Salvatore Virga, Jakob Weiss, Ardit Ramadani and Maria Tirindelli, for all the hard work, success, setbacks, celebrations and long nights we shared working together.

Finally, I want to thank my father, for his support and always encouraging my curiosity.

Contents

Part I

Introduction

1 Medical Imaging

1.1 Background

Medical imaging allows physicians to asses the status of tissues non-invasively. It is used to diagnose and treat many diseases and injuries, including cancer, cardiovascular diseases, and musculoskeletal injuries. It is also used to monitor the progress of treatments or track disease progression. Medical imaging is essential for the diagnosis of a wide variety of conditions and for helping physicians determine the best course of treatment for their patients. Imaging modalities include X-ray, computed tomography (CT), magnetic resonance imaging (MRI), nuclear medicine such as positron emission tomography (PET), and ultrasound (US).

Depending on their working principle, different modalities allow for structural or functional imaging. Structural imaging provides anatomical information, while functional imaging can be used to measure and monitor physiological processes, such as blood flow or metabolic activity.

Since X-ray and CT imaging are based on the absorption of X-ray radiation, their contrast is related to the tissue densities. This makes them well suited for the diagnosis of bone and joint disorders. With the administration of contrast-agent they can also be used for the assessment of cardiovascular structures, such as the abdominal aorta. In contrast, MRI and PET do not rely on the absorption of X-ray radiation and can provide more details of soft tissues than X-ray and CT. MRI allows for higher soft tissue contrast by measuring different tissue properties, such as the proton density, or diffusion velocities of liquids. PET imaging is often used in cancer treatment, as it can detect tumors not visible on X-ray or CT images. It does, however, require the administration of radioactive tracers, targeted at the specific tissues in question. US imaging reveals structural and functional information using high frequency sound-waves, which are reflected by the tissue interfaces. It can be used to interactively assess the dynamic motion of structures, and provides a good overview of soft tissues, such as the vascular system and many of the abdominal organs. Besides their use for diagnosis, certain imaging modalities, such as X-ray imaging and US can be used for image guidance during interventions for the navigation of tools such as catheters during a procedure, or to minimize risks to the patient, *e.g.* to localize and avoid major vessel structures during liver resections. This allows physicians to perform interventions less invasively.

The interpretation of acquired images requires experience with the modality and in depth knowledge of the anatomy and physiology. For this reason, physicians work closely with specially trained radiologists and technologists to analyze medical imaging studies and to formulate the most accurate diagnosis and treatment plan for their patients. Medical imaging techniques can involve radiation exposure, primarily from X-Rays, CT scans, and nuclear

medicine scans. While the amount of radiation exposure is generally small, the risk of harm from radiation increases with repeated exposures. Furthermore, some medical imaging modalities, such as US imaging, rely on the operator's skill and experience, and poor image quality can hamper an accurate diagnosis. Therefore, it is important for healthcare providers to weigh the risks and benefits of the modalities, before their application.

Medical imaging is an important part of modern medicine, and will continue to be an essential tool. This thesis focuses on ultrasound imaging and advances specifically with respect to image quality.

1.2 Essentials of Ultrasound Imaging

In this section, we provide an overview of the essentials in US image formation as an introduction to the following chapters.

1.2.1 Hardware setup

An US imaging system begins with an US probe, that emits ultrasonic pulses and receives the echoes from tissue boundaries and inhomogeneities therein. The probe is often also referred to as transducer, as it generates a mechanical wave when excited by an electric pulse and vice versa. The transducer is connected to a pulse generator, which is responsible for creating the electric pulses. The transducer also is connected to an US frontend, that amplifies and conditions the received electric waveforms caused by the reflected ultrasonic waves. The signals are then sampled by analog-to-digital converters (ADCs) at an appropriate sampling frequency, followed by digital processing to form the images, which are then shown to the operator on a screen. The operator generally can influence major properties of the imaging process, such as imaging depth, pulse frequency, and visible field of view, through controls on the imaging system.

These components are most often integrated in a mobile imaging console to which one or multiple different probes can be connected, but there are also even more mobile realizations that are integrated more tightly. This can range from US systems in laptop like form factors, to all electronics and processing being integrated in the probe handle that then transfers the US images to a mobile phone for display and operator interaction.

1.2.2 Transducers

Transducers are the most crucial component of the imaging system, as their construction defines the applicable imaging protocols and limitations. They are constructed from one or more US elements. US elements perform the conversion from electrical signals to a pressure wave and back. They can be bulk piezoelectric crystals [35], piezoelectric micromachined ultrasonic transducers (PMUT) [24] or even be constructed from silicon as is the case for capacitive micromachined ultrasonic transducers (CMUT) [82].

Based on the number and arrangement of the elements there are a few general classes of transducers, which can be selected by the operator depending on the imaging task at hand.

- Single element transducers: They are mechanically focused to be sensitive only in a narrow beam. They are often combined with a mechanical motion system to cover an area.

- Linear arrays: A number of elements (usually 64-256) in a straight line, often used for smaller field of views, imaging perpendicular to the line.

- Phased arrays: Similar to linear arrays, but since elements are spaced more closely imaging happens in a fan shape starting at the transducer. They are mainly applied when there is only a small space available for placing the transducer, *e.g.* in cardiac imaging.

- Curvilinear arrays also known as convex arrays: An array of elements placed on a curve with usual radii around 2 to 7 cm. Imaging happens in a fan shape around the transducer. They are often used for very deep acquisitions.

- Matrix arrays: Different from the one-dimensional arrays above, which image within a single plane, matrix arrays allow for volumetric imaging, by combining a large number of elements in a two-dimensional arrangement.

When an array of transducer elements is used, it is common to use multiple elements at the same time during pulse emission or echo reception, in order to improve the imaging performance.

1.2.3 Imaging

The general operation of a medical US system is independent of the transducer type. First, an ultrasonic pulse with a frequency in the MHz-range is generated by applying an electric waveform to the element(s). This pulse travels through the tissue, away from the transducer. The wave is then partially reflected at interfaces between tissues (such as the boundary between sub-cutaneous fat-tissue and a skeletal muscle) and at inhomogeneities within tissues. This can be for example the different fibers in muscle tissue, or blood cells in the otherwise homogeneous plasma.

The reflections can then be detected by the transducer, which converts them to a voltage waveform that is recorded. Knowing the average speed of sound in tissue, the depth at which a reflection occurred can then be deduced by the time difference between the pulse and the reception of the echo. This recorded signal represents a measure of the tissue reflectivity modulated with the US pulse frequency. The reflectivity signal can be recovered by computing the instantaneous amplitude (also called envelope) of the modulated signal. An example for both these signals is shown in Figure 1.1.

For single element transducers, as the spatial sensitivity of the element is designed to be very narrow, this signal is considered to originate from one line, a scan-line. After acquisition of

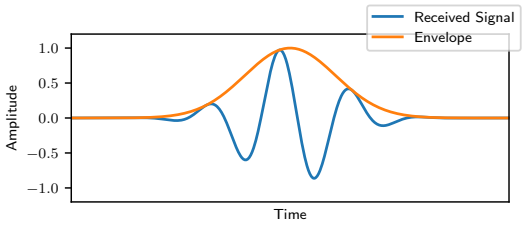


Fig. 1.1. **Example echo signal from a single scatterer** and the corresponding envelope as reflectivity measure.

one scan-line, the transducer can then be repositioned to repeat the process of emitting a pulse and recording the echos to acquire a next line. When arranging the signals received from a number of scan-lines next to each other, the result is an image of the tissue reflectivity.

With array transducers, multiple elements are generally used at the same time to generate a shaped pressure wave and to receive the echos, a process often referred to as beamforming. In the transmission of the pulse, the temporal relationship between the elements' excitations influences the resulting transmit wave, as the waves generated by the individual elements interfere with each other as they travel through the tissue. If the waveforms are delayed such that the spherical waves emitted by each element all constructively interfere in one location, analog to the focusing of a convex lens, the imaging scheme is said to employ focused beams. This narrows the overall spatial sensitivity of the system (during that pulse-echo event) at that depth drastically, increasing the resolution. The signals received by multiple elements are then combined in a process called receive beamforming, in order to further narrow the array sensitivity. Receive beamforming applies delays to the signals, that are similar to the transmit delays and then combines them, effectively emulating interference. A more detailed explanation of receive beamforming is given in chapter 2. Similar to single element transducers, this results in one scan-line. Multiple scan-lines are acquired by selecting a different set of elements to transmit a pulse or by steering the focal point in a different direction. Fig. 1.2 shows the wavefronts for three differently steered focused beams.

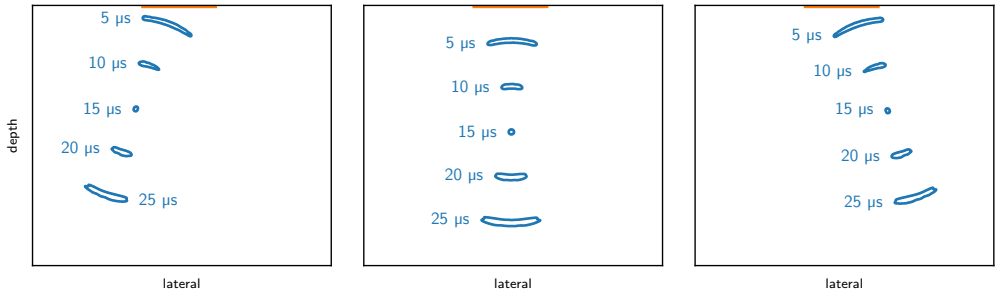


Fig. 1.2. **Focused beam wavefront propagation over time** of three differently steered beams using the same linear array. Contours show the 6 dB level of the instantaneous pressure amplitude at each time point. The transducer location is indicated in orange at the top of the figures.

Besides focused imaging, there are additional schemes that utilize a multitude of elements in an array transducer:

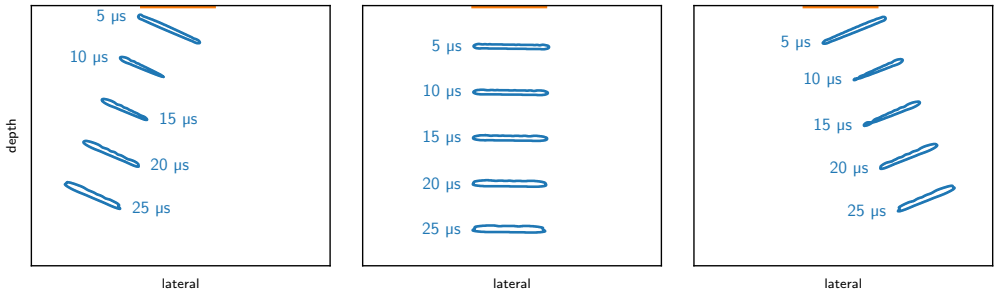


Fig. 1.3. **Planewave propagation over time** of three differently steered planewaves using the same linear array. Contours show the 6 dB level of the instantaneous pressure amplitude at each time point. The transducer location is indicated in orange at the top of the figures.

- Synthetic aperture imaging schemes cover the area of interest by performing a sequence of pulse-echo events, that each emit a pulse from one of the transducer elements followed by the reception of the echos on all transducer elements. By then combining the recorded waveforms from all pulse-echo events with appropriate delays, it is possible to *virtually* focus the transmit wave on each point separately, improving the focusing compared to focused beams.

- Planewave imaging intentionally broadens the transmit beam by emitting (tilted) planar waves using all elements. This allows for the imaging of a larger region of the tissue than just a line, reducing the number of required pulse-echo events to cover a certain area drastically. Fig. 1.3 shows three planewaves with different steering angles. While a single planewave could cover the area of interest, planewave imaging does reduce the image quality compared to focused imaging. This is counteracted by using a few planewave events with different tilts, and combining the resulting images, leading to a net decrease in the required pulse-echo events, increasing the frame rate.

- Diverging wave imaging is similar to planewave imaging in that it broadens the transmitted beam, but instead of creating planar waves, it creates spherical waves where the source of the wave is placed behind the array. Since this is only emulated by delaying the single element transmissions appropriately, this source behind the array is called a virtual source. Its distance from the array and tilt relative to it defines the opening angle of the beam and its direction. Fig. 1.4 shows the beam patterns of three diverging waves. As with planewave imaging, diverging wave imaging allows the reduction of the number of pulse-echo events, leading to a higher imaging frame rate.

These schemes as well as focused imaging can be applied to one-dimensional arrays as well as matrix arrays to form planar and volumetric images respectively.

1.2.4 Challenges and Limitations

For all its advantages, medical US imaging is not without challenges. US is subject to a number of effects inherent to the modality.

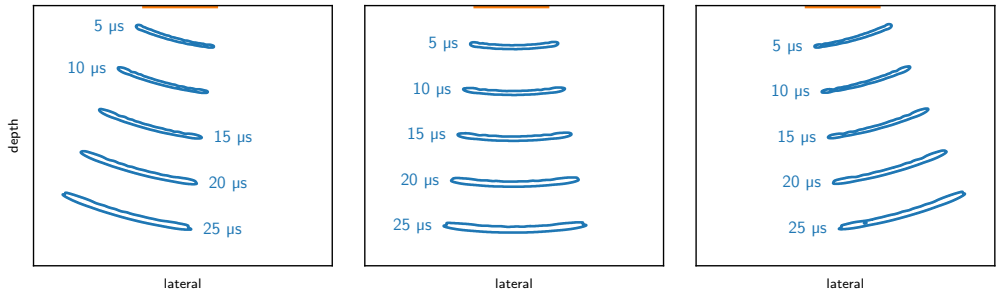


Fig. 1.4. **Diverging wave propagation over time** of three differently steered beams using the same linear array. Contours show the 6 dB level of the instantaneous pressure amplitude at each time point. The transducer location is indicated in orange at the top of the figures.

First, as the pulse is emitted from the same location that records the echos, the tissue appearance depends on the direction of imaging. This can require the operator to manipulate the probe until the optimal direction is found, which can be time consuming in certain clinical tasks.

Second, acoustic shadows caused by high tissue density differences can obscure parts of the image. This heavily influences structures located near and especially behind bone, calcified tissue or air filled cavities and can make them difficult to image.

The third effect inherent to US imaging is speckle, caused by the constructive and destructive interference of multiple reflected waves. This creates a granular texture in the image, even for tissues that appear homogeneous on the macro-scale.

Finally, since virtually all imaging schemes require the recording of multiple pulse-echo events to form a single image, the resulting images can be degraded by tissue motion during the execution of the sequence of pulse-echo events. This is of particular importance for fast moving organs such as the heart, often limiting the acquisition schemes that can be employed.

This thesis investigates a number of approaches to alleviate some of these limitations, in order to advance the technology and its use. The remainder of the manuscript is organized as follows: First, we perform a comprehensive review of techniques for US image reconstruction in chapter 2, specifically focused on receive beamforming. This is followed by the presentation of an open framework for software-based US imaging that aims at facilitating research in US imaging algorithms in chapter 3. In chapter 4, we present an approach for synthetic data-generation and training considerations of a deep-learning based speckle filter that improves the appearance of tissue structures after acquisition. Finally, we investigate in chapter 5 an automatic acquisition planning approach for robotic US imaging using a target visibility model based on ultrasound-physics.

2 Review of Receive Beamforming Techniques

2.1 Introduction and Scope

With the advent of software-based US imaging pipelines and the increase in readily available compute performance, research in medical ultrasound imaging has made significant progress in the last years. The developments reached all parts of the imaging pipelines, from beamforming to image filtering, with the most notable change being ultrafast imaging based on plane-wave (PW) or synthetic aperture (SA) imaging.

In most applications, medical US is performed using transducer arrays to allow for electronic scanning—that is the imaging of a larger area than possible with only a single element—and to improve the image quality and field of view by using multiple transducer elements—the so-called aperture—to send and receive ultrasonic pulses, generally addressed with the terms beamforming and beam steering. Beamforming is applied twice in one pulse-echo event; First, during transmit, it is used to shape the created pulse by using interference within the medium to be imaged and exciting the individual transducer elements with different waveforms and delays w.r.t. each other. Second, after reception of the resulting echos through multiple transducer elements, by adequately delaying and combining the received signals, influencing the spatial sensitivity of the reception, among other things. Consequently, the first is usually referred to as transmit beamforming, while the latter is called receive beamforming. Transmit beamforming creates outgoing waveforms via the physical interference of each transducer element's spherical wave. Receive beamforming uses computational interference across the aperture to amplify in-phase signals and dampen out-of-phase interference and noise. This principle of (computational) interference gives rise to the well-known delay and sum (DAS) beamformer, emulating the physical superposition. It is however possible to combine the delayed signals across the aperture in different ways in order to change the outcome of the beamforming process.

An overview of the development of transmit and receive beamforming systems from analog to digital implementations is given by Thomenius *et al*. [110]. Lu *et al*. [68] review different methods of beam focusing and the influence different parameters have on the beam shape. Demi [25] gives an introduction to the different commonly used image acquisition schemes ranging from multi line acquisition and transmission to PW and SA imaging and their trade-offs. Wang *et al*. [118] summarize several manuscripts dealing with aperture data (*i.e*. data at the same state as this review), specifically around phase aberration correction, sidelobe reduction, and vector velocity estimation. Hasegawa provides an overview of recent developments in beamforming on a broader scope [36], not limited to aperture data processing only.

The focus of this review is on methods that are applied to the delayed signals after reception, also called aperture data processing. This excludes the fixed transmit beamforming of *e.g.* focused imaging (calculation of delays and waveforms to apply to transducer elements), as well as the calculation and application of delays after reception (dynamic focusing in transmit or receive). In short, for the purposes of this review, the received data has already been shifted appropriately. We want to provide an overview of methods to combining the data, that are more than just the sum.

In section 2.2 we discuss the beamforming algorithms, grouped by their general theory of operation, followed by a comparison and discussion of their respective advantages and potential shortcomings in section 2.3. Finally, in section 2.4 we give a brief conclusion and outline a possible path for future development opportunities.

Work	Group	Main contribution
Hollman, [42]	CF	Introduces coherence factor (CF)
Li, [65]	CF	Generalizing CF to reduce sensitivity to phase abberations
Wang, [117]	CF	Adaptive weighting of CF to reduce sensitivity to noise
Camacho, [13]	CF	Uses coherence of phase instead of signal amplitude (PCF)
Hasegawa, [38]	CF	Non-overlapping subapertures improve speckle appearance in PCF
Xu, [125]	CF	Spatio-temporal smoothing of CF improves speckle visualization
Lediju, [62]	SLSC	Introduces short-lag spatial coherence imaging
Bottenus, [11]	SLSC	Synthetic aperture increases depth-of-field vs. focused beams
Li, [66]	SLSC	Applies concept of SLSC to planewave TX beamforming
Zhao, [131]	SLSC	Combines SLSC in RX with MV beamforming to improve resolution
Hyun, [45]	SLSC	Demonstrates advantages of SLSC for 3D imaging
Hyun, [44]	SLSC	Modifications that reduce computational complexity drastically
Matrone, [75]	DMAS	Introduces delay multiply and sum to ultrasound imaging
Matrone, [74]	DMAS	DMAS with THI increases resolution
Matrone, [73]	DMAS	DMAS suppresses multi line acquisition artifacts
Matrone, [76]	DMAS	Synthetic aperture increases homogeneity vs. focused beams
Shen, [102]	DMAS	DMAS after IQ-demodulation reduces computational complexity
Synnevåg, [107]	MV	Applies and adapts minimum variance to medical ultrasound
Vignon, [112]	MV	Multiple focused beams for cov. matrix estimation
Synnevåg, [106]	MV	Temporal averaging of cov. matrix improves speckle appearance
Salari, [96]	MV	Compute diagonal loading, subarray and temp. averaging lengths
Deylami, [28]	MV	Solve MV optimization iteratively to reduce comp. requirements
Nilsen, [81]	MV	Solve MV in low-dim. beam-space for reduced computations
Deylami, [27]	MV	Use DCT matrix to span beam-space for increased performance
Kim, [54]	MV	Construct subspace through PCA of conventional MV weights
Asl, [4]	MV	Project MV weights on cov. matrix eigenvectors for contrast
Synnevåg, [105]	Other	Reduce MV solution space to discrete set for performance
Chernyakova, [16]	Other	Fast statistical beamforming solution iMAP improves contrast
Lucies, [70]	DNN	Use neural networks to suppress clutter in aperture data
Hyun, [43]	DNN	DNN trained on DAS and subaperture data to reduce speckle
Luijten, [71]	DNN	FCNN to approximate MV weights and reduce comp. requirements
Wiacek, [123]	DNN	FCNN to estimate SLSC provides speedup

Tab. 2.1. Selection of discussed works with their main contributions.

2.2 Beamforming Techniques

Included in the review are a number of beamforming methods that can be grouped based on their underlying working principles. Besides *DAS* that emulates physical interference through

the (weighted) sum, we identified families of algorithms based on signal *coherence* within the aperture data (such as short-lag spatial coherence (SLSC)) and the related *delay multiply and sum (DMAS)* that employs cross-correlations of the signals to improve image contrast. The *minimum variance (MV)* methods perform an optimization of the weight vector in order to reject signals not originating from the desired location. Finally, we also discuss a number of works that stand on their own, either because they are application-specific or have not been explored deeply in the body of work as of now. Table 2.1 provides an at-a-glance overview of select discussed works and their main contributions.

Notation used throughout this chapter: Matrices $\mathbf{A}$ and vectors $\mathbf{x}$ are denoted by boldface letters, uppercase and lowercase, respectively. Scalar quantities are kept in regular font weight. The expressions A_{ij} and x_i with indices i, j denote the element (i, j) of matrix $\mathbf{A}$ and the i-th element of vector $\mathbf{x}$.

More specifically, we use $\mathbf{s}(t)$ to describe the aperture signal vector for one reconstruction sample in the output image. It is the data per element, after the application of focusing delays, where $s_i(t)$ is the signal of the i-th transducer element and the number of receive channels is N (*i.e.* $\mathbf{s}(t)$ has dimension $(N, 1)$). The result of a beamformer is usually denoted with $y(t)$. We may omit the time parameter t where it is not required.

Unless noted otherwise, all shown images are scaled symmetrically around zero to show the range $[-i_{\text{ext}}, i_{\text{ext}}]$, where $i_{\text{ext}} = \max_{x\in\Omega} |I(x)|$ is the extreme value of the image I and the lowest and highest values in the range correspond to black and white respectively.

2.2.1 Delay and Sum

As mentioned before, DAS is the simplest beamformer. It gained its popularity due to the possibility of simple implementation, first, through analog circuitry which was later replaced by digital implementation [110]. It can be expressed as

$$y_{DAS}(t) = \sum_{i=1}^{N} s_i(t), \tag{2.1}$$

although it is common to weight the contributions of the different elements through so-called apodization

$$y_{\text{DAS}}(t) = \sum_{i=1}^{N} w_i(t) s_i(t), \tag{2.2}$$

where $\mathbf{w}$ is a predefined weight or apodization vector. In essence, DAS emulates a single large mechanically focused US element with the advantage of allowing for dynamic focusing. Figure 2.1 shows the application of DAS to example data, starting from synthetic raw channel data, through aperture data to the beamformed result.

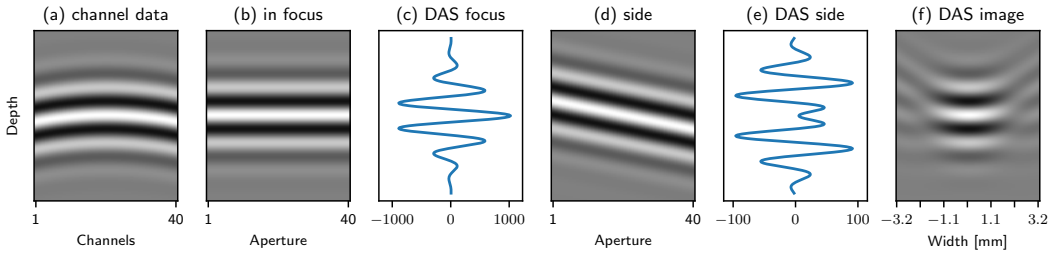


Fig. 2.1. **Visual illustration of DAS. (a)** Channel data from echo of a point reflector, as if received by a linear ultrasound transducer. **(b)** Section of aperture data after application of dynamic receive focus delays for central broadside beam, the reflector is in focus. **(c)** Result of DAS for reflector in focus (one scanline in depth direction). **(d)** Aperture data section with reflector $2.2\,\mathrm{mm}$ laterally of the focal point. **(e)** Result of DAS for out of focus reflector. **(f)** Resulting DAS image. The shown aperture data is used in the subsequent beamformer illustrations as input as well.

2.2.2 Coherence based Methods

Coherence based methods can be placed into two groups. The coherence factors on the one hand use some measure of coherence across the aperture to weight the result of another beamformer. SLSC methods on the other hand are presented as standalone beamformers, their output is considered the image.

Coherence Factors The coherence based beamforming methods generally use the coherence across the aperture as a measure of focusing quality. The coherence factor (CF) can be introduced as

$$\mathrm{CF}(t) = \frac{\left|\sum_{i=1}^{N} s_i(t)\right|^2}{N \sum_{i=1}^{N} |s_i(t)|^2}, \tag{2.3}$$

representing the ratio of coherent energy and incoherent energy [42]. The CF assumes a

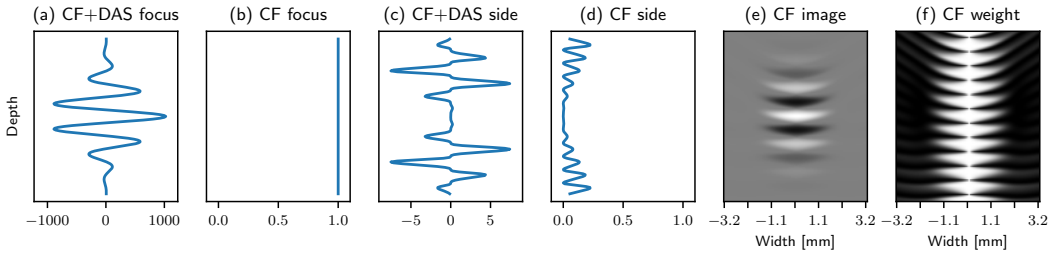


Fig. 2.2. **Visual illustration of CF. (a)** Result of CF for reflector in focus. **(b)** Weighting factor for reflector in focus. **(c)** Result of CF for out of focus reflector. **(d)** Weighting factor for out of focus reflector. **(e)** Resulting CF-weighted image. **(f)** Weight image (range $[0, 1]$).

value of one for fully coherent signals (*e.g.* when all s_i are equal) and its value is reduced in the presence of phase errors or sidelobe signals. While it was suggested originally as means to quantify focusing quality for use in aberration correction, several methods employ it as adaptive weighting together with a different beamformer. Figure 2.2 shows the use of CF in this fashion as well as the resulting weighting.

The CF on its own can be sensitive to small phase aberrations and for diffuse scattering it changes the appearance of speckle when used as weighting factor. To ease this constraint the generalized coherence factor (GCF) was introduced [65], based on the Fourier transform of the aperture data $S_k(t)$, where k is the spatial frequency. Instead of only considering the coherent energy, the GCF includes the energy of the low frequency regions $-M_0 \leq k \leq M_0$

$$\mathrm{GCF}(t) = \frac{\sum_{k=-M_0}^{M_0} |S_k(t)|^2}{\sum_{k=-N/2}^{N/2-1} |S_k(t)|^2}. \tag{2.4}$$

Based on their results, the authors suggest to choose the cutoff frequency M_0 in the range

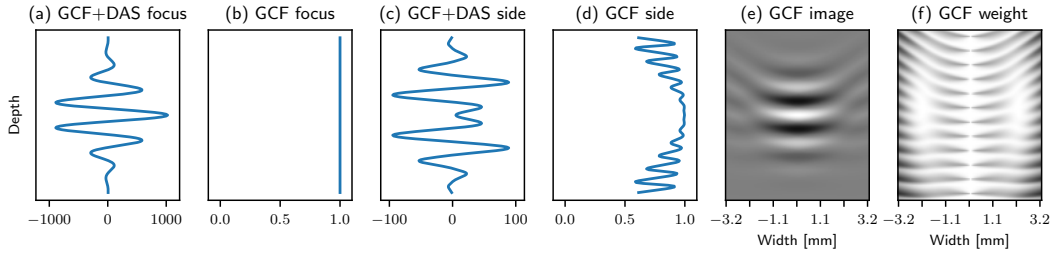


Fig. 2.3. **Visual illustration of GCF** with $M_0 = 1$. **(a)** Result of GCF for reflector in focus. **(b)** Weighting factor for reflector in focus. **(c)** Result of GCF for out of focus reflector. **(d)** Weighting factor for out of focus reflector. **(e)** Resulting GCF-weighted image. **(f)** Weight image (range $[0, 1]$).

of 1-3 for diffuse scattering, for $M_0 = 0$ it is equivalent to the CF. Figure 2.3 illustrates the method and the resulting weights. The GCF was evaluated for use in SA imaging [64] with the effect of improved contrast. In breast imaging the GCF showed great promise in tackling the artifacts present in this challenging anatomy [119].

A different approach to preserve the image appearance in speckle regions was proposed by Wang *et al.* [117]. They provide a method to adaptively strengthen or weaken the application of the CF based on the local signal to noise ratio (SNR), where the signal power is estimated from the low frequency regions $-M_0 \leq k \leq M_0$ of the aperture spectrum (similar to GCF) and the noise power from the remaining aperture spectrum.

Camacho *et al.* propose the phase coherence factor (PCF) that aims to make coherence estimation more robust by using phase information instead of the signal amplitude [13]

$$PCF(t) = max\left(0, 1 - \frac{\gamma}{\sigma_0}\sigma(\varphi_i(t))\right), \tag{2.5}$$

where γ is a sensitivity parameter, $\sigma_0 = \pi/\sqrt{3}$ is the standard deviation of a uniform distribution in $[-\pi, \pi]$, and $\sigma(\varphi_i(t))$ is the standard deviation of the phases of $s_i(t)$ after eventual corrections for phase wraparounds. In their evaluation, the authors choose $\gamma = 1$. This value is also used in the example plots in Figure 2.4. With the sign coherence factor (SCF), they also propose a simplification with still good performance yet allowing for simplified implementation

$$\mathrm{SCF}^p(t) = \left|1 - \sqrt{1 - \left[\frac{1}{N}\sum_{i=1}^{N} sgn(s_i(t))\right]^2}\right|^p, \tag{2.6}$$

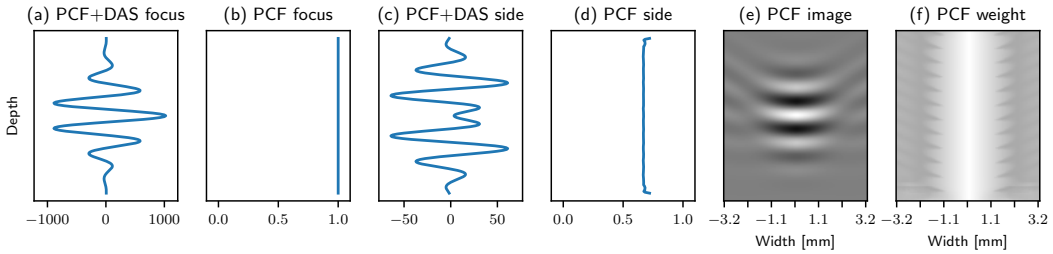


Fig. 2.4. **Visual illustration of PCF** with $\gamma = 1$. **(a)** Result of PCF for reflector in focus. **(b)** Weighting factor for reflector in focus. **(c)** Result of PCF for out of focus reflector. **(d)** Weighting factor for out of focus reflector. **(e)** Resulting PCF-weighted image. **(f)** Weight image (range $[0, 1]$).

where p is a sensitivity parameter. Figure 2.5 illustrates SCF on example data. Unfortunately,

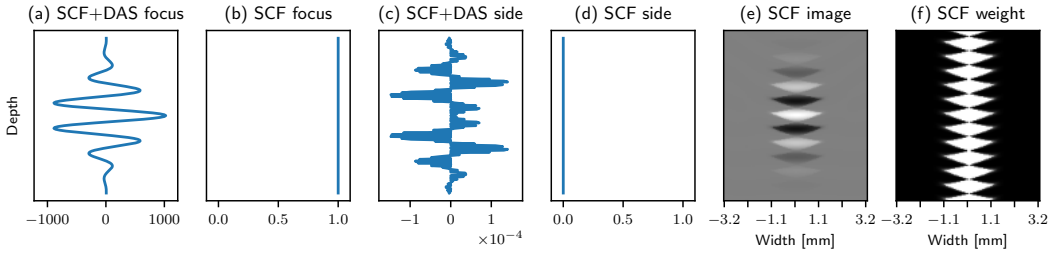


Fig. 2.5. **Visual illustration of SCF** with $p = 1$. **(a)** Result of SCF for reflector in focus. **(b)** Weighting factor for reflector in focus. **(c)** Result of SCF for out of focus reflector. **(d)** Weighting factor for out of focus reflector. **(e)** Resulting SCF-weighted image. **(f)** Weight image (range $[0, 1]$).

the authors did not directly compare the PCF or SCF with CF.

Hasegawa *et al.* investigated the influence of (overlapping and not overlapping) subaperture beamforming on the undesired suppression of echoes from diffuse scattering regions in PCF [38]. They found that by using not overlapping apertures, the speckle suppression can be avoided without degradation of the spatial resolution. Besides the linear nature of the PCF w.r.t. the phase standard deviation in (2.5), two other functions were suggested [37].

To address the dark regions in the proximity of strong scatterers (*e.g.* strongly hyperechoic inclusion) observed with CF, the use of spatio-temporal smoothing of the CF terms was suggested [125]. The authors propose the averaging of the numerator and denominator of (2.3) over several time points t and the use of prebeamforming of the signals with overlapping subapertures. They report a decrease in the dark region artifacts and a reduction in speckle variance. The influence of these modifications on the spatial resolution were not evaluated.

Short-lag Spatial Coherence Imaging Contrary to the coherence based weighting factors, the SLSC is proposed as an imaging method by itself [62]. It uses the normalized spatial correlation $\tilde{R}(m, t)$ between elements of distance m

$$\tilde{R}(m,t) = \frac{1}{N-m} \sum_{i=1}^{N-m} \frac{\sum_{t=t_1}^{t_2} s_i(t) s_{i+m}(t)}{\sqrt{\sum_{t=t_1}^{t_2} s_i^2(t) \sum_{t=t_1}^{t_2} s_{i+m}^2(t)}}, \tag{2.7}$$

where the correlation kernel length $t_2 - t_1$ is suggested to be around one wavelength. The spatial correlation is then accumulated over the first M lags, giving rise to the SLSC

$$R_{\text{sl}}(t) = \sum_{m=1}^{M} \tilde{R}(m, t). \tag{2.8}$$

The SLSC and the spatial correlations are shown in Figure 2.6 for the example data. While point targets are challenging for it, the authors show great improvements in contrast to noise ratio (CNR) and SNR compared to DAS in focused imaging. This advantage was found to

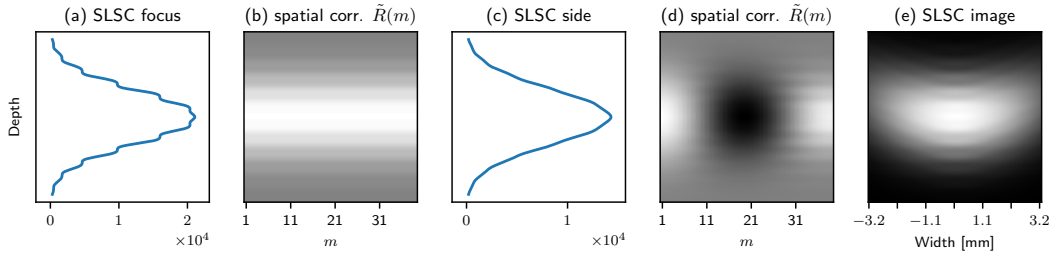


Fig. 2.6. **Visual illustration of SLSC** with $M = 8$. **(a)** Result of SLSC for reflector in focus. **(b)** Spatial correlation (eq. (2.7)) for all lags, reflector in focus. **(c)** Result of SLSC for out of focus reflector. **(d)** Spatial correlation for all lags, out of focus reflector. **(e)** Resulting SLSC image (asymmetric range $[0, max_{x \in \Omega} I(x)]$).

persist even in the presence of high amplitude channel noise [20].

When combined with tissue harmonic imaging (THI) (*i.e.* computing (2.8) on the filtered signals containing the harmonics) in in-vivo settings, it was found to outperform fundamental SLSC as well as THI B-mode imaging w.r.t. clutter removal, especially on patients where B-mode imaging was difficult [21, 47].

As SLSC depends on the focus quality, it can show significant brightness variations across the image depth in focused imaging. In SA this effect is not as predominant [11], where a great improvement in the depth-of-field of the SLSC imaging is shown compared to scanline imaging. Li *et al.* [66] suggested an approach orthogonal to that. After discussing the theory of angular coherence, they propose an application with the short-lag angular coherence (SLAC) beamformer that uses the coherence across different individually DAS beamformed PW images—in essence using SLSC for TX beamforming. While this is only a brief excursion in their theoretically focused manuscript, the results seem very comparable to using SLSC in receive (RX). Combining SLSC as RX beamformer with MV as TX beamformer improves the lateral resolution compared to the DAS TX beamformer with RX SLSC, bringing it roughly on par with that of DAS for TX and RX, while generally retaining the advantages of SLSC [131]. Using the full SA dataset in a dual aperture approach, that is MV in TX with SLSC in RX and MV in RX with SLSC in TX followed by averaging the two resulting $R_{\text{sl RX}}$ and $R_{\text{sl TX}}$ was shown to further improve contrast and CNR [87].

When applying SLSC to volumetric imaging with matrix arrays, its benefits are even more pronounced than for one-dimensional arrays [45]. The authors observed a slight decrease in CNR when prebeamforming (not overlapping) is applied—as is common in matrix array implementations—but it resulted in a slight increase in SNR and most importantly in a

significant reduction in the required computations. The benefits of SLSC in 3D imaging were confirmed in phantom and hepatic in-vivo imaging [46], where a more clear depiction of vessel structures was observed compared to DAS. Addressing the computational complexity of the SLSC, Hyun *et al.* [44] evaluated modifications to the method—namely: a shorter temporal kernel, a downsampled receive aperture and a modified correlation estimator—that reduced the improved computational speed by a factor of 20 in their in-vivo evaluation while the image quality was not impacted significantly.

2.2.3 Delay Multiply and Sum

Similar to the use of the normalized correlation in the SLSC, DMAS uses the auto-correlation across the aperture to create an image. It was first introduced to US for application on radio-frequency (RF) signals [75]

$$\begin{aligned} \hat{s}_{ij}(t) &= sgn(s_i(t)s_j(t)) \cdot \sqrt{|s_i(t)s_j(t)|} \\ y^*_{\text{DMAS}}(t) &= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{s}_{ij}(t) \\ y_{\text{F-DMAS}} &= y^*_{\text{DMAS}} * h_{\text{BP}} \end{aligned} \tag{2.9}$$

using $\hat{s}_{ij}$ as "signed" square root of the product of s_i and s_j, and the bandpass filter h_{BP} to obtain the beamformed signal $y_{\text{F-DMAS}}$. The signal y^*_{DMAS} is considered an intermediate, as it contains both a DC component as well as the desired component at $2f_0$. This is caused by the (multiplicative) mixing of two signals with a center frequency of f_0. For that reason, it is suggested to filter it with a bandpass centered at $2f_0$ with a bandwidth of at least the probe bandwidth, giving rise to the filtered delay multiply and sum (F-DMAS). Figure 2.7 shows this and the resulting F-DMAS image based on the example data, highlighting the increase in spatial frequency. It is worth noting, that before applying (2.9), one needs to ensure that $s_i(t)$

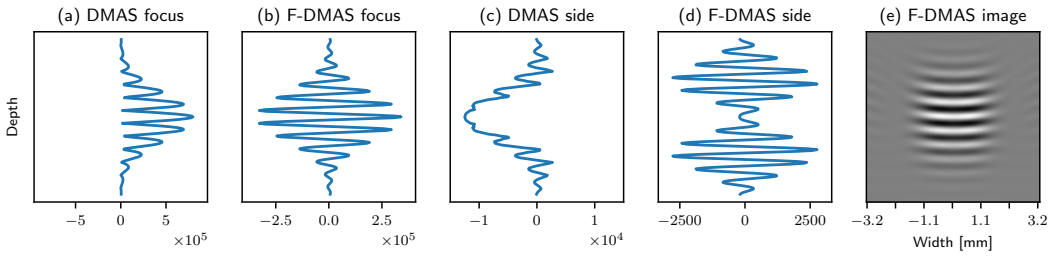


Fig. 2.7. **Visual illustration of DMAS. (a)** Result of unfiltered DMAS for reflector in focus. **(b)** Result of F-DMAS for reflector in focus. **(c)** Result of unfiltered DMAS for out of focus reflector. **(d)** Result of F-DMAS for out of focus reflector. **(e)** Resulting F-DMAS image.

is sampled with a sampling frequency sufficient for the increased signal frequencies resulting from the mixing. Compared to DAS, the authors show an increase in contrast ratio (CR) and lateral resolution in focused imaging, the CNR however was decreased by the increased speckle variance. Applying this concept to THI logically shifts the band of interest to the fourth harmonic. Matrone *et al.* [74] show that in THI the advantages of DMAS carry over from fundamental imaging, resulting in an similar improvement of CR and lateral resolution compared to DAS-based THI. It was also found beneficial in multi line acquisition schemes,

where DMAS beamforming of data containing four transmit beams was found to compare favorably to single line transmission data processed with DAS [73], with the benefit of reduced acquisition time.

Like SLSC, DMAS is influenced by the focus quality, which in focused imaging leads to a change in the speckle pattern in regions besides the focal point. In SA this effect is alleviated, leading to more homogeneous brightness and speckle pattern in the whole image, while retaining the advantages of DMAS [76]. The general influence of signal coherence and the resulting DMAS image properties have also been confirmed theoretically [86].

Su *et al*. [103] extend the concept of DMAS to include (overlapping) subaperture averaging, which reduces speckle variance and thus increases the CNR while also reducing the computational requirements. Another modification to reduce the processing time is to to apply DMAS in the baseband, *i.e*. on in-phase and quadrature (IQ) data, instead of RF-data [102]. Since no shift in frequency is induced in the baseband, the sample rate requirements are relaxed and no bandpass is necessary.

2.2.4 Minimum Variance

The MV or Capon beamformer [15] (named after J. Capon) was introduced for narrowband applications like radar. It formulates the beamforming process as an optimization problem, where the overall output signal power is minimized while ensuring the signal from the desired direction is kept unchanged (distortionless), thus aiming to reject all other signals. This is achieved by computing a weight vector $\mathbf{w}(t)$ for each output sample depending on the aperture data $\mathbf{s}(t)$

$$\begin{aligned} \mathbf{s}(t) &= [s_1(t), s_2(t), \cdots, s_n(t)]^T \\ y_{\text{MV}}(t) &= \sum_{i=1}^{N} w_i(t) s_i(t) = \mathbf{w}(t)^H \mathbf{s}(t). \end{aligned} \tag{2.10}$$

The optimization problem is then formulated in terms of the spatial covariance matrix $\mathbf{R}(t) = E[s(t)s(t)^H]$ of the received aperture data $\mathbf{s}(t)$ as

$$\begin{aligned} &\arg\min_{\mathbf{w}(t)} \mathbf{w}(t)^H \mathbf{R}(t) \mathbf{w}(t), \\ &\text{s.t. } \mathbf{w}(t)^H \mathbf{a} = 1, \end{aligned} \tag{2.11}$$

where $\mathbf{a}$ is the steering vector. In narrowband applications this vector would contain the per element phase shifts that lead to the desired focal point. When working on already delayed aperture data, the array steering direction of interest is the broadside, thus $\mathbf{a}$ is a vector of N ones. The solution to this constrained optimization problem is given by

$$\mathbf{w}(t) = \frac{\mathbf{R}(t)^{-1}\mathbf{a}}{\mathbf{a}^H \mathbf{R}(t)^{-1}\mathbf{a}}. \tag{2.12}$$

When trying to apply this method directly to US imaging, one faces two challenges. First, the spatial covariance matrix $\mathbf{R}(t)$ is unknown and needs to be estimated from the available data. A good estimate is required for robust performance of the beamformer. The second

challenge is computational complexity, which stems from the non-stationary nature of the covariance matrix, that makes it necessary to estimate $\mathbf{R}(t)$ and compute (2.12) for almost every temporal sample. A lot of the works applying MV consequently focus on these two aspects.

Du *et al*. [29] give an overview of user parameter free adaptive beamforming algorithms from the general point of view, *i.e*. not limited to US. After presenting the algorithms, they also analyze the influence of imperfections and noise that are common when applying the algorithms in practice. Here, we focus on applications of these and further methods to medical US and developments made specifically for it.

In their very influential work, Synnevåg *et al*. apply the minimum variance beamformer to medical US imaging [107]. To achieve robust operation, they employ subarray averaging and diagonal loading of the covariance matrix and explore the effects thereof, in settings with and without speed of sound errors.

The concept of subarray averaging can be interpreted as dividing the aperture data of N elements into $N - L + 1$ overlapping arrays of length L

$$\bar{\mathbf{s}}_l(t) = [s_l(t), s_{l+1}(t), \cdots, s_{l+L-1}(t)]^T \tag{2.13}$$

which then can be viewed as multiple observations of aperture data of that reduced length. The sample covariance matrix $\hat{\mathbf{R}}(t)$ is then computed from these observations

$$\hat{\mathbf{R}}(t) = \frac{1}{N-L+1} \sum_{l=1}^{N-L+1} \bar{\mathbf{s}}_l(t)\bar{\mathbf{s}}_l(t)^H, \tag{2.14}$$

with the matrix $\hat{\mathbf{R}}(t)$ being of size $L \times L$. To ensure $\hat{\mathbf{R}}(t)$ is invertible, they suggest to choose $L \leq N/2$.

In addition to the subarray averaging, they apply so called diagonal loading to the covariance matrix, proportional to the power of the received signals in order to increase the robustness of the beamformer

$$\begin{aligned} \tilde{\mathbf{R}}(t) &= \hat{\mathbf{R}}(t) + \epsilon\mathbf{I} \\ \epsilon &= \Delta \; tr(\hat{\mathbf{R}}(t)) \end{aligned} \tag{2.15}$$

where $\mathbf{I}$ is the $L \times L$ identity matrix and Δ the constant loading factor, chosen lower than $1/L$.

Using the covariance matrix estimate $\tilde{\mathbf{R}}(t)$ in (2.12) results in a weighting for the subaperture data—$\mathbf{w}(t)$ is of length L. The beamformed result is then computed as the average of the beamformed subapertures

$$y_{\text{Synnevåg MV}}(t) = \frac{1}{N-L+1} \sum_{l=1}^{N-L+1} \mathbf{w}(t)^H \bar{\mathbf{s}}_l(t). \tag{2.16}$$

They show that MV applied in such a manner increases spatial resolution and reduces sidelobe clutter compared to DAS. Moreover they found that large L can lead to an underestimation of the peak amplitudes. This can be addressed by choosing a smaller subaperture length, which also has the effect of increasing the robustness, *e.g.* against speed of sound errors, but it limits the lateral resolution gains. Increasing the diagonal loading had similar effects. In addition to the effects on the resulting image, the application of subarray averaging also reduces the size of the matrix in the optimization problem, reducing the computational requirements drastically.

In order to further improve the robustness and speckle appearance, Synnevåg *et al.* suggest to employ temporal averaging of the covariance matrix estimate $\hat{\mathbf{R}}(t)$ over a temporal window of approximately one wavelength [106], in addition to the spatial averaging and diagonal loading

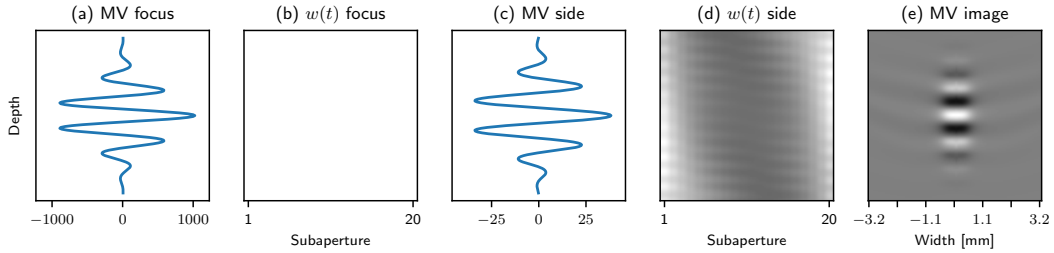


Fig. 2.8. **Visual illustration of MV** with temporal averaging [106], $N = 40, L = 20, \Delta = 1/(100L)$. **(a)** Result of MV for reflector in focus. **(b)** Computed weights $w(t)$ for reflector in focus ($\mathbf{w}(t) = [\frac{1}{L}, ..., \frac{1}{L}]$ uniform across the aperture). **(c)** Result of MV for out of focus reflector. **(d)** Computed weights $w(t)$ for out of focus reflector. **(e)** Resulting MV image.

$$\hat{\mathbf{R}}_{\text{temp}}(t) = \frac{1}{2K+1} \sum_{k=-K}^{K} \hat{\mathbf{R}}(t-k), \tag{2.17}$$

where K is the half-length of the temporal window (*i.e.* half the wavelength) and $\hat{\mathbf{R}}_{\text{temp}}(t)$ the temporally averaged matrix. After application of diagonal loading—in their experiments with $\Delta = 1/(100L)$—the estimate is then used in (2.12). This is illustrated in Figure 2.8 for the example data.

Vignon *et al.* [112] propose to apply MV to aperture data from focused acquisitions as well. However, instead of subarray averaging, they suggest using data from multiple focused transmits in a virtual source imaging setting to estimate the signal covariance matrix. In their evaluation, they employ $N/4$ transmits for that purpose and report increased spatial resolution and contrast compared to DAS.

In [78], Mehdizadeh *et al.* present an adaptive technique to detect portions of the aperture that are affected by shadowing (*e.g.* from proximity of a bone) and to remove those from the MV calculations, which in their experiments improves its performance in such situations especially by reducing signal cancellation. Element shading is detected from the temporally averaged per element energy

$$p_i(t) = \frac{1}{2K+1} \sum_{k=-K}^{K} \left| s_i(t-k) \right|^2, \tag{2.18}$$

from which a new contiguous aperture to be used in MV is selected with a threshold of $p_i > P_{\text{lim}} = -10\,\text{dB}$ relative to the highest power in the aperture.

Liu *et al*. [67] propose a method for adaptive determination of the diagonal loading factor through shrinkage, that allows to reduce the intrinsic trade-off when choosing a constant loading factor for all samples. The presented simulation studies demonstrate an increase in robustness without loss of contrast or resolution compared to MV with low constant diagonal loading. Using this concept, Salari *et al*. [96] investigate several improvements to MV, namely data-dependent loading instead of fixed diagonal loading, which—combined with the shrinkage for estimation of the loading strength—improves resolution. Additionally, they propose to determine the subarray length and temporal averaging kernel length dynamically based on the CF

$$\begin{aligned} L(t) &= \left(\frac{N}{2} - 1\right) CF(t) + 1, \\ K(t) &= 2E_p(1 - CF^8(t)), \end{aligned} \tag{2.19}$$

where E_p is the length of the excitation pulse in samples. They found this approach to improve contrast. Together, these modifications lead to a MV beamformer without user parameters.

Most methods specify the solution of the MV problem as the closed-form solution (2.12), suggesting they compute the inverse of $\mathbf{R}(t)$ or employ a direct solver for $\mathbf{R}(t)\mathbf{v} = \mathbf{a}$, where $\mathbf{v}$ can then be used in (2.12). Deylami *et al*. [28] consider the practical implications of this, namely computational requirements, and suggest to instead solve the MV problem iteratively. By using the resulting weight vector of already computed samples (in range direction) and a stopping criterion based on the beamformer output energy, the iterative solution is close to the closed-form solution, while reducing the computational complexity. They report a reduction of required operations by a factor of 7.1 for a sample case.

Another approach that avoids the computational cost incurred by inversion of the covariance matrix is presented by Asl *et al*. [3]. By using a technique called dominant mode rejection, they suggest a approximation of the covariance matrix $\hat{\mathbf{R}}$ using only a small number D_m of the eigenvectors associated with the largest eigenvalues while preserving the observed power, *i.e*. the covariance matrix trace

$$\begin{aligned} \tilde{\mathbf{R}} &= \sum_{i=1}^{D_m} \lambda_i \mathbf{v}_i \mathbf{v}_i^H + \alpha \sum_{i=D_m+1}^{L} \mathbf{v}_i \mathbf{v}_i^H \\ &= \alpha \left(\mathbf{I} + \sum_{i=1}^{D_m} \frac{\lambda_i - \alpha}{\alpha} \mathbf{v}_i \mathbf{v}_i^H \right), \\ \alpha &= \frac{1}{L - D_m} \left(tr(\mathbf{R}) - \sum_{i=1}^{D_m} \lambda_i \right), \end{aligned} \tag{2.20}$$

where $\mathbf{v}_i$ and λ_i are the i-th eigenvector and eigenvalue respectively. In this model, the weight vector can be computed efficiently without numerical matrix inversion and the resulting beamformer becomes computationally very favorable. Its complexity is similar to that of the subspace methods, while the imaging performance is very close to when using the full

covariance matrix. The considered numbers of accurately retained eigenvectors D_m ranged from 1 up to 5 in their evaluation.

Subspace Minimum Variance Subspace minimum variance methods are a family of approaches that perform the beamforming in a subspace of the N-dimensional space of the aperture data. The space of the aperture is also called the element-space in some works. By approaching the beamforming problem in a subspace with as little as three dimensions, the computational complexity (especially of the matrix inversion) is drastically reduced compared to the N-dimensional problem in the element-space.

In subspace methods, the MV problem is usually first reformulated in terms of a N-dimensional unitary transform $\mathbf{B}$ with

$$\begin{aligned} \mathbf{B} &\in \mathbb{C}^{N\times N} \\ \mathbf{B}\mathbf{B}^H &= \mathbf{B}^H\mathbf{B} = \mathbf{I} \\ \mathbf{s}_B(t) &= B\mathbf{s}(t) \\ \mathbf{w}_B(t) &= B\mathbf{w}(t) \\ \mathbf{a}_B(t) &= B\mathbf{a}(t). \end{aligned} \tag{2.21}$$

Using these definitions, the optimization problem can be expressed as

$$\begin{aligned} \mathbf{w}^H\mathbf{R}\mathbf{w} &= \mathbf{w}^H E[\mathbf{s}\mathbf{s}^H]\mathbf{w} \\ &= \mathbf{w}^H E[\mathbf{B}^H\mathbf{B}\mathbf{s}\mathbf{s}^H\mathbf{B}^H\mathbf{B}]\mathbf{w} \\ &= \mathbf{w}^H\mathbf{B}^H E[\mathbf{B}\mathbf{s}\mathbf{s}^H\mathbf{B}^H]\mathbf{B}\mathbf{w} \\ &= \mathbf{w}_B^H E[\mathbf{s}_B\mathbf{s}_B^H]\mathbf{w}_B \\ &= \mathbf{w}_B^H\mathbf{R}_B\mathbf{w}_B, \end{aligned} \tag{2.22}$$

where $\mathbf{R}_B$ is the covariance matrix in the space $\mathbf{B}$. When using the full matrix, the original problem and the reformulated one are equivalent. However, depending on the construction of $\mathbf{B}$, it can be feasible to only consider a certain number of columns of it to span a subspace. Since that reduces the dimensionality of $\mathbf{R}_B$, subspace methods can substantially lower the computational complexity and potentially increase robustness. The subspace methods mainly differ in how the subspace is constructed and in the dimensionality of said subspace.

Nilsen *et al.* [81] propose to perform the MV in the so-called beam-space. That means that instead of computing how to combine the signals from the elements (conventional element-space MV), the goal is to find an optimal weighting between the results of differently steered (DAS) RX beams. This is achieved using the normalized N-point discrete Fourier transform (DFT) matrix

$$B_{l,m}^{\text{DFT}} = \frac{1}{\sqrt{N}} e^{-j2\pi lm/N}. \tag{2.23}$$

The resulting beams then have the main directions (for N even)

$$\theta_n = \begin{cases} \arcsin(2n/N) & n = 0, 1, \cdots, N/2 \\ \arcsin(2n/N - N) & n = N/2+1, \cdots, N-1. \end{cases} \tag{2.24}$$

By using an orthogonal (w.r.t. the steering vectors) beam basis, it fulfills (2.22). In pulse-echo imaging the majority of the energy is transmitted and received in the steering direction, which corresponds to the broadside direction after application of receive delays. Thus, the most influential beam directions are those around θ_0. Consequently, the concept in beam-space MV is to consider only those beams with strong response in the directions the desired signal or interference was received from. The method also employs subaperture and temporal averaging and diagonal loading, with similar parametrization to element-space MV. They demonstrate in focused imaging that it is possible to consider only a subset of as little as three of those beams with only a slight loss in imaging performance, yet drastic reductions in computational complexity. The beam-space MV was shown to be suitable for cardiac imaging using focused beams [2] and its requirement for real-time execution with the implementation presented by the authors.

Deylami *et al*. [27] suggest to use the discrete cosine transform (DCT) matrix $\mathbf{B}^{\text{DCT}}$ to construct the beam-space, instead of the DFT matrix from the original formulation [81]

$$B_{l,m}^{\text{DCT}} = \begin{cases} \frac{1}{\sqrt{N}} & l = 0,\ 0 \leq m \leq N-1 \\ \frac{2}{\sqrt{N}} \cos \frac{\pi(m+\frac{1}{2})l}{N} & 1 \leq l \leq N-1,\ 0 \leq m \leq N-1. \end{cases} \tag{2.25}$$

In their results, this shows improvements in MV estimation accuracy and robustness against focus errors compared to the DFT-based subspace when using the same number of columns. The symmetry of the beam patterns of the DCT beam-space also allow for the usage of

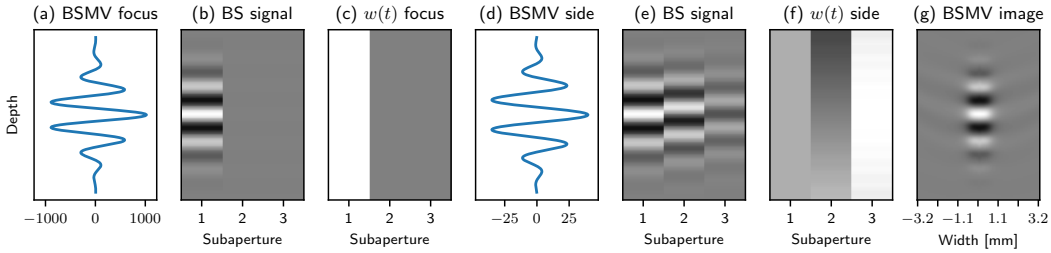


Fig. 2.9. **Visual illustration of beam-space MV** [27], $N = 40, L = 20, \Delta = 1/(100L)$ and subspace dimension 3. **(a)** Result of beam-space MV for reflector in focus. **(b)** Beam-space signal of the central subaperture for reflector in focus. **(c)** Computed weights $w(t)$ for reflector in focus. **(d)** Result of beam-space MV for out of focus reflector. **(e)** Beam-space signal of the central subaperture for out of focus reflector. **(f)** Computed weights $w(t)$ for out of focus reflector. **(g)** Resulting beam-space MV image.

even numbers of columns—*e.g*. 2—without introducing asymmetric distortions, offering more flexibility in choosing the subspace dimensionality. Figure 2.9 shows the DCT-based beam-space MV on the example data.

In an approach related to the beam-space MV, Kim *et al*. [54] propose a dimensionality reduction of the covariance matrix by projection into a signal subspace. This space is determined through principal component analysis (PCA) of exemplary MV weights, as opposed to the the DFT matrix in beam-space. The result is an approximation of the standard MV with drastically reduced computational requirements, while the resulting images are very close to the element-space MV and outperform the beam-space method in their experiments when using the same subspace dimensionality.

Bae *et al*. present an alternative projection space based on Legendre polynomials [7]. They show that in this basis, the imaging performance is comparable or even outperforms the beam-space or PCA projection methods with the same number of subspace dimensions.

Park *et al*. [83] provide a modified derivation for the MV process, that promises to reduce the complexity. While not strictly a subspace method—since no dimensionality reduction is involved—they employ a transform into a different space, namely one in which the covariance matrix is a scalar multiple of I by means of QR decomposition, removing the need to compute a matrix inverse. In the simulated and experimental data they show, the results of conventional MV and the QR based method are virtually identical.

Contrasting the subspace methods above, the eigenspace MV described by Asl *et al*. [4] computes the MV weight conventionally first and then projects the weight vector into a subspace spanned by the most influential eigenvectors of the covariance matrix before using it to compute the beam sum. To obtain a suitable covariance matrix estimate, they employ subarray and temporal averaging (similar to Synnevåg). The projection matrix is constructed from the eigenvectors associated with a certain energy, by retaining those with eigenvalue above a threshold δ relative to the largest eigenvalue. They show a significant sidelobe reduction using $\delta = 0.5$, that benefits image contrast while not compromising the resolution of the MV in their experiments. It is worth noting, that the eigenspace MV does not reduce the computational requirements compared to element-space MV. However, it seems plausible that the projection after computation of the weight vector could be combined with the eigenspace covariance matrix approximation presented in [3].

Multistage Beamforming Sakhaei *et al*. [95] utilize the robustness increase of the eigenspace MV for an approach that promises to alleviate the loss of beamforming performance that occurs when MV needs to be applied with spatial smoothing. This is achieved by applying a second stage MV to combine the beamformed subapertures, replacing the otherwise employed sum over them. In their evaluation on simulations using the eigenspace MV for both stages resulted in the best performance for their method, outperforming the subaperture averaged MV.

A different combination of two beamformers is suggested by Shamsian *et al*. [100]. It uses DAS as the first stage on small (*e.g*. $J = 8$) subapertures followed by a decimation (*e.g*. $D = 4$), effectively prefiltering the aperture data with a stride > 1 leading to a considerable reduction in the aperture dimension. Using MV or beam-space MV as the second stage is then considerably less computationally expensive. Their results suggest that this method is a good approximation of the MV and beam-space MV results.

2.2.5 Hybrid Methods

Asl *et al*. [5] combine the MV with CF weighting (both computed on $s_i(t)$)

$$y_{\text{Asl}}(t) = \text{CF}(t) \cdot y_{\text{Synnevåg MV}}(t). \tag{2.26}$$

Doing this, they show an increase in contrast compared to DAS and MV in simulations.

The opposite has been investigated by Wang *et al.* [120]. They propose a method to estimate the coherence factor based on the principles of the MV and use it as weighting of DAS results

$$\begin{aligned} y_{\text{Wang}}(t) &= \text{CF}_{\text{MVDR}}(t) \cdot y_{\text{DAS}}(t) \\ &= \frac{1}{\mathbf{a}^H \mathbf{R}^-1(t)\mathbf{a}} \frac{1}{tr(\mathbf{R}(t))} y_{\text{DAS}}(t). \end{aligned} \tag{2.27}$$

They also extend this similar to the GCF by taking the energy from multiple closely spaced steering vectors $\mathbf{a}(\theta)$ into account.

Su *et al.* [104] combine the DMAS beamformer with SCF weighting for each individual low resolution image in a PW sequence and their results suggest this as an interesting avenue to increase contrast and resolution compared to DMAS alone. The images they present however exhibit depth-dependent brightness changes reminiscent of those observed when applying SLSC to focused imaging.

2.2.6 Others

A low complexity adaptive beamformer was suggested by Synnevåg *et al.* [105] that is inspired by the working principle of the MV. By limiting the solution space of the MV to a set of preselected weight vectors, this approach presents a very compelling trade-off between the imaging performance of the MV, computational complexity and ease of implementation. Instead of minimizing (2.11), the low complexity variant selects the weight from P preselected windows, that minimizes the output variance

$$\arg\min_p E[\left|\mathbf{w}_p^H \mathbf{s}(t)\right|^2], p = 1, \cdots, P, \tag{2.28}$$

where the variance is estimated over a temporal window for each weight vector

$$\hat{\sigma}_p^2 = \frac{1}{2K+1} \sum_{k=-K}^{K} \left|\mathbf{w}_p^H \mathbf{s}(t-k)\right|^2, \tag{2.29}$$

followed by selecting the p with lowest variance $\hat{\sigma}_p^2$ and using the weight $\mathbf{w}_p$ for beamforming.

Zahnd *et al.* [129] introduce a DAS weighting scheme, that uses a bimodal Gaussian function symmetric around the aperture center. This induces transversal oscillations in the resulting beamformed data—in addition to the implicitly present longitudinal oscillations—that enable phase based motion tracking in two dimensions. They evaluate this approach in-vivo with intima-media tracking and observed an increase in tracking accuracy compared to intensity based tracking on conventional images.

Zhuang *et al.* [132] propose a beamforming method that uses the accumulated phase change across the aperture to highlight specular or bone interfaces while strongly suppressing soft-tissue responses.

In some way similar to CF and GCF, Wang *et al.* [121] propose an image weighting scheme targeted at sidelobe suppression that is based on the consistency of the aperture data. This is achieved by considering the ratio of mean to standard deviation either on the aperture data itself or on the spectrum of the aperture data, resulting in a weighting whose effects are similar to CF and GCF respectively.

Chernyakova *et al.* [16] present a statistical interpretation of the beamforming process. By modeling the aperture data as $\mathbf{s} = \mathbf{1}y + \mathbf{n}$, with the signal of interest y and a noise vector $\mathbf{n}$ both assumed to be Gaussian random variables ($y \sim \mathcal{N}(0, \sigma_y^2)$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$), they propose an iterative beamformer based on alternating maximum-likelihood estimation of the distribution parameters σ_y^2 and σ_n^2 and maximum-a-posteriori estimation of the signal of interest, as shown in Algorithm 1. The resulting iMAP beamformer shows, in their experiments, resolution similar

Algorithm 1 iMAP beamforming [16]

1: Initialize $y_{\text{iMAP},(0)} = \frac{1}{N}\mathbf{1}^H\mathbf{s}$
2: Update distribution parameters
$\left\{\hat{\sigma}_y^2, \hat{\sigma}_n^2\right\}_{(i)} = \left\{y_{\text{iMAP},(i)}^2, \frac{1}{N}\left|\mathbf{s} - \mathbf{1}y_{\text{iMAP},(i)}\right|^2\right\}$
3: Update signal estimate
$y_{\text{iMAP},(i+1)} = \frac{\hat{\sigma}_{y,(i)}^2}{\hat{\sigma}_{n,(i)}^2 + N\hat{\sigma}_{y,(i)}^2}\mathbf{1}^H\mathbf{s}$

to DAS, contrast improvements better than CF, while suppressing the speckle less than CF. This appears to be advantageous in acquisition schemes with low transmit focus, *i.e.* ultrafast

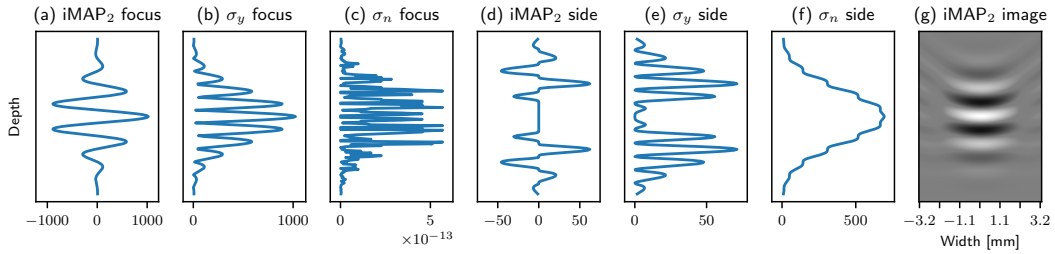


Fig. 2.10. **Visual illustration of iMAP** with two iterations. **(a)** Result of iMAP for reflector in focus. **(b)** Determined signal standard deviation for reflector in focus. **(c)** Determined noise standard deviation for reflector in focus. **(d)** Result of iMAP for out of focus reflector. **(e)** Determined signal standard deviation for out of focus reflector. **(f)** Determined noise standard deviation for out of focus reflector. **(g)** Resulting iMAP image.

imaging, especially when a limited number of transmissions is employed. Figure 2.10 shows the result of iMAP when executing two iterations, including the determined signal and noise parameters.

Polichetti *et al.* [85] propose a generalization of DAS that is also related to DMAS. They present p-DAS which uses the (signed) p$^{\text{th}}$-root before summation, and p-power scaling afterwards to introduce nonlinearity

$$\begin{aligned} \tilde{y}_{\text{p-DAS}}(t) &= \sum_{i=1}^{N} sgn(s_i(t))\ |s_i(t)|^{\frac{1}{p}} \\ y_{\text{p-DAS}}(t) &= sgn(\tilde{y}_{\text{p-DAS}}(t))\ |\tilde{y}_{\text{p-DAS}}(t)|^p. \end{aligned} \tag{2.30}$$

This is followed by a bandpass centered at the probe center frequency f_0. The strength of the effect is tunable through the choice of p, the authors evaluate $p = 2$ and $p = 3$, for $p = 1$ it is equivalent to DAS. The result is a promising increase in resolution and contrast in

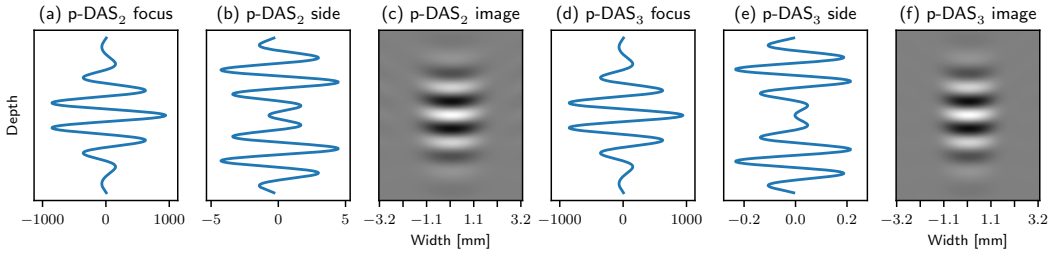


Fig. 2.11. **Visual illustration of p-DAS.** **(a)** Result of p-DAS ($p = 2$) for reflector in focus. **(b)** Result of p-DAS ($p = 2$) for out of focus reflector. **(c)** Resulting p-DAS ($p = 2$) image. **(d)** Result of p-DAS ($p = 3$) for reflector in focus. **(e)** Result of p-DAS ($p = 3$) for out of focus reflector. **(f)** Resulting p-DAS ($p = 3$) image.

their simulated and experimental phantom data, albeit with a change in speckle appearance compared to DAS. The effects of p-DAS are illustrated in Figure 2.11.

Deep Learning based Methods A method for suppressing off-axis scattering in aperture data is presented by Luchies *et al.* [70]. It uses a set of neural networks that are tasked to remove echos from scatterers outside of the mainlobe. The input to the networks are the short time Fourier transform (STFT) coefficients of each input channel (STFT applied to the time-domain delayed data) and the output are the coefficients after removal of the undesired signals. When followed by inverse STFT, the resulting aperture data is summed to create the image. They trained the networks on single scatterer simulations where the scatterer was either placed in the mainlobe and its echo should consequently be retained, or outside of the mainlobe, where the desired output was zero. In phantom and in-vivo data, they show contrast improvements similar to or outperforming CF, while generally retaining the speckle appearance of DAS. They extend this work with a detailed investigation of the learning process [69]. The effects of training set size and the other hyperparameters are evaluated, while also probing the method for robustness against channel noise. Additionally, they extend the training data to contain multiple scatterers.

Hyun *et al.* [43] utilize a deep neural-network (DNN) for speckle reduction from aperture data, where the network input are prebeamformed RF signals from several subapertures as well as the DAS image. The network is then trained on simulations and the echogenicity maps used to create the simulations serve as the training targets. They show promising performance and outperform spatial compounding and other speckle reducing postprocessing filters in a number of cases. In their investigation of the training loss, they found a combination of the L_1 norm and the multiscale structural similarity index to perform best on log-compressed images.

Complimentary to these deep learning based methods that aim at novel image characteristics, there is a movement that employs machine learning with the goal of runtime performance improvements of non-learned methods.

Luijten *et al.* [71] propose the use of fully-connected neural networks to learn a fast approximation of a computationally complex beamformer. The method called ABLE (adaptive beamforming by deep learning) operates on each output sample independently and its output is a weight vector to be used analog to (2.2). It is a lightweight network trained with an intriguing combination of losses: First, an image loss of the beamformed result against the MV output after application of the logarithm to better capture the deviations that would be perceptible after log-compression. The second term $\mathcal{L}_{unity}(\mathbf{w}_{\text{ABLE}}) = \left|\mathbf{1}^T\mathbf{w}_{\text{ABLE}} - 1\right|^2$ penalizes deviations of the output weight vector $\mathbf{w}_{\text{ABLE}}$ from unity gain. This aims at promoting a distortionless response, *i.e.* the signal amplitude in the direction of interest is not reduced. In their evaluation, it performed on par with, or better than the eigenspace MV beamformer similar to [4] used as training target. In terms of computational complexity, ABLE compares favorably against the MV, the first being in $\mathcal{O}(N^2)$, whereas MV generally falls into $\mathcal{O}(N^3)$, potentially alleviating performance constraints for the application of MV-methods.

In a similar spirit, Wiacek *et al.* [123] present with CohereNet an architecture for the speedup of correlation estimations. It is based on a fully connected neural network that estimates all spatial correlations $\tilde{R}(m,t), m \in \{1, \cdots, N-1\}$ in the aperture data, with the numeric version (2.7) as training target. Using the spatial correlation estimates to compute the SLSC, they show a significant speedup of their method compared to the conventional CPU implementation and better accuracy compared with a previously proposed GPU implementation of the SLSC, which makes use of simplifying assumptions.

2.3 Comparison and Discussion

We compare a number of the highlighted beamforming algorithms on two datasets qualitatively and through contrast and resolution measurements. For this, we chose with DAS, CF [42], GCF [65], PCF, SCF (both [13]), SLSC [62], F-DMAS [75], element-space MV [106], beam-space MV [27], iMAP [16] and p-DAS (for $p \in \{2, 3\}$) [85] the foundational methods from each group. We execute them on two focused imaging datasets from the UltraSound ToolBox [92, 134], that were acquired with an Alpinion ultrasound machine using a L3-8 probe. The datasets depict different regions of a CIRS General Purpose Ultrasound Phantom Model 054GS (CIRS Inc., Norfolk, VA, USA). One consists of a homogeneous background with an anechoic and a hypoechoic inclusion relative to the background. The second contains a hyperechoic inclusion (approx. $12\,\text{dB}$ relative to the background), two isolated wire targets, and a resolution group. The phantom was imaged with 256 focused emissions and a transmit focus at $30\,\text{mm}$ depth.

The scan lines were delayed appropriately, followed by the application of the different beamformers on the resulting aperture data. For display, the beamformed images were envelope detected using a Hilbert-FIR filter followed by logarithmic compression except for SLSC, where only logarithmic compression is applied. The code to apply the beamformers, create the illustrations and perform the comparisons is available at https://github.com/goeblr/rxbf_review. Figure 2.12 shows the result of the beamformers on the region with the anechoic inclusion. Compared to DAS, using CF, SCF and p-DAS$_3$ drastically changes the speckle appearance, making the background appear as less homogeneous—with SCF even to the point where subsequent speckle filtering could become challenging. For GCF, iMAP, F-DMAS, and p-DAS$_2$

this effect is less pronounced and with PCF, or the MV variants the texture of the background region is very similar to the DAS result. In the SLSC image, the background appears even smoother than all other images. The anechoic inclusion should—ideally—be depicted as black, however there is noise visible in the DAS, PCF and MV results. While the others show no noise (within the shown range of $60\,\mathrm{dB}$), the CF, SCF, and GCF images stand out in their sharp delineation of the inclusion boundary. It appears as if there is a trade-off between speckle appearance and noise in the anechoic region for the shown beamformers.

In Figure 2.13 we focus on the appearance of the wire targets that are incorporated into the phantom and have a diameter of $100\,\mu\mathrm{m}$. In it, the SCF and the MV variants show the most narrow response. CF, PCF, F-DMAS and the p-DAS images also exhibit an improvement in the width of the targets compared to DAS, while the iMAP image response appears comparable to it. In the GCF and SLSC the point response is wider than in DAS. It is noteworthy, that most of the coherence based methods (CF, GCF, SCF, and SLSC) suppress the background in the lateral vicinity of the wires, and iMAP and p-DAS$_3$ show this to some degree as well.

We also quantitatively compare the selected methods on the datasets shown in Figures 2.12 and 2.13. They are investigated w.r.t. the contrast ratio (CR)

$$\mathrm{CR} = \begin{cases} \frac{\mu_B}{\mu_T}, \text{for known dark targets} \\ \frac{\mu_T}{\mu_B}, \text{for known bright targets,} \end{cases} \tag{2.31}$$

relative to the background, and the contrast to noise ratio (CNR)

$$\mathrm{CNR} = \frac{|\mu_B - \mu_T|}{\sqrt{\sigma_B^2 + \sigma_T^2}}, \tag{2.32}$$

where μ_T and σ_T are the mean an standard deviation of the uncompressed envelope (except for SLSC, where no envelope detection is required) in a target region, while μ_B and σ_B are computed on an homogeneous background region. The speckle signal to noise ratio (SNRs)

$$\mathrm{SNRs} = \frac{\mu_B}{\sigma_B} \tag{2.33}$$

is used as a measure of homogeneity in the background region. These metrics are determined for the anechoic and the hyperechoic inclusion. Finally, the full width half maximum (FWHM) of the wire target responses is used as a measure of the beamformer resolution capabilities.

Table 2.2 shows the results of the evaluation. The CR measurements of the anechoic inclusion (ideally infinity, for an area completely void of echos) mostly agree with our observations regarding the noise within the inclusion. With CR around $30\,\mathrm{dB}$, DAS, PCF and the MV variants show noise well within the shown dynamic range. The highest CR values result from the application of SCF, CF, and GCF, which had the clearest boundary definition visually.

For the hyperechoic inclusion with echogenicity (as given by the manufacturer) of "about $12\,\mathrm{dB}$" (without a specified tolerance), there are two methods for which the CR is the furthest from $12\,\mathrm{dB}$, SCF and F-DMAS, both of which seem to suppress the background speckle disproportionately.

	Anechoic inclusion			**approx. +12 dB inclusion**			**Wires**	
	Circles in Figure 2.12			Circles in Figure 2.13			In Figure 2.13	
	CR [dB]	CNR [dB]	SNRs	CR [dB]	CNR [dB]	SNRs	FWHM 1 [mm]	FWHM 2 [mm]
DAS	28.51	**4.33**	**1.71**	13.06	**2.84**	**1.81**	1.28	2.16
CF+DAS	**58.43**	-0.44	0.95	14.95	-1.32	0.99	0.99	1.65
GCF+DAS	**50.49**	2.75	1.38	14.61	1.80	1.42	1.35	2.16
PCF+DAS	29.95	3.39	1.53	13.42	1.87	1.63	1.21	2.01
SCF+DAS	**88.44**	-6.99	0.45	20.91	-5.69	0.43	**0.77**	**1.13**
$iMAP_2$	43.21	3.84	1.57	13.12	2.07	1.63	1.28	2.16
SLSC	44.70	**7.53**	**2.40**	**12.26**	**5.03**	**2.69**	2.38	4.35
F-DMAS	38.55	1.40	1.19	8.17	-2.21	1.24	0.99	1.57
$p\text{-}DAS_2$	40.18	2.15	1.29	14.26	0.83	1.33	**0.91**	1.65
$p\text{-}DAS_3$	44.47	0.48	1.06	15.36	-0.57	1.08	**0.91**	1.50
MV	28.81	3.89	1.63	**11.72**	1.94	1.78	1.72	**1.13**
BS-MV	29.30	**4.07**	**1.66**	**11.96**	**2.20**	**1.82**	1.65	**1.13**

Tab. 2.2. **Contrast and resolution metrics** computed on the highlighted areas in Figure 2.12 and Figure 2.13. FWHM 1 determined on the wire at 20 mm, FWHM 2 at 40 mm. In bold: For the contrast and noise ratios the three highest values - except for the CR of the +12 dB inclusion, where smallest differences to 12 dB are highlighted. For wire FWHM the three lowest values are emphasized.

The SNRs measurements of the background regions describe the appearance of the speckle within them. High values indicate smooth appearance, while low values are the result of high intensity variations. They generally match between the two acquisitions, especially on the relative ordering of the methods and agree with the visual impressions of speckle suppression or the lack thereof.

The choice of a beamforming method is influenced by a number of factors. The first constraint to consider is the amount of required computations. Given real-time nature of US and its interactive use, a beamformer that cannot be applied fast enough simply cannot be used.

In situations where computational resources are limited, *e.g.* to fit a given power envelope, or when ultrafast imaging should to be performed at the rate limit given by the physical wave propagation time, one might consider the less computationally demanding approaches such as CF, PCF, SCF, and p-DAS. These are followed by the more complex methods GCF and iMAP. Should the available compute not be a limiting factor, one might consider the application of MV (likely in a subspace variant), DMAS, or SLSC. In terms or computational complexity, learned methods can span a wide range. The approaches aimed at replicating the results of a different method tend to be designed to be faster than the originals, while the techniques that introduce novel imaging methods can be more computationally demanding.

Given this initial selection, the clinical application dictates which approach is to be selected. If the task requires accurate representation of image contrast, *e.g.* to show and grade inclusions accurately, SLSC should be a prime candidate. Should it be necessary to achieve maximum resolution, then a MV method or their learned approximations can be an excellent fit. If, however, the images are mainly degraded by clutter, the coherence based weighting methods or F-DMAS should be considered.

Generally, there is no single best beamformer, the suitability of a method depends on the requirements of the clinical application, the anatomy to be visualized and available perfor-

mance. Given these constraints, one has to select the optimal beamformer for the specific circumstance.

2.4 Conclusion

In this chapter, we gave an overview of the developments in US beamforming methods that perform aperture data processing. We grouped the works by their main working principle and compared the results of select methods on phantom data. Although there are a few clusters of works w.r.t. their underlying principles, such as aperture data coherence, all methods have their own strengths and weaknesses. Among the most important aspects to consider are the beamformers resolution capabilities, the accuracy in reproducing tissue contrast, the appearance of the resulting images, and of course the computational requirements. This makes the selection of beamforming method ultimately dependent on the clinical application of the system and the trade-offs acceptable in each specific setting.

With our focus on processing methods that operate on the data after delay application, effects from speed-of-sound errors are not considered in our evaluation. A detailed investigation of the influence of speed-of-sound mismatch and phase-abberations (as caused by speed of sound inhomogeneities in tissue) on the methods would be of interest.

Within the body of work on ultrasound beamforming, there has been a shift in presented algorithms towards data-driven approaches, specifically utilizing deep learning. We believe this to be a very promising avenue for further imaging performance improvements, the reduction of computational requirements of existing beamformers, and most importantly the development of entirely new methods.

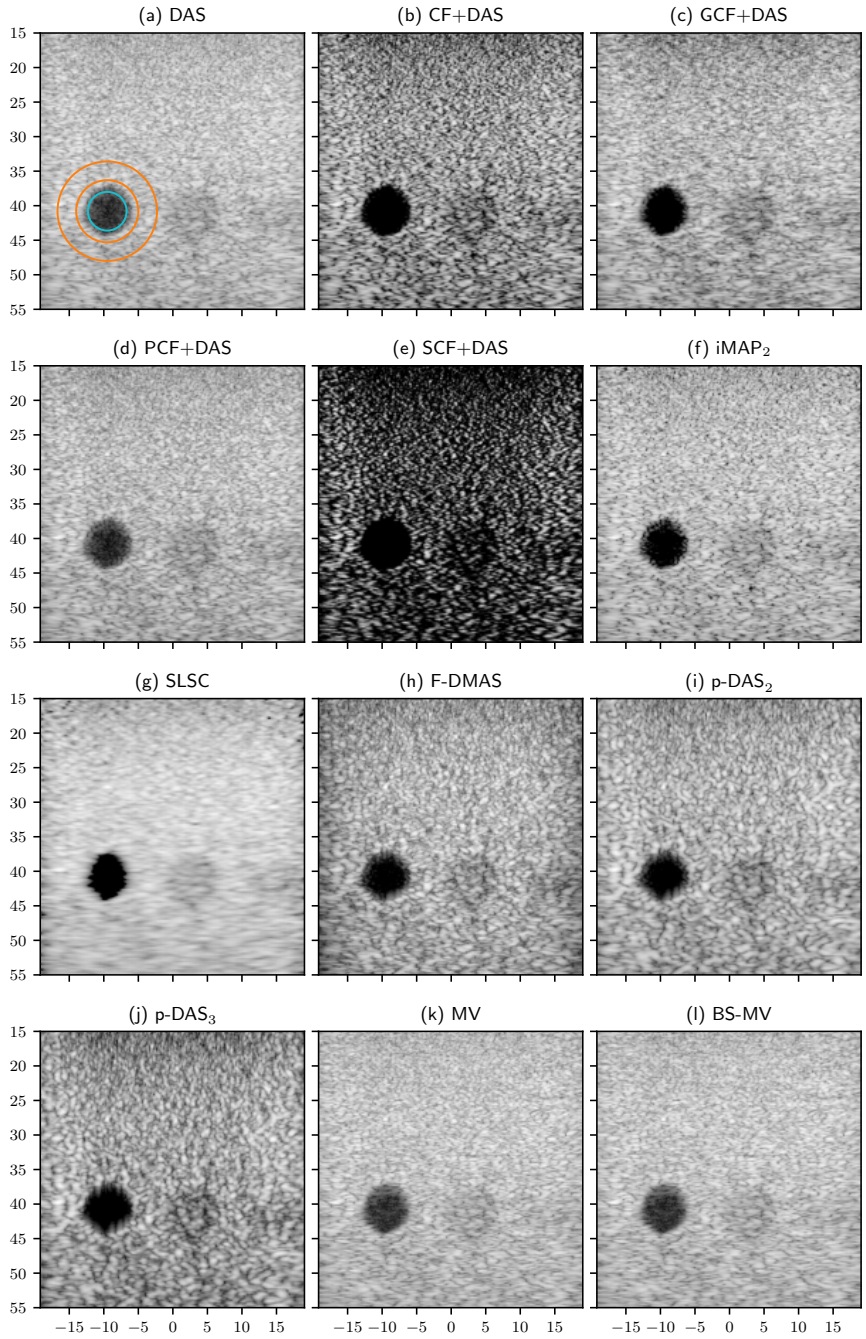


Fig. 2.12. **Qualitative comparison between beamformers** on phantom data. The images show 60 dB of dynamic range. Highlighted areas are used for contrast measurements, with the blue circle defining the target, the ring between the orange circles the background region. Raw ultrasound dataset: "Alpinion L3-8 FI hypoechoic", v1.0.2 from [134]

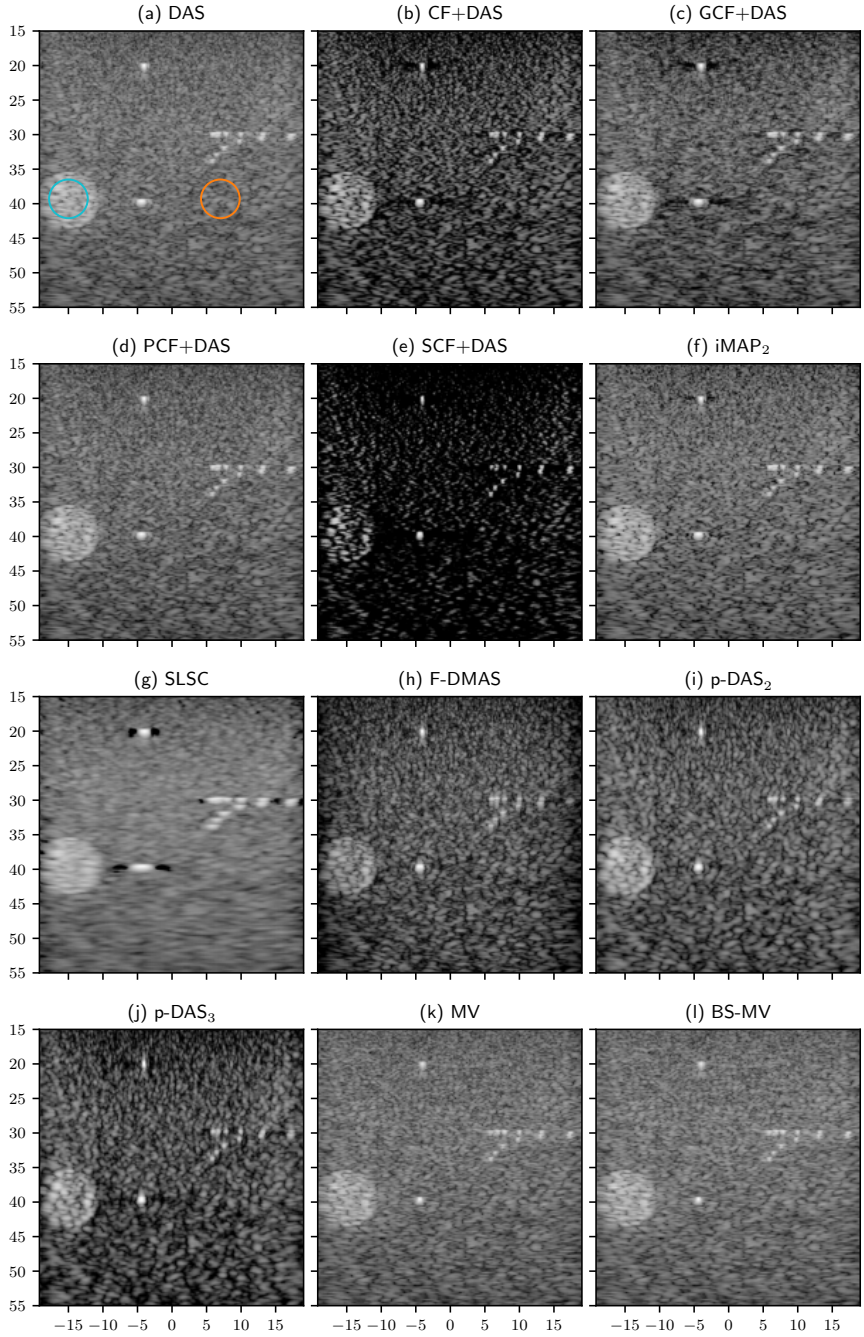


Fig. 2.13. **Qualitative comparison between beamformers** on phantom data. The images show 60 dB of dynamic range. Highlighted areas are used for contrast measurements, with the blue circle defining the target, the orange circle the background region. Raw ultrasound dataset: "Alpinion L3-8 FI hyperechoic scatterers", v1.0.2 from [134]

Part II

Methodology and Contributions

3 Software based Ultrasound Imaging

In this chapter we present an open framework for software-based ultrasound imaging, aimed at being flexible to extend and real-time use. The goal is to close the gap between research around US image formation that often operates offline on pre-recorded data and closed commercial imaging systems that allow interactive, real-time application.

Section 3.1 provides the motivation and reviews the related work. In section 3.2 we present the design considerations behind and properties of the proposed framework and in section 3.3 we investigate the imaging performance and run-time behavior of it.

Substantial parts of this chapter have been published and are quoted verbatim:

[32] **Rüdiger Göbl**, Nassir Navab, and Christoph Hennersperger. “SUPRA: open-source software-defined ultrasound processing for real-time applications”. *International Journal of Computer Assisted Radiology and Surgery (Mar. 2018)*.

3.1 Motivation and Related Work

Ultrasound (US) imaging is used in a wide variety of applications and complements modalities such as computed tomography (CT) and magnetic resonance imaging (MRI). It enables non-invasive, low cost acquisition of anatomical, dynamical, as well as, functional information while being highly portable. Because of this, there is a trend to replace MRI and especially CT with US imaging when applicable.

The image acquisition process of US can be broken down into several steps forming a pipeline.

- First, an ultrasonic pulse modulated with a chosen frequency is generated in the tissue through precisely orchestrated electrical excitation of piezo transducer elements (transmit beamforming).

- The echos induced by this pulse in the tissue are converted to electrical signals by the transducer elements and commonly stored digitally (receive).

- After that, the signals are used to compute how they would have been received from one single line, where the data from different channels is delayed such that scattered signals from that line are intensified by constructive interference and echoes from other positions are reduced through destructive interference (receive beamforming) [101].

- The result of this operation is called the radio frequency (RF) data, because the echos are still modulated with the transmit frequency.

- Since only the scattered intensity is of interest to the user, the RF-data is demodulated (envelope detection), leaving the signal amplitude. As this reduces the maximum signal frequency, it is usually decimated during this step.

- Before these amplitudes are shown on a screen they undergo a non-linear compression stage to match their dynamic range to the perceived dynamic range (log-compression).

- Finally, the single lines that have been reconstructed and which are not necessarily parallel to each other are interpolated to create an image representing physical dimensions (scan conversion).

Those steps typically have been performed in specialized hardware, field programmable gate arrays (FPGAs), or digital signal processors (DSPs). Implementations like this have been necessary in the early years of US imaging. With the increase of computing power available in modern workstations, this need has been relaxed, especially when using GPUs to perform the numerical calculations. When implemented in separate hardware, modifications to the algorithms are hardly possible, as programming interfaces might not be available and the development for FPGAs and DSPs is highly complex and time-consuming. Especially for research applications, such limitations often lead to the utilization of frame-grabbers to allow for a retrieval of US images from clinical scanners [130] even in recent publications [9, 91], which is not only limiting reproducibility but also potentially hampers image quality.

There already exist some US systems, in which all of the processing happens on GPUs, thus allowing the manufacturer of the device to implement changes to their pipeline more rapidly. Yet, as vendors protect their intellectual property, it is difficult for independent research groups to use those machines for research purposes. On the other hand, there have been efforts to give researchers access to parts of the US pipelines they employ, as for example by the PLUS toolkit [61]. While this project is tailored to tracked and freehand 3D-US applications, it does not provide US beamforming or low-level processing capabilities. An open source US processing implementation based on ITK [77] has been proposed. While it is real-time capable, it only covers the later processing steps (envelope detection to scan conversion), requiring the use of other (potential proprietary) implementations for the early steps. Besides those efforts, a group of researchers recently announced the UltraSound ToolBox for MATLAB [92]. While this platform provides basic capabilities for beamforming and development of advanced low-level processing methodologies, it is conceptionally not directed towards the use in real-time applications, thus not being suited for many applications in computer assisted interventions.

In [124] Wilson proposes with the Ultrasonix 500 RP a research interface to intermediate data at different stages up to RF as well as control over many system parameters. This allows for the execution of methods other than the proprietary built-in functionality. However, it does not yet enable access to channel data. By using an additional data aquisition module (either SonixDAQ [17] or the real-time capable RX-DAQ [116]) channel data can be captured. Yet, the base implementations are proprietary–and are as such bound the Ultrasonix hardware–and can only be replaced to some extent. In contrast to this, the ULA-OP US system [10] is open and it is possible to implement new methods for it. Yet, as the processing happens in FPGAs and DSPs, this task is complex.

On this basis, we try to close the gap between low-level *offline* US research on the one side, and the processing of *already processed* US images on the other. We propose an open source pipeline for 2D and 3D US imaging called SUPRA (software defined US processing for real-time applications) that can be used to perform all computation-based steps in US and enable researchers and developers to work on all parts of the imaging process. This ranges from beamforming the raw channel data recorded by an US-system to the output of B-mode images. Thus, we hope to integrate efforts of other platforms, and also specifically provide a way towards a stronger integration of high-level processing (targeting *e.g.* at a specific medical image computing or computer aided intervention application) with low-level, US-specific information (*e.g.* raw channel or RF data). This way, SUPRA could for example be used to integrate efforts from image segmentation throughout all levels of the US processing pipeline, potentially leveraging specific first order data not considered so far.

3.2 SUPRA

In the following, we describe our approach towards software defined US processing for real-time applications (SUPRA). The framework is licensed under LGPLv3 and designed as cross-platform solution tested with both Windows and Linux/Ubuntu. It is publicly available on GitHub[1].

The term software defined US is derived from a concept called “software defined radio”. In this field of radio communication, hardware implementations of signal processing components such as filters, amplifiers and modulators are replaced by software implementations. This can help to reduce costs and simplify their development.

Our framework SUPRA follows the same concept. In addition to being fully implemented in software, all the real-time critical processing steps have been implemented in NVIDIA CUDA to achieve high throughput. Figure 3.1 shows a pipeline with the steps as outlined in Section 3.1 and highlights where the respective processing takes place. The transmit beamforming is performed on the CPU, as it is only necessary to compute the transmit parameters once for a fixed acquisition. It is important to note here that *only* the actual transmit and recieve steps require specific hardware with an analog frontend, allowing for the excitation of piezo-elements in the US probe to create acoustic waves. Thus, all other steps in the pipeline (receive beamforming, envelope detection, log-compression, and scan conversion) can be

[1] https://github.com/IFL-CAMP/supra

fully customized in software. Since these are performed repeatedly, we implemented them in NVIDIA CUDA. As a consequence, it is possible to execute the pipeline for 3D US on consumer

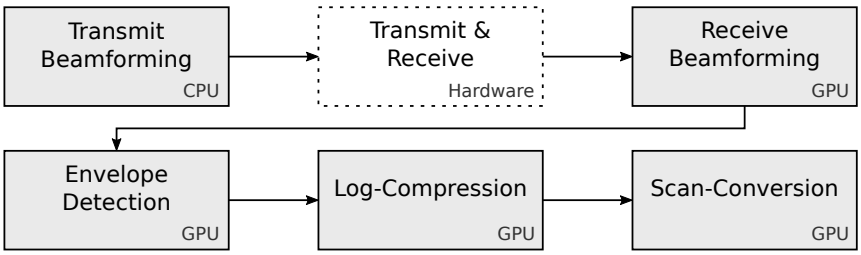


Fig. 3.1. **Implemented pipeline** with the modules realized in SUPRA, hence in software, marked with solid lines and the module that has to be performed in hardware with dashed lines.

graphics cards, as we show in Subsection 3.3.3.

Besides the aim to maximize real-time capability by parallelizing relevant parts of the US pipeline, we also considered the modularity of the pipeline as a major design goal. We achieved this by encapsulating the processing components into nodes on a data-flow graph, realized with the Intel Thread Building Blocks open source library[2]. Based on this architecture, the nodes only exchange shared pointers to data containers (which may reside either on the CPU or the GPU) among each other, eliminating unnecessary memory operations.

In detail, each encapsulated node can utilize an input, provide an output, or both. On this foundation, by placing the processing steps of the US pipeline within the data-flow graph, nodes can be added, removed, or exchanged without recompiling the actual code. Besides this possibility to exchange individual parts of the pipeline, the overall pipeline can also be altered completely using the input-output mechanisms. In this way, it is for example possible to perform two differently parametrized beamforming runs in parallel on the same input data to extract different information. In a hardware-based pipeline this is not possible, while with SUPRA such considerations are limited only by the computational power.

In view of our efforts to provide a fully functional basic US beamforming and processing pipeline, the following methods are available in SUPRA:

1. Dynamic transmit and receive beamforming for a fully flexible scanline layout and resolution, as well as full control of acoustic wave excitations and multi-line acquisitions

2. Delay and sum beamforming for received raw channel data directly after analogue to digital conversion [101]

3. Envelope detection by IQ-demodulation, including frequency compounding through a bank of configurable bandpasses

4. Configurable log compression for target imaging dynamic range

5. Scan conversion in 2D and 3D, for a wide range of scan-line configurations

[2] `https://www.threadingbuildingblocks.org/`

6. Graphical user interface for online-configuration and real-time visualization of received data

7. Configurable XML-interface for the generation of system parameters and specific imaging pipelines

In addition to the nodes that make up the core pipeline, several input and output nodes are present. Input nodes provide the interface to the actual US system hardware (*i.e.* hardware transmit and receive as indicated above) and are thus vendor or system-specific interface implementations. At this stage, Cephasonics US hardware (Cephasonics, Santa Clara, CA, USA) is integrated with the full pipeline for beamforming; and Ultrasonix systems (BK Ultrasound, Peabody, MA, USA) can be interfaced using the proprietary ulterius software interface. It should be noted, however, that the integration of other hardware-platforms would only require the implementation of a new input node within the data graph, providing a respective interface to the system-specific transmit and receive hardware.

Output nodes provide a way to either forward information at any stage of the overall pipeline to a dedicated interface, or provide a means to save data to a file on the hard-disk respectively. Implemented output nodes represent at this stage:

- ROS bridge for interaction with robotic environments,
- OpenIGTLink bridge for exchange of image and tracking data,
- Storage of information as meta images (mhd) for offline use.

It is worth noting that output nodes are not limited to the last step of the pipeline, but can also be used with any intermediate data stream present in the system. Following the generic and modular software architecure, nodes for input and output can also contain other information. For the use in interventional settings, tracking information can be attached to the images. In this view, SUPRA provides generic interfaces for tracking in- and output via OpenIGTLink and as ROS messages.

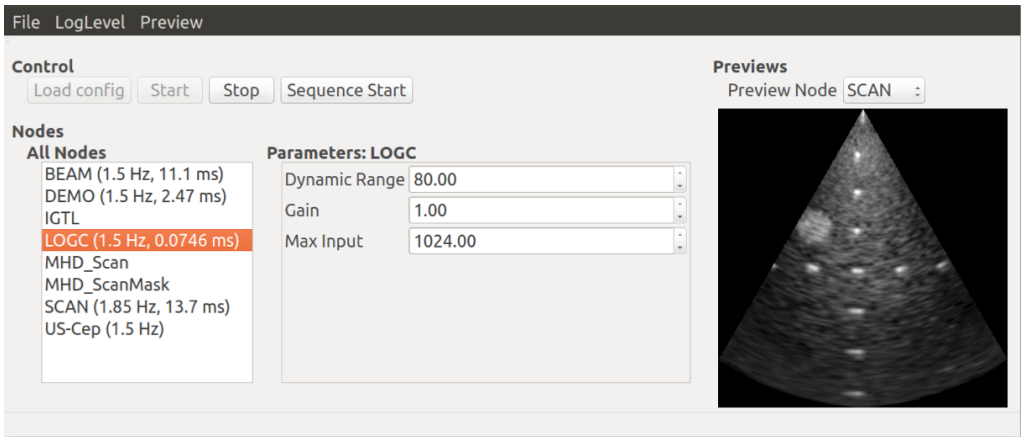


Fig. 3.2. **Graphical user interface** with the processing nodes in the left column, the parameters of the selected node in the central column and a live preview of one data stream on the right.

The online-configuration mentioned before is realized via a generic parameter system, that enables all processing nodes to define parameters with default values and valid value ranges (continuous, discrete), where applicable. Through this parameter-system the respective nodes are notified of parameter changes and can react accordingly. The parameters can be inspected and modified during run-time by the user in the GUI, which is shown in Figure 3.2. Additionally, for the use in automated systems, the parameters can be accessed through a ROS service, allowing fully dynamic imaging.

To this end, in order to use the full pipeline as introduced above, a respective US system is required, providing access to the beamforming parameters and the raw data collected at channel level has to be available. However, even without this access, a user can still apply the other parts of the pipeline to harness the full control over those steps, *e.g.* by injecting previously acquired data into the pipeline.

Additionally, with the planned inclusion of image post-processing to SUPRA, such as speckle reduction techniques [57], or advanced imaging protocols like harmonic imaging [6] and planewave imaging [108], the current baseline provided by the framework will allow researchers to evaluate their methods in a more meaningful manner.

3.3 Comparison

For evaluation we employ the proposed SUPRA pipeline with a 384 channel cQuest Cicada™ from Cephasonics, CA, USA and a Cephasonics CPLA12875 transducer with $7\,\mathrm{MHz}$ center frequency, 128 elements, and $0.3\,\mathrm{mm}$ pitch and work with the raw channel data collected by this system. For this transducer only 64 channels can be used. The resulting images are compared qualitatively as well as quantitatively to the clinical pipeline native to the Cephasonics Cicada with the same transducer. Both systems use the following parameters: frequency: $7.6\,\mathrm{MHz}$, depth: $4.5\,\mathrm{mm}$, focal depth: $30\,\mathrm{mm}$. Additionally, we show acquisitions performed with a 2D matrix probe with all 384 channels of the Cephasonics interface.

3.3.1 Qualitative Comparison

For a qualitative comparison of the image quality between the two systems, we show in-vivo images of the carotid artery and the biceps tendon in combination with the brachialis of a healthy volunteer, as well as phantom data acquired with a CIRS multi purpose phantom (Model 040GSE). Due to tissue deformations and limited probe placement reproducibility, the anatomy shown is not exactly the same. Nevertheless, in the in-vivo images in Figure 3.3 it can be observed, that the tissue texture is comparable in both systems, although it appears less blurred with the Cephasonics pipeline.

This blur in the SUPRA images is likely caused by settings of the frequency compounding during envelope detection, as the filters used to separate different frequencies of the RF lines can cause blurring.

In the phantom images the tissue texture in the image acquired with SUPRA appears more consistent than with Cephasonics. Additionally, the wires in the lower part of the images exhibit a higher contrast in SUPRA and seem less blurred as well, while being more spread in the upper part. The qualitative comparison thus shows, that the image quality of SUPRA is comparable to what can be achieved with the built-in Cephasonics pipeline.

As pointed out earlier, SUPRA is not limited to classical 2D US, but is also capable of 3D imaging. Slices of a volume acquired with a 32×32 element Vermon matrix probe are shown in Figure 3.4. It was connected to a Cephasonics cQuest Cicada with 384 channels. The volume shown was acquired with $70\,\mathrm{mm}$ depth, a frequency of $7\,\mathrm{MHz}$ and with 512 (32×16) scanlines over a field of view of $60°$. The top-left image shows a volume slice perpendicular to all internal structures and it is clearly visible that the resolution decreases with increasing depth, as is expected for a phased array. The image in the top-right shows a 3D rendering of the volume, where the different lengths of the the hyperechoic inclusion and the wires are apparent. Bottom left shows a slice perpendicular to that, including a longitudinal view of the hyperechoic inclusion and the image in the bottom-right shows a longitudinal view of the horizontal wires. It is worth noting that the wires are longer than is visible from the slice, but due to the increasing scanline angles to the image boundaries natural to phased arrays in combination with the highly specular reflectivity of the wires, their visibility quickly falls off with distance from the center.

3.3.2 Quantitative Evaluation

To complement the qualitative comparison with a quantitative evaluation of both beamformers, we estimate the point-spread functions (PSFs) of both systems, following the approach of Jeong [50]. For this purpose, we imaged a wire target in a tank with distilled water at $48\,°\mathrm{C}$ (resulting in a speed of sound of $1540\,\mathrm{m/s}$ [8]) at different depths. We acquired the RF data after beamforming, both for the Cephasonics and SUPRA. Afterwards we performed envelope detection through the hilbert transform in a numerics software followed by a log compression to a dynamic range of $50\,\mathrm{dB}$.

Given the Dirac-like reflector, images representing the PSFs at depths 10, 15, 20 and $25\,\mathrm{mm}$ are retrieved. Figure 3.5 shows in the top row the PSFs of both systems at $20\,\mathrm{mm}$ and $50\,\mathrm{dB}$. While the lateral extent of Cephasonics appears larger than that of SUPRA combined with the Cephasonics system and probe, their height is comparable. This becomes even more clear in the bottom row of Figure 3.5, where the lateral PSF full-width-half-maximum (FWHM) of SUPRA is smaller in most depths, except for close to the transducer and the focal depth, whereas the axial FWHM of SUPRA and the Cephasonics beamformer differ only very little.

This result only partially agrees with the observations from section 3.3.1, where a more pronounced blurring could be observed for SUPRA, even in axial direction. This difference can be explained by the fact that the qualitative evaluation takes the complete imaging pipeline into account, while the PSF evaluation only considers the pipeline up to the beamforming. Overall, however, it shows that SUPRA can provide comparable image quality, following the purely software-based approach.

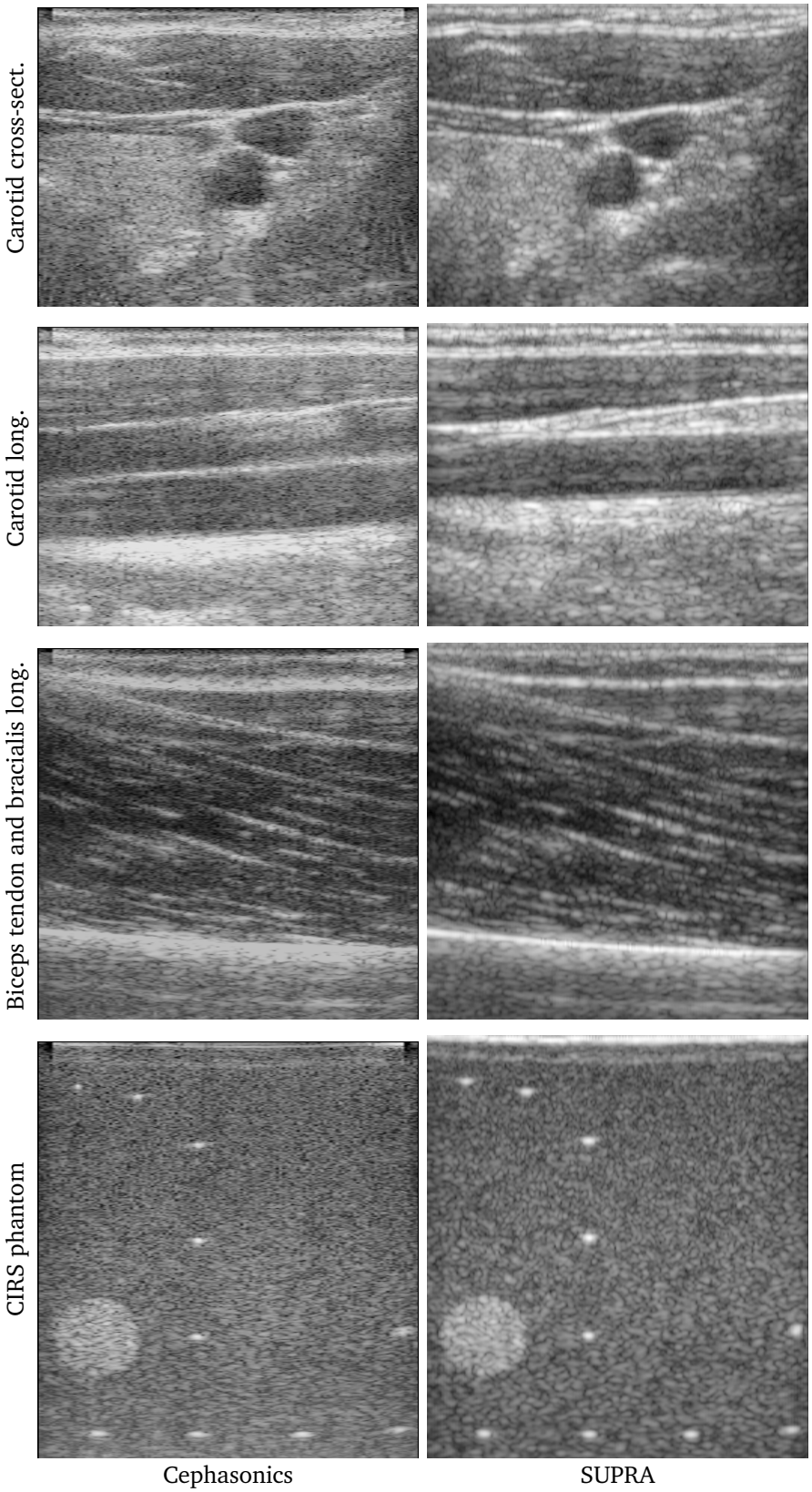


Fig. 3.3. **Qualitative comparison** of Cephasonics cQuest Cicada (left) and SUPRA with a Cephasonics cQuest Cicada (right) equipped with a linear transducer at $7.6\,\mathrm{MHz}$, depth $45\,\mathrm{mm}$. The first three rows show in-vivo acquisitions of the carotid (cross-sectional and longitudinal) and the biceps / brachialis of a healthy volunteer. The last row shows images of a CIRS multi purpose phantom (Model 040GSE).

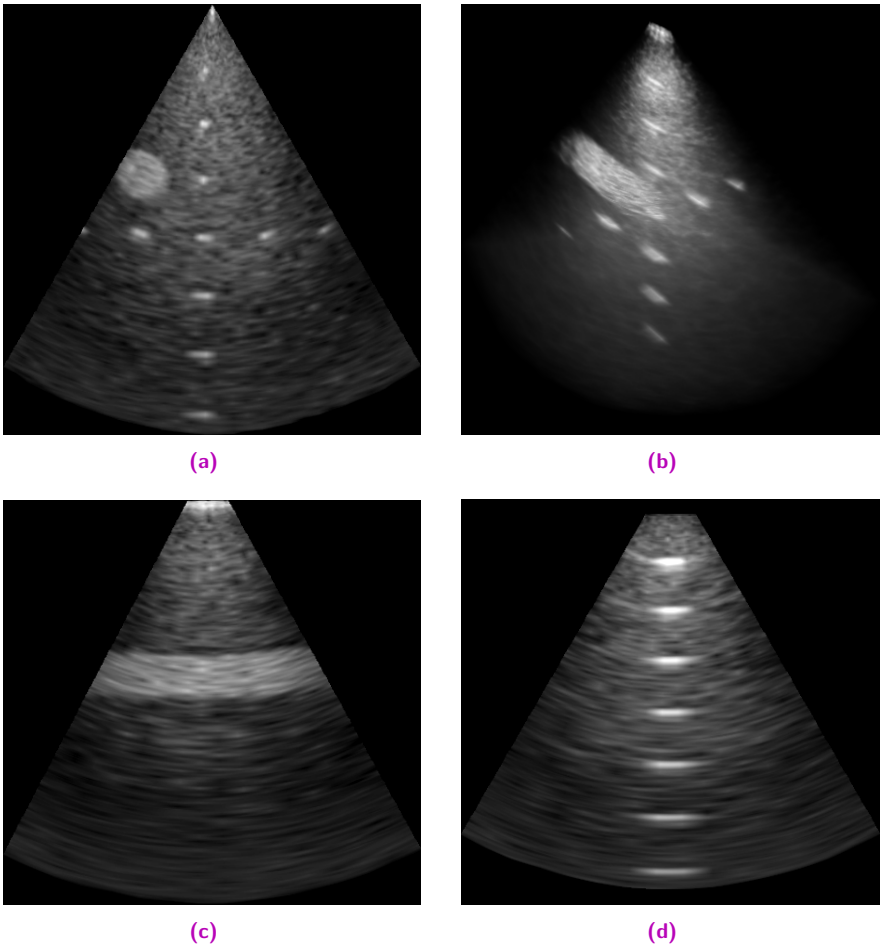


Fig. 3.4. **3D US volume acquired with SUPRA** of a CIRS multi purpose phantom (Model 040GSE). **(a)** shows a cross-section of the structures in the phantom. A rendering of the volume is shown in **(b)**. The second row shows two longitudinal views, **(c)** of a hyperechoic region, **(d)** wires. Note the limited visibility of the horizontal wires in **(d)**, caused by their highly specular reflectivity.

3.3.3 Performance

In the following, we present a run-time analysis of a SUPRA pipeline consisting of beamformer, envelope detection, log compression and scan conversion. We performed this evaluation on a number of computers

- Dedicated workstation with a NVIDIA GeForce GTX 1080 / 8 GB (Ubuntu Linux 14.04, Intel Xeon E5 - 1660 v4, 3.2 GHz, 8 core, 32 GB RAM)
- Notebook with a NVIDIA GeForce GTX 960M / 2 GB (Windows 10, Dell XPS 15 9550, Intel Core i7 6700 HQ, 2.6 GHz, 4 core, 16 GB RAM)
- Jetson TX2, an embedded SoC with a total power target of 15 W, includes a NVIDIA GPU sharing the main memory (Ubuntu Linux 16.04, ARM A57, 2.0 GHz, 4 core, 8 GB RAM)

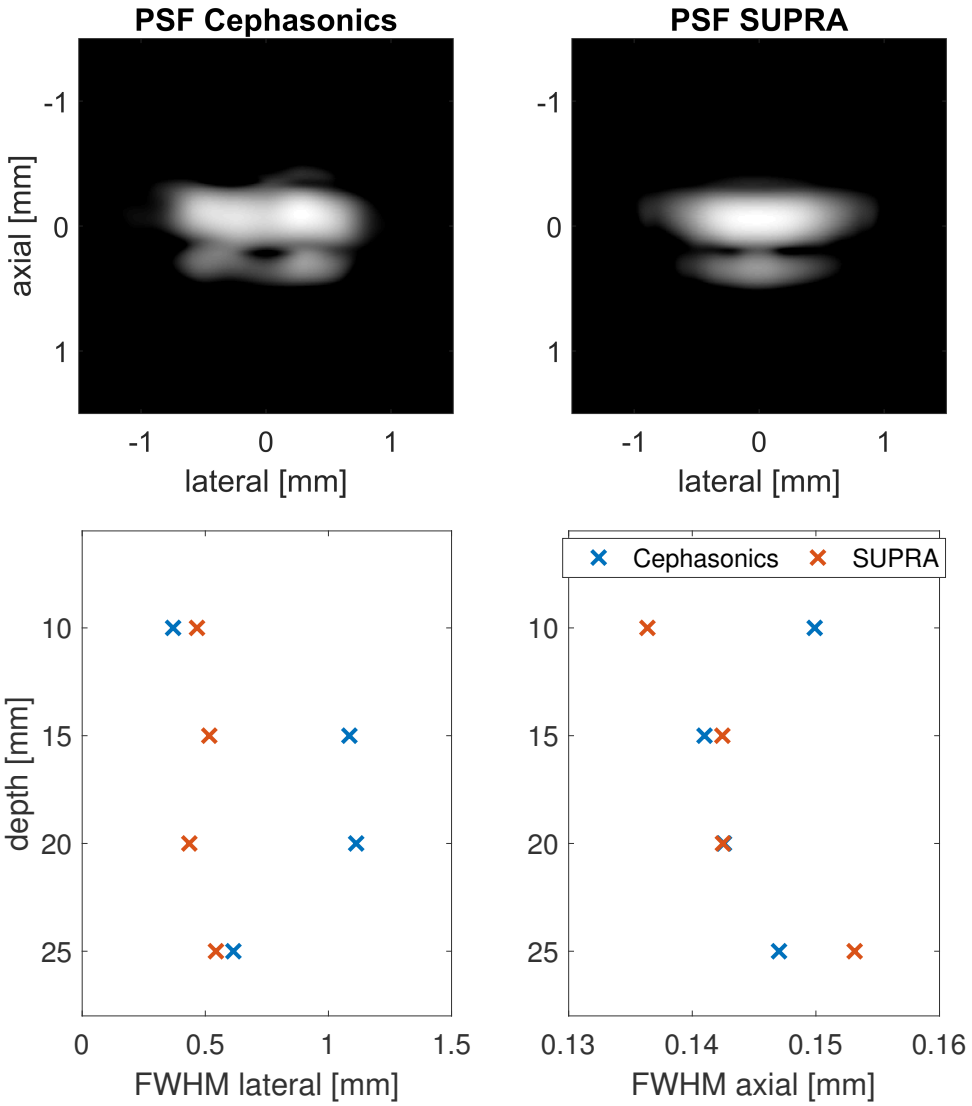


Fig. 3.5. **Point spread functions** (PSFs) measured for the Cephasonics and SUPRA beamformers with a linear probe. The top row shows exemplary PSFs at $20\,\mathrm{mm}$ depth and $50\,\mathrm{dB}$ dynamic range, the bottom row measurements of the full-width-half-maximum (FWHM) in lateral and axial directions.

As stated before, one of the design goals of SUPRA is its interactive use, consequently limiting the run-time of all nodes of a pipeline. Table 3.1 shows the node and pipeline run-times we observed in milliseconds on the different hardware configurations. Note that SUPRA used previously recorded raw channel data as input for the beamforming on the Windows laptop and the ARM SoC. From this table a number of observations are noteworthy. The benchmarks on the NVIDIA Jetson TX2 show that a pure software 2D US pipeline can be executed on an embedded device with reasonable frame rates. This has the potential to enable mobile US-systems based on software beamforming. It can be seen, that the log compression and scan conversion nodes exhibit only limited variation for the different scanline configurations. This is caused by two circumstances: While the log compression run-time is governed by the CUDA management overhead as it performs only limited computations, the run-time of the scan-conversion depends mostly on the resolution and size of the output.

	Scanlines	Beamformer	Envelope	Log comp.	Scan conv.	Total
Jetson TX2 SoC, Integrated GPU / 8 GB shared						
2D linear	(64 / 1)	7.02 ±0.89	9.00 ±1.31	0.41 ±0.16	20.91 ±2.88	37.64 ±8.22
	(64 / 2)	11.21 ±1.09	12.80 ±1.47	1.01 ±0.24	21.45 ±2.19	46.66 ±5.31
	(128 / 1)	11.09 ±0.48	11.32 ±0.96	0.75 ±0.14	19.68 ±1.19	43.43 ±5.30
	(128 / 2)	20.27 ±0.57	16.80 ±1.10	0.66 ±0.21	19.27 ±0.80	57.44 ±6.40
Notebook, NVIDIA GeForce GTX 960M / 2 GB						
2D linear	(64 / 1)	3.35 ±0.22	5.08 ±0.64	0.13 ±0.03	6.68 ±0.26	15.25 ±1.01
	(64 / 2)	4.97 ±0.28	6.60 ±1.00	0.70 ±0.22	7.03 ±0.45	19.28 ±2.07
	(128 / 1)	5.39 ±0.23	9.28 ±1.31	0.80 ±0.17	7.18 ±0.32	22.70 ±2.01
	(128 / 2)	9.04 ±0.24	11.04 ±0.90	0.74 ±0.19	7.14 ±0.26	28.15 ±2.65
Dedicated workstation, NVIDIA GeForce GTX 1080 / 8 GB						
2D linear	(64 / 1)	1.54 ±0.21	1.68 ±0.19	0.08 ±0.02	2.06 ±0.19	5.37 ±0.44
	(64 / 2)	0.93 ±0.04	0.98 ±0.17	1.10 ±0.19	1.23 ±0.03	4.24 ±0.47
	(128 / 1)	1.03 ±0.03	0.98 ±0.17	1.15 ±0.19	1.22 ±0.03	4.38 ±0.49
	(128 / 2)	1.68 ±0.03	1.98 ±0.29	0.08 ±0.03	1.23 ±0.02	5.00 ±0.48
3D phased		11.15 ±0.65	2.49 ±0.28	0.08 ±0.03	13.79 ±0.31	27.49 ±1.04

Tab. 3.1. **Observed pipeline and node run-times** [**ms**] for 2D and 3D imaging (mean and standard deviation). The 2D imaging was performed in different scanline configurations ranging from 64 reconstructed scanlines without multiline receive (64 / 1) to 255 scanlines reconstructed from data of 128 transmit events (128 / 2) scanlines with depth $45\,\mathrm{mm}$ and isotropic image resolution $0.0225\,\mathrm{mm}$. The 3D pipeline was parametrized as described in section 3.3.1.

As the GPU in the tested laptop is significantly faster than that present in the Jetson TX2, it is not surprising that the node run-times are lower. On the dedicated workstation with a NVIDIA GTX 1080 it is clear that even the 2D beamforming is limited by CUDA management operations as its run-time is not influnced by the scanline configuration. The overall 2D pipeline should thus be able to perform significantly faster than $100\,\mathrm{Hz}$.

In addition to the 2D pipeline profiled on all three machines, we executed a 3D pipeline on the dedicated workstation and measured its run-time as well. Although the number of scanlines was only twice as large as with the largest 2D pipeline, the beamforming took significantly more time. This is caused by the increased number of raw channels (384 vs. 64) and the resulting need to take those into account during beamforming. It can furthermore be observed, that the 3D scan conversion requires more time, which is caused by the addition of a whole dimension to its output. Due to the significant memory requirements and number of neccessary operations of the 3D pipeline, we did not execute it on the laptop and the Jetson TX2.

This run-time analysis is shows that a purely software based US pipeline as implemented with SUPRA can in fact be used for real-time imaging, even on commodity graphics hardware.

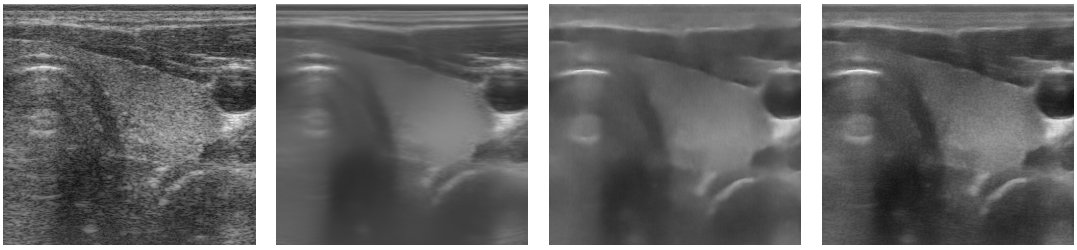
4 Deep Speckle-Removal

This chapter introduces our work in advancing the ultrasound image quality after acquisition, specifically the removal of speckle from images through a deep-learning based method. Sections 4.1 and 4.2 provide the motivation and place our work in the literature. The method for data generation and model training is described in section 4.3 and evaluated in section 4.4. In sections 4.5 and 4.6, we discuss the proposed method and provide an outlook.

4.1 Motivation

US imaging is a valuable tool to support non-invasive diagnosis and for use during interventions, as it is portable, affordable and does not subject the patient or staff to ionizing radiation. It is not, however, without problems. One of the major differences between CT and MRI on the one hand and US on the other hand is the presence of speckle, a grainy, noiselike image degradation that even affects tissue regions that are homogeneous in their microstructure such as liver or thyroid tissue (*e.g.* see Figure 4.1a). This speckle can make interpretation of the images difficult—especially for users with limited experience working with US. Consequently, most—if not all—US systems apply speckle reduction techniques after image formation, where the amount of "graininess" that is retained usually can be controlled by the user through a linear combination or blending of filtered and unfiltered data, depending on the clinical task and user preference. Figure 4.1 shows a thyroid image before and after filtering with our method compared to a state-of-the-art method and outlines a way to preserve a controlled amount of speckle. While for some tasks, such as the characterization or classification of masses, the presence of speckle is advantageous [80, 90], for others it can be detrimental. The tasks that benefit from speckle removal include interventional navigation using 3D visualization [94, 99] and the imaging of tissue boundaries [30].

The speckle patterns that emerge in US images are directly influenced by the location of subresolution scatterers in the medium. Since the distances between the scatterers are below the resolution limit, the reflections caused by the individual scatterers cannot be separated by the US imaging system. The result is interference between the echoes of different amplitude and phase, leading to an apparent echogenicity that changes depending on the spatial configuration of the scatterers within each resolution cell and the orientation of transmitting and receiving elements. In addition to speckle, there is also electronic noise that degrades US images, which is not the focus of this work. Speckle, while sometimes referred to as noise, is a deterministic phenomenon. In fact, if it were possible to place an US transducer at the exact same location when repeating a scan, the resulting speckle pattern would be exactly the same.



(a) Unfiltered **(b)** Filtered: OBNLM **(c)** Filtered: Ours **(d)** Blending

Fig. 4.1. Comparison of image appearance of a thyroid. Unfiltered **(a)**, despeckling with a state-of-the-art method (OBNLM) **(b)**, despeckling with our method **(c)** and a linear combination of **(a)** and **(c)** in **(d)**.

4.2 Related Work

Conventional despeckling approaches such as speckle-reducing anisotropic diffusion (SRAD) [128] and the Optimized Bayesian non-local means filter (OBNLM) [19]—itself building upon non-local means (NLM) [12]—have been studied extensively. We refer the interested reader to the review from Mohd Sagheer and George [79]. Especially the latter method OBNLM exhibits excellent despeckling performance, but requires application-specific tuning and significant processing time.

Cammarasana et al. [14] alleviate both these concerns by using a regression CNN that is trained against images despeckled with a tuned, computationally expensive conventional method, thereby allowing the CNN to replicate its results while lowering the execution time. Yang et al. [126] propose a dual-branch network operating on high- and low-spatial frequency components of the image (and feature-maps). They use clinical images despeckled with some other method as training target and add speckle following a simplified model to generate the input data for training. The performance of their approach is comparable with OBNLM. Khor et al. [53] apply a dual-path network after wavelet-transform, treating the low and high-frequency components of the images separately. In the absence of noise-free images, they propose an approach using a generative adversarial network, where the generator and discriminator are first trained on synthetic images of limited realism—paired noise-free and high-noise image, generated by the application of multiplicative noise to optical images. The networks are then fine-tuned with phantom US image pairs, where only high-noise and low-noise (images despeckled with another method) pairs are available for training. During fine-tuning, the networks task is to predict the low-noise image based on the high-noise image.

A number of approaches in the literature use learned methods that are trained on simulated US data. They then use the echogenicity map that was used to generate the simulated image as the training target, effectively creating a method for regression of the tissue echogenicity. Ando et al. [1] use Field-II simulated images as the network input and a spatial map of the scattering strength as the target image. This approach was also investigated by Jin and Palmeri [51] in their work on the effects of US data representation on the performance of deep learning methods. In a similar approach, Hyun et al. [43] use Field-II simulations to train a network whose output is a speckle reduced image. The echogenicity map is used as target

image. Contrary to the methods working on beamformed images, the input for the network is the individual US elements signals (channel data) as well as a beamformed image, limiting it's applicability to systems with access to the channel data. The use of the echogenicity map as training target, however, means that some phenomena specific to US imaging can not be directly represented in the data, leading to the potential removal of these image aspects. This includes shadowing, *e.g.* around stones, which can be used for detection and treatment decision based on their size [22]. It also becomes difficult to realistically account for direction dependant visibility of structures such as vessel lumen, bone interfaces or most prominent needles.

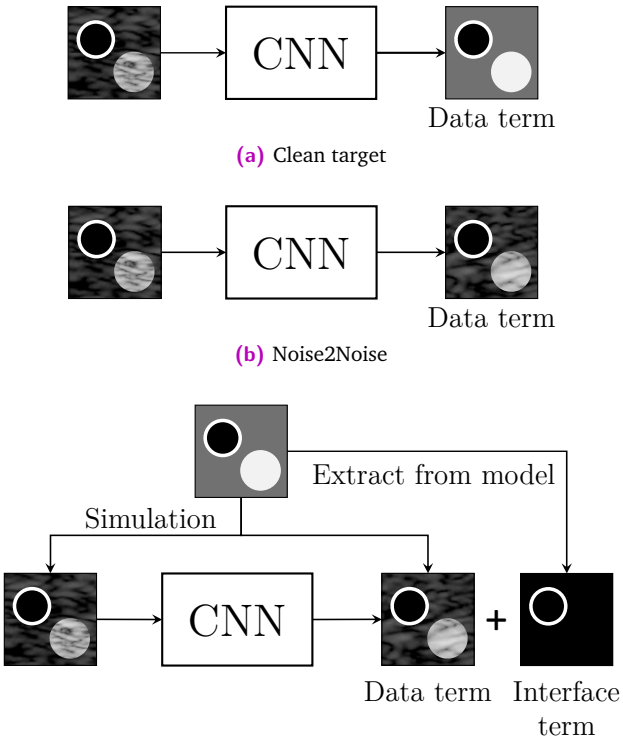
In other areas different to US imaging, there exists a body of work on the task of image denoising without requiring perfectly matched pairs of corrupted and clean data. In Noise2Noise [63], a CNN learns to restore images from a number of different corruptions, by using a corrupted image as input and a second image of the same "scene" with a different realization of the corruption as training target. The authors show that this can reach the performance of methods training against clean images. Noise2Void [58] takes this even further by requiring only unpaired corrupted images, it requires however, strong assumptions on the corruption, namely zero-mean and per-pixel independent noise.

Given the difficulties of acquiring clean images in many medical imaging modalities, these general approaches have been further specialized. In optical coherence imaging, a few well-chosen modifications to the acquisition setup can allow for the acquisition of two independent speckle instances with otherwise unchanged scanning setup [127], effectively allowing for direct use of Noise2Noise. However, this method cannot be directly applied to US imaging, as these modifications are not possible in US scanning setups.

In this chapter, we build on the concept of exploiting two independent *corrupted* observations with the same scanning setup and geometry, and transfer this to a generalizable method for US imaging. In this way, we present a technique for learning a CNN-based US despeckle filter only from images containing speckle. To do so, we apply the concept of Noise2Noise [63], combined with a scheme for US simulation for data generation, and modifications to the training loss to enhance the image appearance in a method we call Speckle2Speckle. By simulating different observations of the same general anatomical geometry but with different scatterer locations, we create images with uncorrelated speckle patterns. In our qualitative comparison and quantitative evaluation with state-of-the-art approaches, we show that the method can remove speckle from real US data efficiently while not relying on manual data annotation or acquisitions, while the application of the filter is several orders of magnitude faster.

4.3 Method

In this section we first summarize the original formulation of Noise2Noise, followed by a description how we create simulated US images to be used with a denoising approach like this, and the modifications that were required to account for the peculiarities of US. Figure 4.2 shows a comparison between the concepts for conventional (clean-target) [14], Noise2Noise and Speckle2Speckle filter learning.



(a) Clean target

(b) Noise2Noise

(c) Ours

Fig. 4.2. **Overview of training** a model with corrupted data input to clean data output **(a)**, Noise2Noise training with both corrupted data inputs and outputs of different realizations **(b)**, and Speckle2Speckle training with corrupted realizations of US data simulated from the same tissue model **(c)**.

4.3.1 Noise2Noise

Noise2Noise [63] is a deep-learning based technique for general image restoration that does not require clean, uncorrupted data for training. In a conventional restoration approach, one would use corrupted images as input and clean images of the same scene as the training target. In their work however, Lehtinen *et al*. derive, from a statistical perspective, that a network trained with infinite samples can learn to estimate the expectation of the target samples. This allows the input images as well as the target images to be corrupted with noise, like image noise observed with digital image sensors. The authors show that the networks trained in this manner perform on-par with networks trained against clean data.

These relaxed requirements on the training targets allow for the application of this technique to situations where the acquisition of clean targets is either costly or impossible all-together. In their work, Noise2Noise is applied to a number of different image corruption scenarios ranging from photographs subjected to Gaussian-, Poisson- and Bernoulli-noise or text-overlays, Monte-Carlo rendered images, and even MRI acquisitions in the randomly sampled k-space. The network architecture is a modified U-Net following [93], see Figure 4.3.

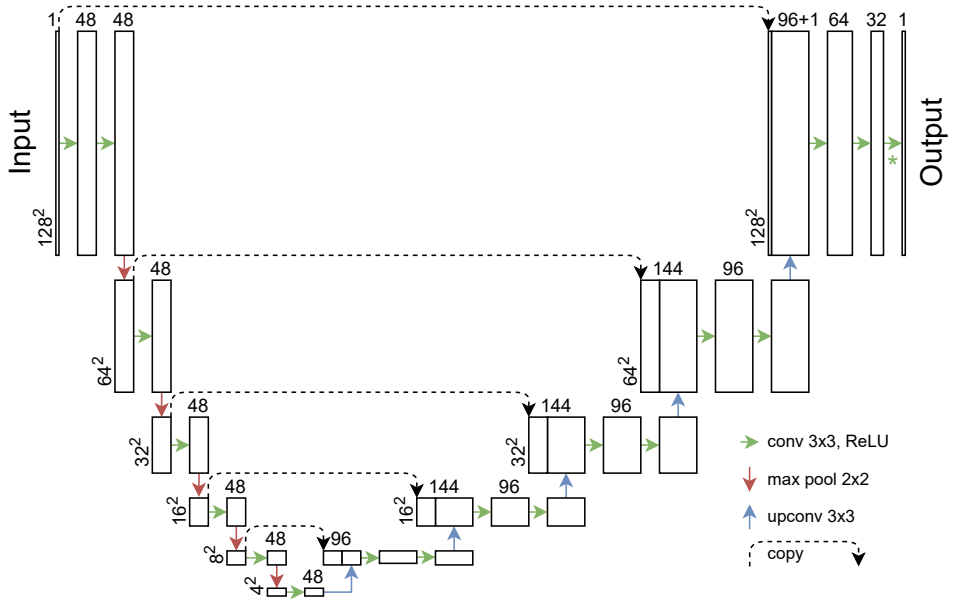


Fig. 4.3. **Network architecture of Noise2Noise and Speckle2Speckle.** Numbers left of each box show the sizes of the feature maps, exemplified with a 128 by 128 pixel input. The number of channels per feature map is given by the numbers above the boxes. Different arrows represent different operations. The rightmost convolution (marked with $*$) has linear activation.

4.3.2 Speckle2Speckle: Data

US imaging is a perfect example for a situation where a noise-free target image is difficult to obtain: since diffuse scattering is present for any US acquisition, speckle patterns remain visible at any time for acquired echo data. Thus, one can argue that in US imaging, there is no image without speckle. The strong correlation between the location and scattering behaviour of the tissue inhomogeneities causing the echos and the locations of extrema in the observed reflections (*i.e.* the speckle pattern), however, makes the acquisition of multiple images with independent corruption (of the desired echogenicity measurement) difficult for real acquisitions.

However, we can instead turn to realistic US simulations, where we can exploit the specific control over individual scatterers to generate an arbitrary amount of corrupted datasets for a given setup. The training data for Speckle2Speckle is generated with the simulation software Field II [48, 49]. Field II is based on the computation of the spatial impulse response of the US transducer elements. Assuming constant speed of sound throughout the medium and linear wave propagation (no generation of harmonic components), it accurately models the propagation of the excitation pulse to the scatterers, the reflections caused by them and the interactions between the reflected signals at the individual transducer elements. The resulting channel data—the signals received from each transducer element over time—is then processed in the same manner as is done with real US acquisitions to form images.

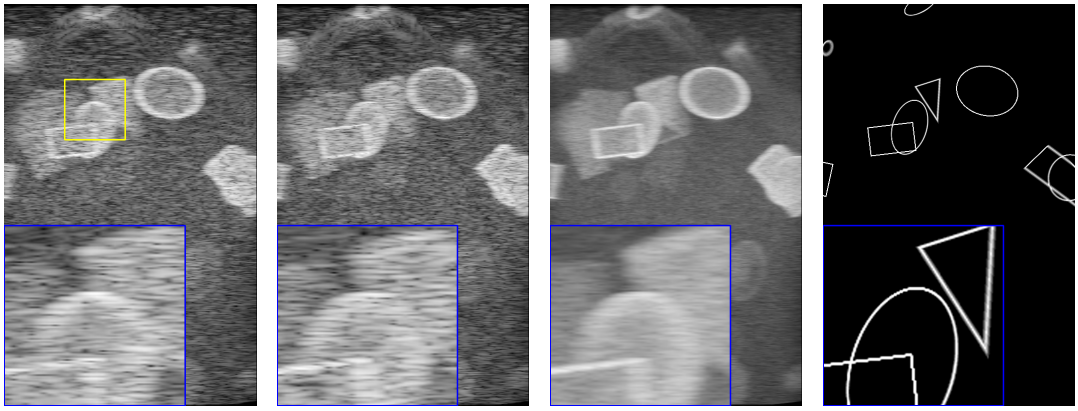
Images are generated in pairs, using independent in-silico scatterer phantom datasets (collections of scatterers with associated scattering strengths). These in-silico scatterer phantom

datasets are derived from the same geometric phantom, which is a collection of 3D shapes each of which with a number of parameters describing their appearance in US imaging and scatterer parameters:

- scatterer density
- scatterer amplitude distribution
- presence or absence of an interface
- interface scatterer density
- interface scatterer amplitude distribution

For the shapes that define the inclusions, we chose spheroids and cuboids. While neither of these shapes are common in biological tissue, the shapes exhibit common local features: varying inclusion surface curvatures including no curvature. Thus, the combination of these structures consider a wide range of interfaces as observed in biological tissue through the different angles, curvatures, and the overlap of multiple shapes.

Based on the these properties, this approach allows for the generation of scatterer phantom datasets with independent scatterer locations and is thus ideally suited for approaches such as Noise2Noise. Scatterer locations can be generated randomly, but since the "macro-scale" geometry (*e.g.* areas of higher scattering) is the same, the US images created from the different scatterer phantoms look structurally identical and even the speckle distributions are identical. The main difference, however, is that the speckle pattern "instance" is different, for which an example is shown in Figure 4.4a and Figure 4.4b.



(a) Instance 1 **(b)** Instance 2 **(c)** Average **(d)** Interface map

Fig. 4.4. **Comparison of the different image types created through simulation:** US images of two different scatterer instances from the same phantom geometry **(a)** and **(b)**, average of nine US images from the same geometry **(c)**, interface map derived from the phantom geometry **(d)**.

While the training is performed with pairs of scatterer instances, we use a different target for the testing and validation. In order to facilitate the evaluation of speckle removal in the absence of noise-free images, we compare the output of our method with an image that is composed by averaging multiple US images of the same phantom geometry but different scatterer instances. For the test and validation sets, we generated 10 instances per phantom,

one of which is used (randomly) as input, the other nine are averaged. Figure 4.4c shows the average of nine images corresponding to the same geometry as in Figure 4.4a and Figure 4.4b.

In addition to the US images, we also derive an interface map directly from the geometric phantom. Being of the same size and spatial origin as the simulated US images, it contains whether an interface of a shape was present at each location. Figure 4.4d shows the interface map associated with the geometry used for the simulations of the US images in Figure 4.4. This is later utilized to adapt the training loss to the challenges of US imaging in order to achieve enhancements of prominent interfaces in the output.

Following this general scheme, a dataset was generated with a training set consisting of 1000 image pairs while the validation and test sets each contain 100 phantom geometries with 10 speckle instances per phantom geometry, resulting in 4000 simulated images in total. Fundamentally, the size of the dataset was limited by the simulation time, where each image required roughly between one and two days on a single CPU core to simulate (depending on the concrete number of scatterers and the CPU performance), although parallelization between images was trivial.

Conventionally, the scatterer density is chosen at or above 10 scatterers per resolution cell λ^3 (where λ is the wavelength of the transmit pulse) to ensure fully developed speckle [115]. If the scatterer density is not high enough, there is the possibility of spurious dark-regions forming in areas that are intended as homogeneous medium. The run-time of the chosen simulation method, however, is linear to the number of scatterers in the scatterer phantom dataset. So in order to limit the simulation run-time, the scatterer densities values were empirically determined as high as required such that further increase did not cause a change in the speckle patterns.

The probabilities of geometries being anechoic and the scatterer amplitude distributions were determined arbitrarily to ensure visibility of the different regions. The specific parametrization used to determine the macro-scale geometric phantom properties is listed in Table 4.1, where "probability anechoic" represents the probability that a given geometric region (background or inclusion) is anechoic, *i.e.* devoid of scatterers.

The simulated imaging protocol used focused imaging and it followed the same restrictions as an actual focused scan using a CPLA12875 probe with a cQuest Cicada (both Cephasonics, CA, USA) would be subject to. Table 4.2 contains the detailed acquisition parameters.

The output of the simulations was US radio-frequency channel data, which was subsequently processed to US images using SUPRA [32] with parameters as in Table 4.3. After delay-and-sum beamforming, envelope detection and log-compression, the resulting images were converted to PNGs and interface indicator maps for inclusions with an interface were created by examining the geometry at every pixel location for the presence of an interface.

Region	Parameter	Value
Background	probability anechoic	0.4
	scatterer density	$0.333 \times 10^{10}\,\mathrm{m}^{-3}$
	scatterer amplitude	normal distributed $\mathcal{N}(0, \sigma_{\mathrm{bg}}^2)$; σ_{bg} drawn per phantom from $\mathcal{N}(1, 0.5^2)$
Inclusion	random objects	spheroids & cuboids, uniformly distributed
	number per phantom	100
	extents	drawn per axis uniformly from $[1, 5]\,\mathrm{mm}$
	probability anechoic	0.4
	scatterer density	$0.333 \times 10^{10}\,\mathrm{m}^{-3}$
	scatterer amplitude	normal distributed $\mathcal{N}(0, \sigma_{\mathrm{inc}}^2)$; σ_{inc} drawn per inclusion from $\mathcal{N}(4, 2^2)$
Inclusion Interface	probability interface	0.5
	scatterer density	$5 \times 10^{6}\,\mathrm{m}^{-2}$
	scatterer amplitude	constant per inclusion, drawn per inclusion from $\mathcal{N}(14, 2^2)$

Tab. 4.1. Simulated phantom parameters and values

Parameter	Value
transducer	CPLA12875
number elements	128
element pitch	$0.30\,\mathrm{mm}$
elevation focus	None
transmit frequency	$7\,\mathrm{MHz}$
transmit pulse	bipolar
imaging depth	$60\,\mathrm{mm}$
focus depth	$30\,\mathrm{mm}$
steering	$0°$
number scanlines	128
TX aperture size	32
RX aperture size	64

Tab. 4.2. Simulated acquisition parameters and values

4.3.3 Speckle2Speckle: Learning

The network architecture employed in Speckle2Speckle follows Noise2Noise (see section 4.3.1). The training loss, however, was adjusted for the peculiarities of US imaging. While the original Noise2Noise approach used different loss data terms depending on the noise, we extended the data term L_2 with a term to enhance the appearance of interfaces in the images, as interfaces retain key information for interpreting US imaging data.

The interface term is based on the known locations of interfaces given by the interface indicator map I_{i}. We chose a smoothed version of the interface indicator map $I_{\mathrm{i}G} = g_{\mathrm{i}} * I_{\mathrm{i}}$ as spatial weight for the interface loss, where g_{i} is a 2D Gaussian with standard deviation σ_i. This serves to i) extend the range of the interface effects from the potentially sharp peaks in the interface map, as well as ii) achieve a smooth transition between the effects of the data term and the

Parameter	Value
beamforming	DAS with hamming window
dynamic range	70 dB
image resolution	0.075 mm
image width	$37.6\,\mathrm{mm} \mathrel{\widehat{=}} 502$ pixels
image height	$60.0\,\mathrm{mm} \mathrel{\widehat{=}} 801$ pixels

Tab. 4.3. Image reconstruction parameters and values

interface term. The data term is spatially weighted with the inverse weight $\tilde{I}_{\mathrm{i}G} = 1 - I_{\mathrm{i}G}$. With that, we formulate the overall loss L_D as

$$L_D(I_\mathrm{o}, I_\mathrm{t}, I_\mathrm{i}) = L_2(I_\mathrm{o} \odot \tilde{I}_{\mathrm{i}G}, I_\mathrm{t} \odot \tilde{I}_{\mathrm{i}G}) + \lambda S_D(I_\mathrm{o}, I_\mathrm{t}, I_\mathrm{i}), \tag{4.1}$$

$$S_D(I_\mathrm{o}, I_\mathrm{t}, I_\mathrm{i}) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \left((I_\mathrm{o} * g_\mathrm{psf} - I_\mathrm{t}) \odot I_{\mathrm{i}G} \right)^2, \tag{4.2}$$

where I_o is the network output, I_t the target image, λ the weighting of the interface term, $\odot$ denotes the Hadamard product (element-wise multiplication) and g_psf a filter kernel of the point spread function (PSF). Ω is the set of pixel coordinates x in the images. The notion for the interface term is: In the target image, the interfaces are smoothed by the PSF by the process of imaging. Whereas in the output image the goal is to reduce the blurring, *i.e.* increase the sharpness. Consequently, the term $I_\mathrm{o} * g_\mathrm{psf} - I_\mathrm{t}$ penalizes deviations in I_o from sharp peaks. The PSF in this case could be the result of a per-image estimation, learned for one specific system, manually determined for one system or treated as hyper-parameter. We assumed the PSF to be an anisotropic Gaussian, with $\sigma^x_\mathrm{psf} = \mathrm{FWHM}_\mathrm{lat}/2$ and $\sigma^y_\mathrm{psf} = \mathrm{FWHM}_\mathrm{ax}/2$ determined experimentally from a simulated point reflector, where we measured the full width half maximum (FWHM) in lateral and axial directions. The hyper-parameters and their values are listed in Table 4.4.

Parameter	Value
λ	500
learning rate	3×10^{-5}
epochs	1000
σ_i (interface)	$0.375\,\mathrm{mm} \mathrel{\widehat{=}} 5$ pixels
σ^y_psf (axial)	$0.075\,\mathrm{mm} \mathrel{\widehat{=}} 1$ pixel
σ^x_psf (lateral)	$0.525\,\mathrm{mm} \mathrel{\widehat{=}} 7$ pixels

Tab. 4.4. Training hyper-parameters and their values

4.4 Evaluation

In order to achieve a fair comparison we evaluate our approach to a group of well-established and state-of-the-art speckle filters, and provide qualitative comparisons and quantitative evaluations for the simulated, phantom, and in-vivo datasets in the following. To this end, we compare the results of our method against SRAD[1] [128], median filter[2], bilateral filter[2] [111],

[1] implementation from [135]

non-local means[2] (NLM) [12], and optimized Bayesian non-local means[3] (OBNLM) [19]. Table 4.5 contains the parametrization used for the mentioned methods. These parameters were obtained through qualitative tuning on the in-vivo image in Figure 4.7a. The in-vivo image was chosen for tuning of the methods, as it represents the eventual use-case for all methods—for any system in clinical use, the processing pipeline would be optimized for in-vivo use. Additionally, while it would have been technically possible to perform quantitative tuning on the inputs and average images of the test set, the required computation time for the used implementations would have been prohibitive.

Method	Parameter	Value
SRAD	number of iterations	200
	lambda	0.1
Median	window size	15
Bilateral	degree of smoothing	0.05
	spatial sigma	5
NLM	degree of smoothing	0.075
	search windows size	101
	comparison window size	21
OBNLM	search area size	101
	patch size	45
	degree of smoothing	1.05

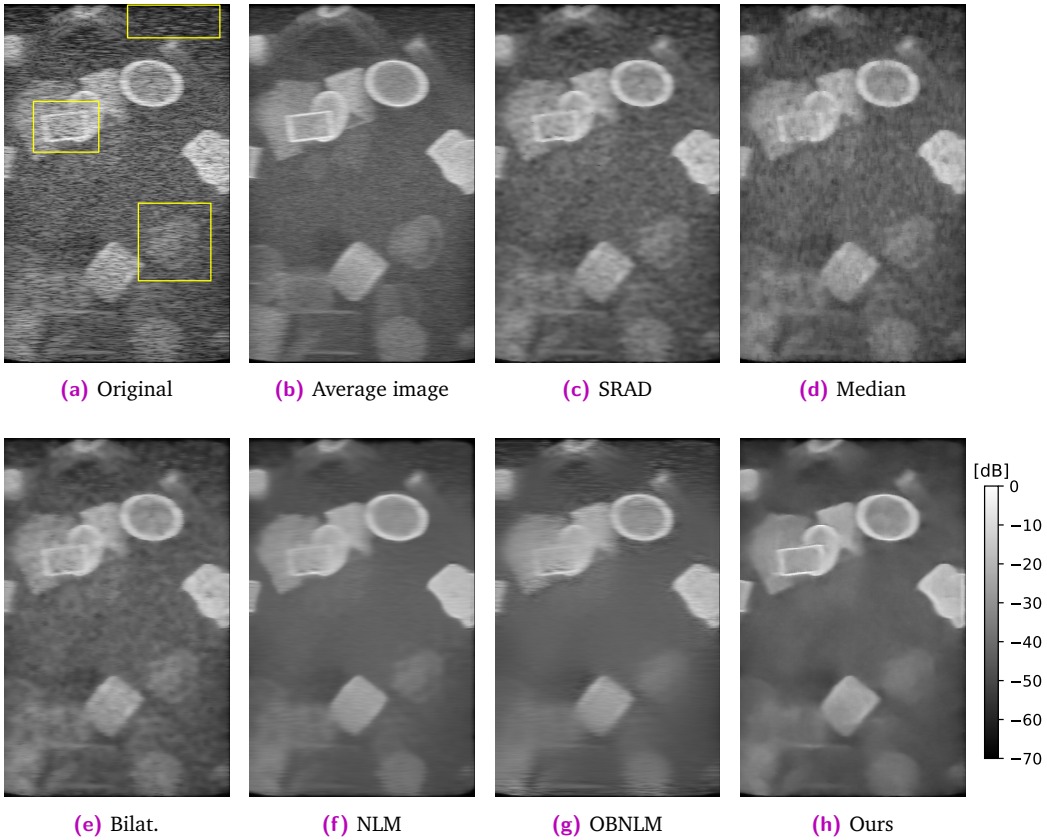
Tab. 4.5. Comparison method parameters and their values after manual tuning on the image in Figure 4.7a.

4.4.1 Qualitative Comparison

Simulated data. In the simulated image from the validation set (see section 4.3.2) shown in Figure 4.5, we can visually compare the result of applying the different methods directly with the average image of nine speckle instances. It is clear that SRAD **(c)**, median filtering **(d)** and the bilateral filter **(e)** mostly change the appearance of the speckle, but do not fully remove it. NLM **(f)**, OBNLM **(g)** and our method **(h)** all remove the speckle very well. In the outlined area in the top of the image, the less dense speckle is not fully removed by OBNLM. Another noteworthy region is the rectangular object in the top left that has a bright interface in the input as well as the average image. NLM and OBNLM do not retain this feature with high contrast, whereas our method shows contrast close to the average image. The interface indicator map corresponding to this image is shown in Figure 4.4d. The elliptic object in the bottom right appears in our method not as clearly as it is visible in the average image, we believe this is due to comparatively low brightness of interface around it. The other methods, however, also struggle to highlight the object.

Phantom data. The experimental image in Figure 4.6 shows an US quality assurance phantom (Model 040GSE, CIRS Inc., Norfolk, VA, USA). The image was acquired with a CPLA12875 probe and a cQuest Cicada (both Cephasonics, CA, USA), with equivalent imaging parameters as used for the simulations (see Table 4.2), but with an imaging depth of $45\,\mathrm{mm}$. Being acquired from a physical phantom, there is no average image as reference, but the distinct regions in the phantom are constructed as regions of homogeneous scattering. The image shows a background region, several string reflectors, as well as two different circular hyper-

[2] implementation from MATLAB R2020b
[3] implementation from [133]



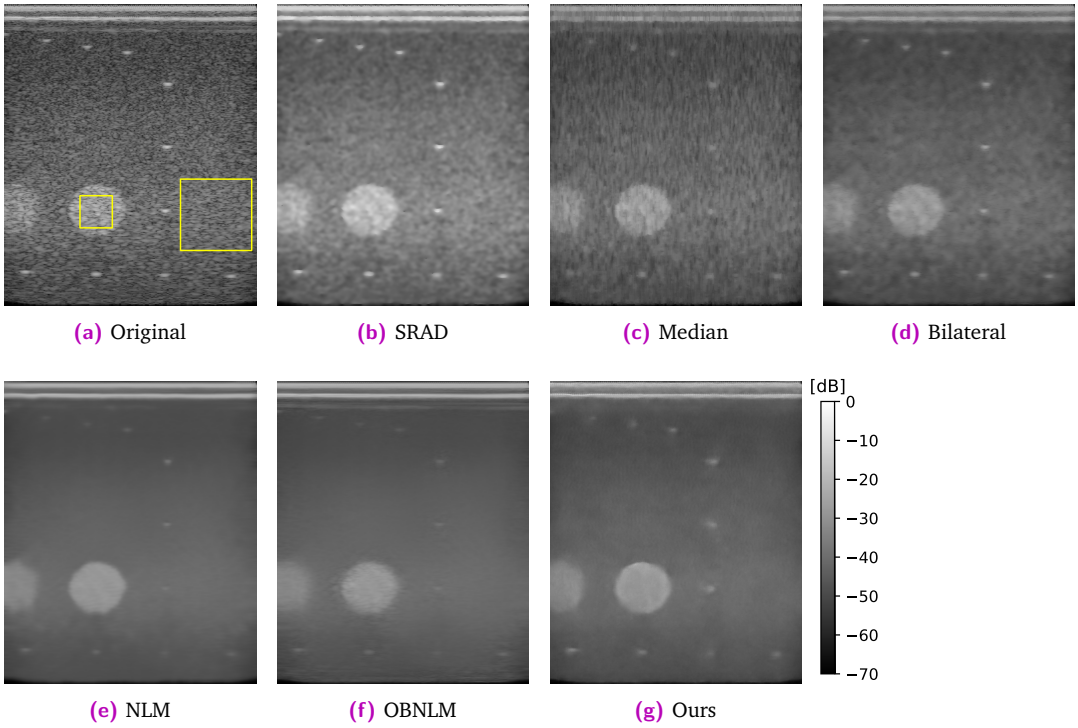
(a) Original **(b)** Average image **(c)** SRAD **(d)** Median

(e) Bilat. **(f)** NLM **(g)** OBNLM **(h)** Ours

Fig. 4.5. Qualitative comparison on a simulated image. The original speckled image (processed following Table 4.3) **(a)**, the average of nine different scatterer instantiations **(b)**, output of SRAD **(c)**, median filter **(d)**, bilateral filtering **(e)**, NLM **(f)**, OBNLM **(g)** and ours **(h)**.

echoic inclusions. The result of SRAD **(b)**, median **(c)** and bilateral **(d)** show the same general patterns as with the simulated data—not succeeding in completely removing the speckle. NLM **(e)**, OBNLM **(f)** and our method **(g)** are again close to that goal. While NLM shows the smoothest image among the three methods and shows a sharp edge of the bright inclusion, it suppresses the string responses more. The result of OBNLM is almost as smooth as NLM overall, but it shows inhomogeneities around the bright inclusion and its outline is not as well defined as with the other methods. Our method—while smooth locally—shows a certain inhomogeneity in the background, but does not suppress the point reflectors as strongly and has a clear demarcation of both inclusions.

In-vivo data. In the in-vivo image in Figure 4.7—a cross-sectional view of the thyroid of a healthy volunteer acquired with the same configuration as the phantom data with $35\,\mathrm{mm}$ depth—we can see that SRAD **(b)**, median **(c)** and bilateral filtering **(d)** perform better than in the simulated and experimental cases, although NLM **(e)**, OBNLM **(f)** and our method **(g)** exhibit a significantly more homogeneous appearance in the thyroid and especially OBNLM and our method retain more sharpness in comparison. While OBNLM better captures the muscle interfaces over the carotid artery (top right of the image) compared to our method, it is also less smooth in the overall area of the carotid. Our method does not retain said interfaces completely, but is not disturbed by the clutter within the artery.



(a) Original **(b)** SRAD **(c)** Median **(d)** Bilateral

(e) NLM **(f)** OBNLM **(g)** Ours

Fig. 4.6. **Qualitative comparison on an experimental phantom image.** The original speckled image (processed following Table 4.3, imaging depth $45\,\mathrm{mm}$) **(a)**, the output of SRAD **(b)**, median filter **(c)**, bilateral filtering **(d)**, NLM **(e)**, OBNLM **(f)** and ours **(g)**. The marked regions in the inclusion and background are used for quantitative evaluation.

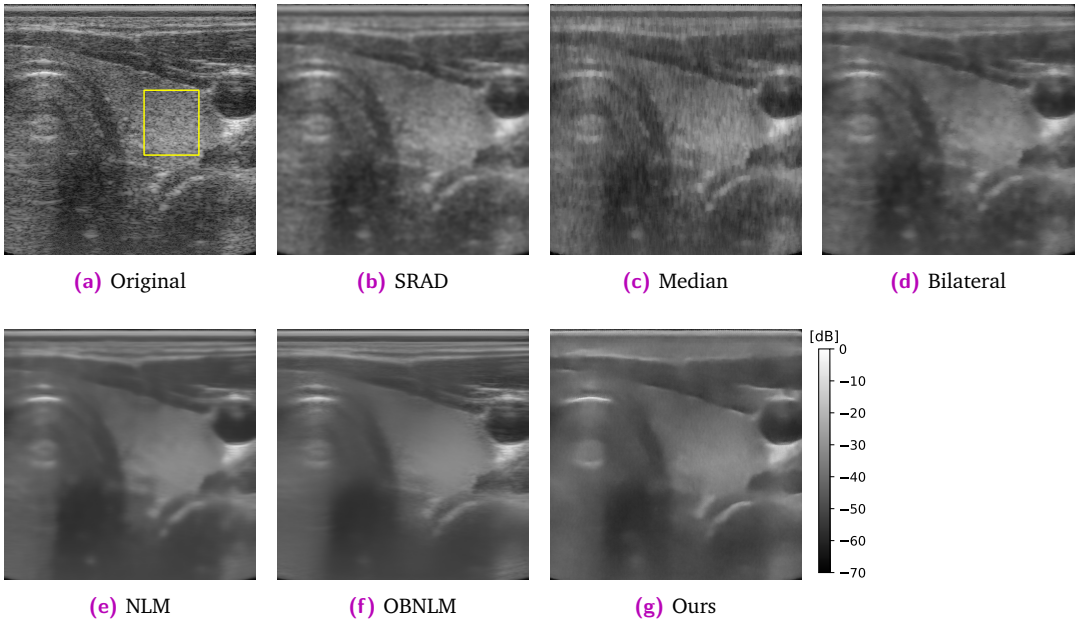
4.4.2 Quantitative Evaluation

Simulated data. For the quantitative evaluation in the simulated case we utilize the average images as an approximation of the speckle-free image. We perform this for all 100 simulated images in the validation set and the corresponding average images. The input images are scaled to the range $[0, 1]$ before application of any of the methods. Table 4.6 shows the mean and standard deviation of the mean squared errors (MSE) and the mean absolute differences (MAD) between the considered methods and the average images. All methods besides SRAD

Method	**MSE** ($\times 10^{-3}$)	**MAD** ($\times 10^{-2}$)
Input	5.87 ± 1.31	5.86 ± 0.87
SRAD	2.33 ± 0.86	3.61 ± 0.76
Median	1.82 ± 0.25	3.25 ± 0.33
Bilateral	1.48 ± 0.20	2.94 ± 0.30
NLM	$1.27 \pm \mathbf{0.09}$	$2.70 \pm \mathbf{0.13}$
OBNLM	1.48 ± 0.15	2.87 ± 0.20
Ours	$\mathbf{1.20} \pm 0.11$	$\mathbf{2.60} \pm 0.20$

Tab. 4.6. **Quantitative comparison on 100 simulated images** from the validation set. It shows the mean and standard deviation of the MSE and the MAD between images processed with each methods (and the unprocessed input) against the average images of nine speckle instances. Lowest error mean and standard deviation for both MSE and MAD are highlighted in bold without implying significance.

and median show a similar performance with respect to MSE and MAD, where our method



(a) Original **(b)** SRAD **(c)** Median **(d)** Bilateral

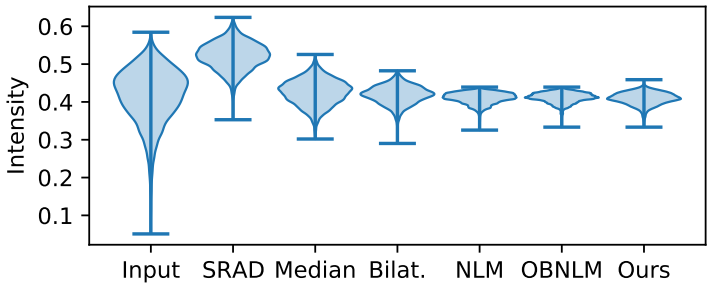
(e) NLM **(f)** OBNLM **(g)** Ours

Fig. 4.7. Qualitative comparison on an in-vivo image. The original speckled image (processed following Table 4.3, imaging depth $35\,\mathrm{mm}$) **(a)**, the output of SRAD **(b)**, median filter **(c)**, bilateral filtering **(d)**, NLM **(e)**, OBNLM **(f)** and ours **(g)**. The marked region in the thyroid is used for quantitative evaluation.

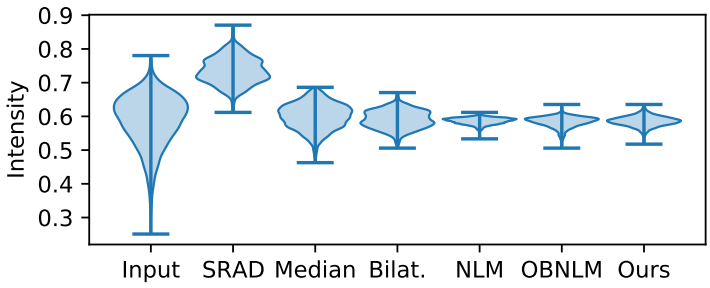
yields the best result. Note, however, that while the proposed method was trained on data from this domain, the comparative methods were tuned on the in-vivo image. In combination with the clearly varying image appearance of the different methods, this emphasizes the trade-offs between the methods compared here.

Phantom and in-vivo data. In the experimental and in-vivo images it is not possible to observe different speckle instantiations and thus no average image can be computed. Consequently, we focus on homogeneous image regions, namely the hyperechoic inclusion and the background for the phantom dataset (see Figure 4.6a) as well as the thyroid in the in-vivo dataset (see Figure 4.7a). Within those regions of assumed homogeneous scatterer distribution the ideal speckle removal filter would result in a flat response. Figure 4.8 shows violin plots of the intensity distribution in the marked regions of the phantom images, while Figure 4.9 shows those of the in-vivo image regions. The results of NLM, OBNLM and our method on the experimental image show very similar distributions; SRAD, median and bilateral-filter however are less tightly packed. This generally is matched by the evaluation for the in-vivo image data, but the distribution of our method is less pronounced than NLM and OBNLM.

In Table 4.7, we show the image intensity standard deviations within the homogeneous image regions. As expected from the violin plots, SRAD, median, and the bilateral-filter exhibit larger standard deviations in the three regions than NLM, OBNLM and our method. While our method does not outperform NLM and OBNLM—which share best performance within the three regions and were tuned on the thyroid image—it shows how well our method translates to real US data, despite not having been trained with real images.



(a) Background



(b) Hyperechoic Inclusion

Fig. 4.8. **Image intensity distribution** within homogeneous background region **(a)** and hyperechoic inclusion **(b)** of the experimental images.

4.5 Discussion

The qualitative and quantitative evaluations show that Speckle2Speckle works well in the simulated images, performing on-par with or even outperforming NLM and OBNLM, as can be seen in Figure 4.5 and Table 4.6. It also performs well in real acquisitions, although no real images have been used for training. There are shortcomings however, especially regarding the overall smoothness as in the background area and the reproduction of the point scatterers in Figure 4.6. Yet, it seems the highlighting of interfaces translates well to real images, with the exception of the linear structures in the muscle layer above the carotid artery in Figure 4.7.

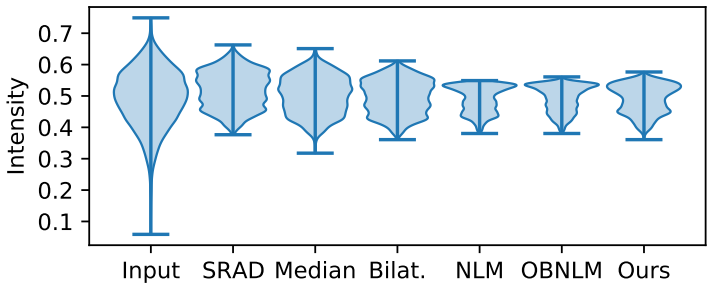


Fig. 4.9. **Image intensity distribution** within homogeneous thyroid region of the in-vivo images.

Region	**Experimental** ($\times 10^{-2}$)		**In-Vivo** ($\times 10^{-2}$)
Method	Background	Inclusion	Thyroid
Input	7.10	7.58	8.96
SRAD	3.55	3.96	5.89
Median	3.37	3.83	6.05
Bilateral	2.39	2.69	5.42
NLM	1.49	**0.96**	4.21
OBNLM	**1.32**	1.60	**4.20**
Ours	1.61	1.51	4.54

Tab. 4.7. **Quantitative evaluation on the experimental and in-vivo images** and shown in Figure 4.6a and Figure 4.7a respectively. The table shows the image intensity standard deviation in homogeneous regions in the background and foreground of the experimental images, and within the thyroid in the in-vivo case. Lowest standard deviation per region highlighted in bold without implying significance.

Since US is mostly used interactively, the run-time of algorithms in the imaging pipeline is of importance. The proposed method can be performed within tens of milliseconds on a recent GPU, making it suitable for real-time use.

4.6 Outlook and Future Work

We have shown with Speckle2Speckle a technique for training despeckling filters purely on simulated data as well as the generation of the simulated data. The concept is intriguing, since it allows for the generation of multiple speckle realizations from the same underlying geometry. The learned filter translates well to real acquisitions, performing similar to NLM and OBNLM, while allowing for real-time execution.

The method also enables the creation of matching maps with high-level information on the structures present in the image that can be used to guide the outcome of the method during training—as we have done with interfaces in this case.

As discussed by [126] in their ablation study, their network architecture that treats the low and high spatial frequency components separately can outperform pure convolutional residual networks in speckle removal. It would be interesting to explore the combination of the training and loss approach proposed here with the architecture proposed by them or the wavelet-based adversarial method described by [53].

Including single point scatterers and/or strings in the simulated images could improve the reproduction of those structures. They could be included in the interface map used for sharpening or placed in a separate map all-together, only depending on the desired appearance. This aspect of using additional geometric maps to tune the result for particular locations could possibly be exploited beyond speckle removal, for example for clutter filtering in blood vessels within the same filter pass.

5 Image Quality in Robotic Ultrasound Imaging

This chapter presents our work on image quality improvement in autonomous robotic ultrasound acquisitions. It is achieved through an US specific planning stage that takes the expected image quality into account.

Section 5.1 describes the motivation and related work on robotic US and planning thereof. In section 5.2 we present the proposed planning approach which is evaluated in section 5.3, followed by a discussion in section 5.4.

Substantial parts of this chapter have been published and are quoted verbatim:

[33] **Rüdiger Göbl**, Salvatore Virga, Julia Rackerseder, Benjamin Frisch, Nassir Navab, and Christoph Hennersperger. "Acoustic window planning for ultrasound acquisition". *International Journal of Computer Assisted Radiology and Surgery (June 2017)*, Creative Commons Attribution 4.0 International License (`http://creativecommons.org/licenses/by/4.0/`).

5.1 Motivation and Related Work

Sonography is a fundamental imaging modality for chronic cancerous [72] and non-cancerous [109] liver diseases. Novel developments in ultrasound (US) research, such as perfusion imaging [18], further contribute to its importance as a screening and interventional imaging device. Its main drawback, high operator variability, could be overcome by a robotic US imaging approach, that would allow for reproducible and precise data acquisition [40]. This will enable improved longitudinal studies and automated interventional US imaging, providing versatilities similar to other interventional imaging modalities such as Cone-Beam CT (CBCT), frequently employed in clinical practice.

Aiming at widespread applications, however, the automatic planning of US trajectories needs to be addressed not only in 2D, but also in 3D. The latter provides crucial information in a number of clinical settings, such as diagnosis of hepatic diseases [59]. For planning, both the US probe position as well as its orientation heavily impact the resulting image quality. Consequently, one has to account for the directional physics of US imaging with acoustic attenuation, potential shadowing, and other imaging artifacts such as reverberations. Beyond the basic US-related constraints, the probe position planning needs to consider anatomical constraints such as the patient surface, and optimize for the resulting (expected) image quality to avoid adverse objects (*e.g.* bones in US). Only this way, an optimal acquisition can be

performed for a given target anatomy. While a generalized planning can still be considered for easily accessible organs and structures such as the carotid artery, here we focus specifically on automatic acquisitions for organs with non-trivial acoustic windows, such as the liver or the heart.

Despite first approaches to US acquisition trajectory planning [34], a full optimization of 2D- and 3D-US acquisition trajectories with respect to the resulting image quality was to our knowledge not considered so far. In this regard, we introduce a novel planning framework for autonomous 2D- and 3D-US, and include geometrical, anatomy-based, and imaging-physics-based constraints to automatically retrieve the optimal position and orientation for a specific target point of interest. To optimize for the image quality, we integrate US attenuation estimates in our planning, which are derived from existing tomographic data such as CT and MRI. Ultimately, we aim at closing the gap for US trajectory planning, making autonomous US imaging more versatile.

In view of prior work covering planning of automatic US acquisitions, a general US probe path planning is proposed in [34]. The method allows for the full coverage of a region of interest, but does not consider the resulting image quality of the planned acquisition to optimize for appropriate acoustic windows. More recently, [40] and [113] showed the feasibility and accuracy of autonomous US acquisitions performed by a robotic system, introducing concepts for constant force acquisitions with lightweight robots. The focus of these studies was, however, not on planning of the US trajectories based on the optimization of acoustic windows, but on their actual execution. The proposed systems employ US confidence maps [41, 52, 55] to estimate the quality of acquired US images, which can only be used during the acquisition itself but not for quality simulation. Thus, they cannot provide a global optimal planning. For a targeted quality optimization in a planning stage, US simulation approaches as in [56, 97, 122] make use of the physics-based properties of US-imaging, and focus on the synthesis of realistic images or images with realistic appearance. They do not provide a measure of image quality to assess the anatomical constraints linked to determining the best acoustic window for an acquisition.

5.2 Acoustic Window Planning

In this section we introduce a method to determine optimized US probe positions for the acquisition of single US images. We show how a US sensor model can be used to integrate hard constraints, allowing the automatic planning of acquisitions for a target point (or structure) with a predefined US probe (Sec. 5.2.1). On this basis, we describe how acoustic transmission estimates can be used to retrieve the best acoustic window for a target structure (Sec. 5.2.2). Finally, the overall probe position planning is described, which incorporates the aforementioned parameters for optimization of the probe position and its orientation (Sec. 5.2.3). An overview of the proposed planning workflow is depicted in Fig. 5.1.

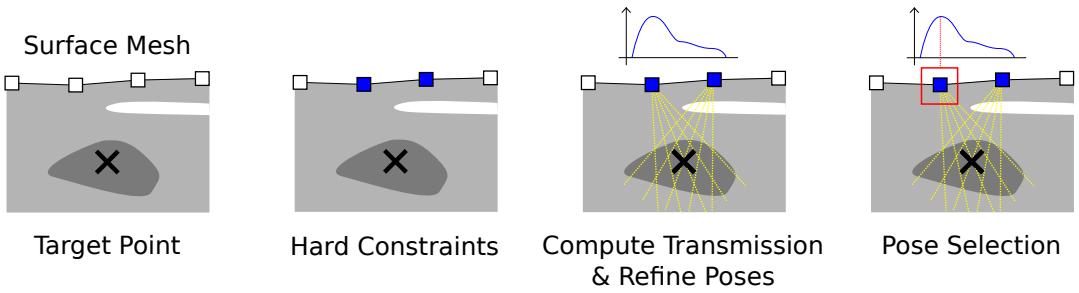


Fig. 5.1. **Schematic workflow of our method**. For a defined target point, pose candidates are selected according to the hard constraints given by the US probe and acquisition properties. Next, the transmission of acoustic waves is optimized such that the best possible pose is selected to cover a defined target point.

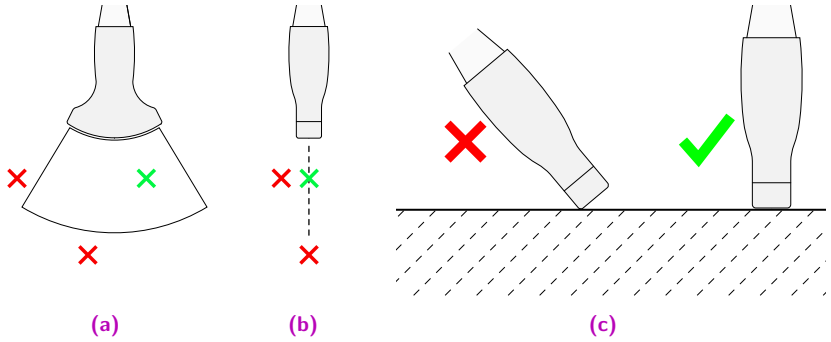


Fig. 5.2. **Constraints for US probe placement**, as determined by the image plane in axiolateral **(a)** and elevational **(b)** directions, as well as by the need for correct acoustic coupling of the US probe **(c)**.

5.2.1 Sensor Model

US imaging imposes certain requirements with respect to the probe positioning based on the underlying US imaging physics (longitudinal acoustic waves), the acquisition parameters (wavelength, field of view) as well as the transducers hardware design (piezo element size and array shape). During their training, physicians learn to intuitively regard for all these parameters to identify reasonable acoustic windows for certain target structures. To mimic a similar behavior for automatized acquisition planning, we model a set of constraints for US pose optimization and evaluate a set of pose-candidates with respect to the expected image quality. While 3D-US allows for improved structural coverage by acquiring volumetric information, at first we specifically focus on the foundation of 2D US imaging, and later extend the concept to 3D-imaging with freehand sweeps in Sec. 5.2.4.

The task of single view probe position planning is to find a corresponding US probe pose T_{US} to allow for *optimal* imaging of a target point P_t defined by the operator. For automatic acquisitions, we assume that pre-operative images are employed for the planning, in our case X-ray Computed Tomography (CT). From these images, a target point or structure is selected, and the probe positioning planned accordingly. For US imaging, resulting images are defined by the system settings (*e.g.* penetration depth d_y, frequency) and the probe geometry. With the discrete patient surface $\mathcal{S}$ extracted from the CT, we select a subset $\mathcal{S}_C$ of points $s \in \mathcal{S}$ for which the target point can be imaged based on these parameters. This is influenced by the

image depth and the axial orientation of the probe with respect to the corresponding surface normals n_s

$$\mathcal{S}_C = \{s \in \mathcal{S} : \ |P_t - s| < d_y \ \wedge \ \langle d_a, n_s \rangle < \alpha\}, \tag{5.1}$$

with the transducers axial orientation d_a restricted to the connection to the target-point $(P_t - s)/|P_t - s|$. By doing this, one ensures that the target is always within the image-plane, given the depth-constraint is fulfilled.

Fig. 5.2 shows intuitively how these constraints are enforced in order to allow for US probe planning. To maintain sufficient acoustic coupling and ensure patient comfort, we restrict the angle between the target (patient) surface and the US probe. This limit depends on a number of factors including the transducers physical geometry and the stiffness of the covering tissue-layers. During our experiments we used $\alpha = cos(30°)$. These constraints thus guarantee that the region of interest is inside the image and that the insonificating pulse can reach the patients skin.

5.2.2 Attenuation Estimation

The identified pose-candidates allow for the acquisition of images for a target region of interest, yet they do not regard for features of the patient anatomy, which could heavily influence the US image quality. In order to identify a suitable acoustic window for a target point, one has to ensure that sound waves reach the target with sufficient wave intensity. This thus corresponds to finding a linear path to the target that exhibits low attenuation while traversing the tissue.

Building on the efforts of Wein et al. [122], we propose an attenuation estimation based on CT, which is employed for acoustic window planning to evaluate possible poses as identified in Sec. 5.2.1 with respect to their expected image quality. The acoustic transmission coefficient of US waves through an interface of two tissues with acoustic impedances Z_1 and Z_2 can be written as

$$t(Z_1, Z_2) = 1 - \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2. \tag{5.2}$$

Using the approximately linear relationship between density ρ and X-ray attenuation coefficient μ in tissues [98, 122], this can be rewritten as

$$\Delta t(x) = 1 - \left(\frac{|\Delta\mu(x)|}{2\mu(x)}\right)^2, \tag{5.3}$$

where a constant speed of sound is assumed for simplicity. US waves can traverse several tissue interfaces, such that the overall transmission from a base point b along a ray of direction v is

$$t(x) = exp\left(-\int_0^a \left(\frac{|\Delta\mu(b + lv)|}{2\mu(b + lv)}\right)^2 \mathrm{d}l\right), \tag{5.4}$$

with $x = b + av, a \in \mathbb{R}^+$. To account for the processing in common US imaging pipelines, a log-compression is applied to the transmission estimate

$$\hat{t}(x) = \frac{log(1 + \nu t(x))}{log(1 + \nu)}, \tag{5.5}$$

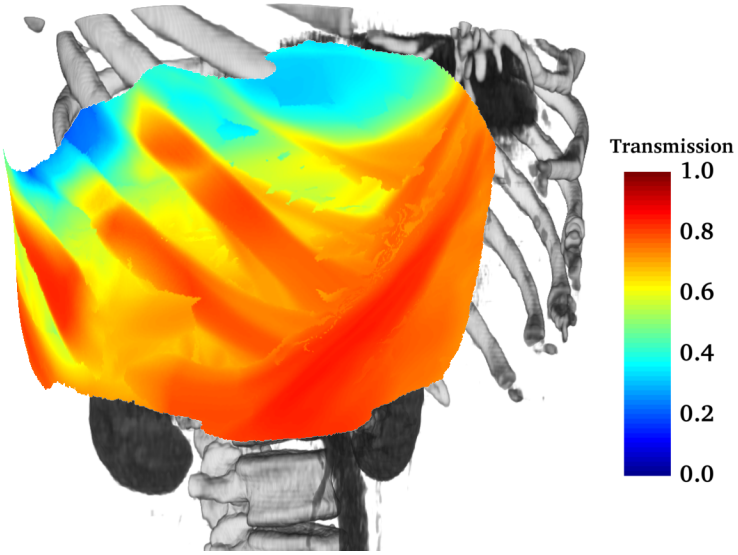


Fig. 5.3. **Mean transmission estimation** $\bar{t}$ for one target point in the liver, drawn for each surface point.

with $\nu = 0.5$ representing a constant compression factor in this work, comparable to the findings of Wein et al. [122]. As the respective US probe geometry and resulting image geometries are known a-priori (number of elements N_{el}, scan-line origins w.r.t. the central element b_i and their direction v_i), we can approximate the US transmission for a pose candidate and retrieve an average transmission value $\bar{t}$ for each relevant surface point $s \in \mathcal{S}_c$ and transducer orientation $R \in SO(3)$

$$\bar{t}(s, R) = \frac{1}{N_{el} d_y} \sum_{i=1}^{N_{el}} \int_0^{d_y} \hat{t}(R(b_i + l v_i) + s) \mathrm{d}l, \tag{5.6}$$

thus computing the mean along all scan-lines. As the average transmission is mainly influenced by strong reflectors (*e.g.* bone) between the target structure and the respective surface point, this allows for the identification of surface points with higher wave intensities (and thus better signal to noise ratio) at the target structure. Fig. 5.3 shows the exemplary mean transmission values estimated for each point on the surface for a target point inside the liver.

5.2.3 Probe Position Planning

On the foundation of the hard constraints and the transmission model, we can retrieve a patient-specific optimal US pose to cover a target point with maximized acoustic intensity. Anatomical factors such as the ribs will cause strong reflections (*i.e.* low transmission) for certain probe orientations, whereas a change in its position potentially has smaller effects on the expected transmission. We propose a two-stage quality maximization, where the best angle is retrieved first, since the transmission values should be almost convex w.r.t. the probe angle. The surface position s is then optimized to retrieve the final US probe pose $T_{US} = (\hat{R}, \hat{s})$.

In our case, the US probe pose is already characterized by a base point on the patient surface $s_i \in \mathcal{S}$ and the axial orientation d_a. Therefore, the goal of the first optimization stage is to find the best transducer orientation R_{US} for the acquisition in one degree of freedom. By defining

the axial direction to be $d_a = (P_t - s_i)/|P_t - s_i|$, we effectively align the transducer center towards the target-point and reduce the space of eligible orientations to the rotations around d_a. The mean transmission $\bar{t}$ can then be employed as quality metric in order to maximize the overall transmission for a given target position

$$(s, \phi) = \underset{(s,\phi)\in\mathcal{S}_C\times[0,\dots,\pi[}{\arg\max} \bar{t}(s, R_{d_a}(\phi)), \tag{5.7}$$

where $R_{d_a}(\phi)$ is the rotation around the fixed axis d_a by the angle ϕ. Intuitively, by maximizing the transmission across one image and different image candidates, those with a low transmission are rated lower and discarded in the optimization.

5.2.4 Planning of 3D-trajectories

Generalizing the proposed method for 3D-trajectories, we optimize the best acoustic window for the acquisition of all the N_t target points P_{ti} within a structure. The 3D-trajectories planned by our method are restricted to one base point $s \in S$, and only vary with respect to their orientation. By choosing a point that allows for high transmissions to all target points, we ensure that all poses are within an acoustic window.

$$(s, (\phi_i, \dots, \phi_{N_t})) = \underset{(s,(\phi_i,\dots,\phi_{N_t}))\in\mathcal{S}_T\times([0,\dots,\pi[)^{N_t}}{\arg\max} \prod_{i=1}^{N_t} \bar{t}(s, R_{d_a}(\phi_i)), \tag{5.8}$$

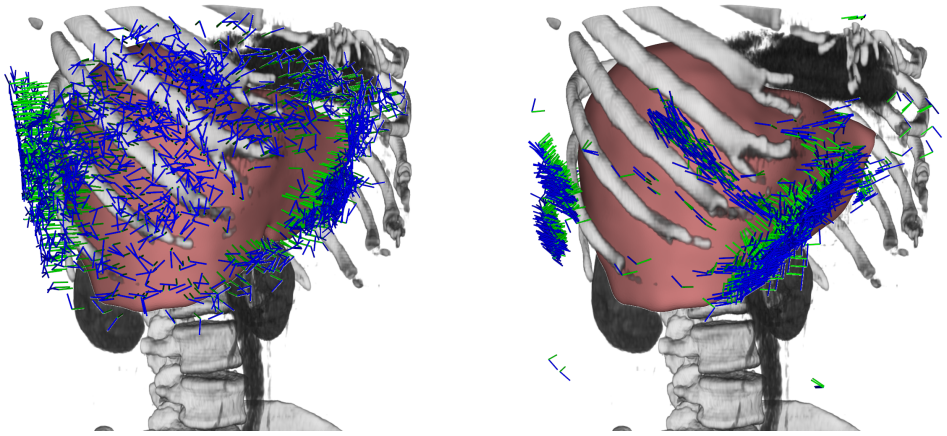
where the set of surface points has to fulfill the hard constraints w.r.t. every target point $\mathcal{S}_T = \bigcap_{i=1}^{N_t} \mathcal{S}_{Ci}$. We follow the same two-step maximization approach as above by first selecting the best rotations for each target point and surface point $s \in \mathcal{S}_T$, followed by selecting the base-point with the overall best transmission.

5.3 Experiments

As acoustic window planning has not been considered so far, we compare the results of the proposed planning framework to a naive planning, comparable to the planning approaches in [34, 40]. This consists of choosing the surface point s_n nearest to the target point P_t as base-point for the acquisition. The transducer orientation d_a is then chosen in the same way as described in Sec. 5.2.3, while minimizing the angle between the transducer and the surface normal n_n at s_n. By doing so, the transducers lateral axis d_l is $(n_n \times d_a)/|n_n \times d_a|$. When multiple views are considered, the naive approach is to choose the base-point and d_a as for single-views and the elevational direction d_e as the rejection of the input-trajectory direction d_T from d_a, aiming at trajectories with parallel image planes.

$$d_e = \frac{d_T - d_a \langle d_a, d_T \rangle}{|d_T - d_a \langle d_a, d_T \rangle|} \tag{5.9}$$

We first perform a set of experiments on publicly available datasets (Sec. 5.3.1), as well as for a torso (rib) phantom, where scans are performed with a robotic US system (Sec. 5.3.2).



(a) (b)

Fig. 5.4. **Comparison of single-view poses for the N_s target points** for one subject from the SLIVER07 dataset [39] together with a render of the underlying CT volume. The orientations of the image axes are visualized with colored lines, **green** for axial and **blue** for lateral directions. **(a)** shows the plans created naively. The results of our method are shown in **(b)**.

The planning was performed on a workstation (Intel i7-4820K, NVIDIA Titan Black) and the computation of the mean transmission was implemented in CUDA. Computing the best poses for 10 target points took on average 356 seconds.

5.3.1 Synthetic Planning

Using a dataset of 20 annotated upper-torso CTs from the SLIVER07 challenge dataset [39] we performed an evaluation of the proposed single and multi view planning method. The volumes featured different portions of the thorax and abdomen respectively, but all contained the liver with some margin. For our evaluation, we chose $N_s = 2000$ random points inside the liver segmentations and manually defined between 4 and 6 plans for each volume to cover the large vessel trees. This resulted in a total of 100 multi-view plans for all evaluated datasets.

Stability

We demonstrate the effectiveness of our acoustic window planning approach qualitatively by independently computing the best poses for the N_s random points of one case. Fig. 5.4 shows the resulting pose in combination with a visualization of the corresponding CT volume. The poses computed with the naive approach are widely spread over the thorax and a significant number of image planes intersect with ribs, resulting in a majority of images with strong reflectors and prominent artifacts in images. The proposed method results in poses that closely follow the acoustic windows in the intercostal spaces. Inferior to the costal cartilage, the image planes are oriented tangent to the rib cage, suggesting that the planned poses are suited better for US-imaging.

Quality of Acoustic Window

A quantitative comparison of both planning methods was performed using the ratio of non-soft-tissue areas r_{CT} and the ratio of depicted liver r_{seg}, as given by the segmentation,

$$r_{\text{CT}} = \frac{1}{n} \left| \{x : \mu(x) < \beta_1 \vee \mu(x) > \beta_2\} \right| \tag{5.10}$$

$$r_{\text{seg}} = \frac{1}{n} \left| \{x : x \in \text{segmentation}\} \right| . \tag{5.11}$$

Where n is the number of pixels in the image, and $\beta_1 = -100\,\text{HU}, \beta_2 = 150\,\text{HU}$ separating soft from hard tissue following [98]. The ratio of non-soft-tissue r_{CT} indicates the fraction of dense tissues (*i.e.* bones or air-filled areas) contained in the image, which impairs the transmission of US pulses across soft-tissues. Consequently, a lower ratio exhibits potentially better overall image quality, as a large r_{CT} potentially also causes US artifacts. Conversely, the ratio of the target organ r_{seg} depicted in a target image serves as measure of how much anatomical context is provided in the resulting US images. A higher r_{seg} provides more context, as a higher fraction of the target organ is covered by the image content.

The evaluation was performed on CT slices at the respective pose with the size of the US image, as shown in Fig. 5.5. We used the single- and multi-view plans created using the segmentations of each of the 20 datasets. Table 5.1 shows the quantitative results for both evaluated measures. As the mean ratio of non-soft-tissues is lower for our method, while the ratios of depicted target anatomy are roughly equal, this effectively shows that the poses resulting from our method are less likely to have shadowing artifacts, while allowing for a similar coverage of the target image region. In Fig. 5.5, the pose obtained with naive planning cuts through several ribs, which would result in strong shadows in an US-image. Our method planned a pose that only contains non-optimal regions at the bottom and is likely to have no shadowing artifacts in the liver.

5.3.2 Robotic Acquisition Experiments

The two methods were evaluated on real acquisitions performed by a robotic US system. Intuitively, the manual execution of planned trajectories is prone to errors and would lack accuracy even in the case of statically placed phantoms. Furthermore, the reproducibility of such trajectories would be limited significantly. Robotic US well tackles these limitations and allows for the autonomous execution of planned probe trajectories. In this view, the robotic system previously presented in [40] and [113] represents a good choice to qualitatively and quantitatively assess the performance of our proposed planning technique for US acquisitions.

	$\bar{r}_{\text{ct}}$ **Ours**	$\bar{r}_{\text{ct}}$ **Naive**	$\bar{r}_{\text{seg}}$ **Ours**	$\bar{r}_{\text{seg}}$ **Naive**
Single view planning	$\mathbf{0.222} \pm 0.163$	0.294 ± 0.174	0.488 ± 0.171	0.477 ± 0.161
Multi view planning	$\mathbf{0.214} \pm 0.175$	0.256 ± 0.179	0.470 ± 0.225	0.456 ± 0.214

Tab. 5.1. **Acoustic window quality** mean and SD for the naive and our planning technique applied to single and multi view settings. Columns **1** and **2** show the ratios of non-soft-tissue values, columns **3** and **4** the ratios of depicted liver.

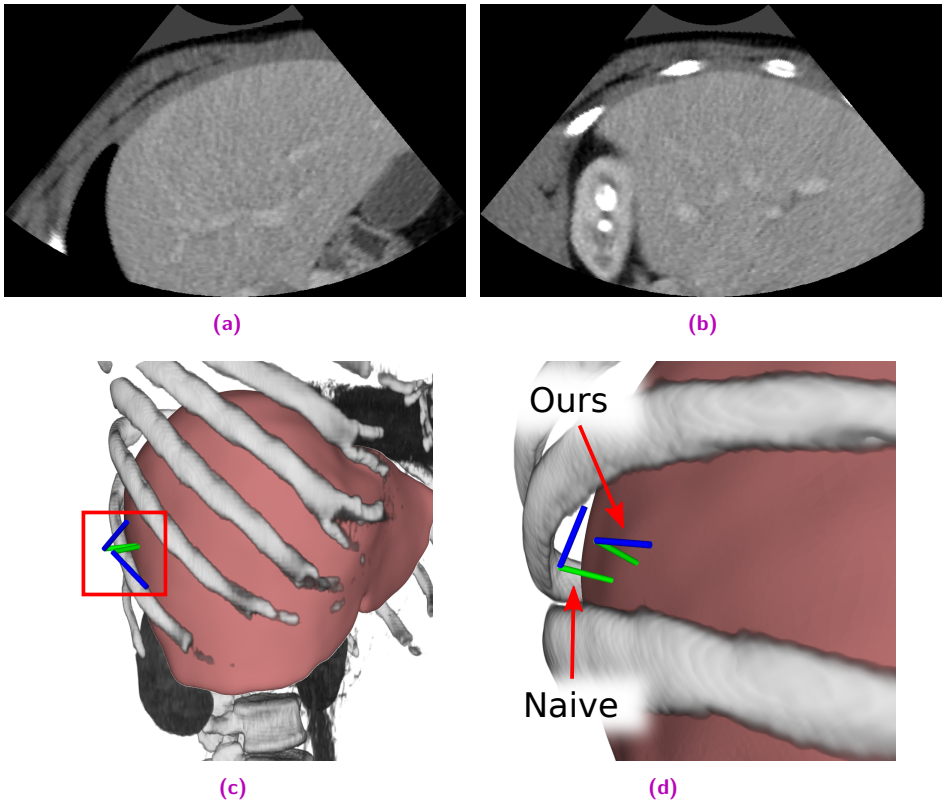


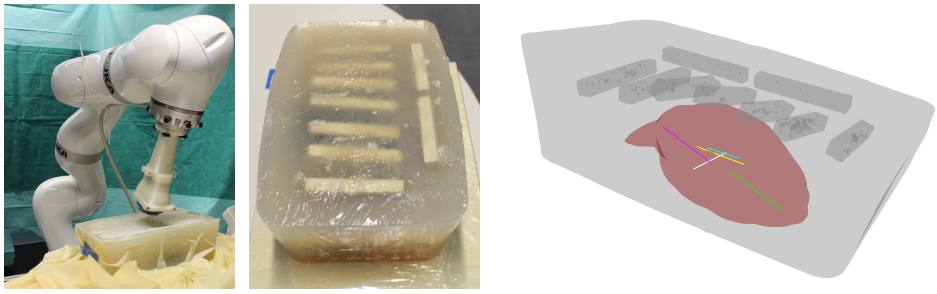
Fig. 5.5. **Comparison of synthetic poses for one target point**. **(a)** shows the CT slice at the proposed pose, **(b)** at the one resulting from naive planning. **(c)** and **(d)** set the poses in relation to the surrounding anatomy.

The system is composed of a robotic arm, KUKA LBR iiwa R800 (KUKA Roboter GmbH, Augsburg, Germany), controlled using a custom software module[1] and the Robot Operating System (ROS) framework. An RGB-D sensor (Kinect, Microsoft Corporation, Redmond, WA, USA) is positioned above the examination bed and used to perform CT to phantom calibration. The US acquisitions are obtained from an Ultrasonix® Sonix RP US system equipped with a 4DC7-3/40 curvilinear transducer, using the following parameters: frequency: 3.3 MHz, depth: 140 mm, gain: 50%. ROS and the respective modules to control the robot ran on a PC (Intel Core i5, NVIDIA GTX 970) communicating with the planning system.

Using this system, the acoustic window planning was evaluated for single and multi-view acquisitions on a gelatin-agar phantom based on [23]. The tissue mimicking material is targeted to multi-modal imaging. A liver was molded from a gel made out of 10 weight percent ($wt\%$) gelatin and $6wt\%$ agar, diluted in water. For enhanced scattering in US $0.6wt\%$ graphite powder were added. The surrounding tissue mimicking gel consisted of 3 and $1.5wt\%$ gelatin and agar, respectively. Shortly before complete solidification of the gel, the liver model and chalk sticks to simulate ribs were added.

Acquisitions were planned in a CT volume of the phantom and target acquisition trajectories were computed both with our and the naive planning method. The 5 target lines (27.0–53.5 mm) defined for the acquisitions with differing orientations w.r.t. the ribs are shown in

[1] `https://github.com/SalvoVirga/iiwa_stack`



(a) (b) (c)

Fig. 5.6. Phantom and robotic system used for experiments. (a) Robotic arm equipped with US transducer approaching the phantom covered by a thin latex sheet. **(b)** Gelatin-agar phantom, chalk bars resembling ribs are visible. **(c)** Rendering of the phantom, the selected trajectories are shown in distinct colors.

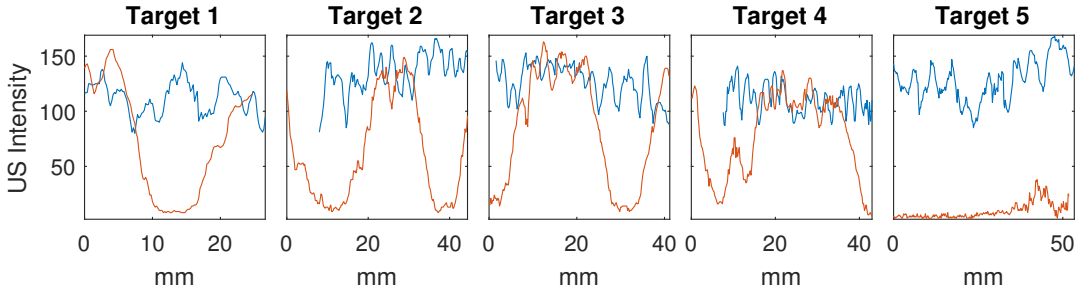


Fig. 5.7. US-intensities along planned trajectories in phantom for naive and the proposed planning technique, ours in blue, naive in orange. For target points that were not covered by the acquisition, no value is shown.

Fig. 5.6c. After planning, the trajectories were executed by the robotic setup as in [40]. A constant force of 5 N onto the phantom's surface was applied for all US acquisitions. They were reconstructed into 3D volumes (compounding) and employed for evaluation.

Fig. 5.7 shows the US intensity profiles measured along the 5 defined target lines. Blue plots display the intensities from acquisitions planned with our method, while the orange plots show the naively planned ones. Where target points were not covered by the US acquisition, no values are shown. Such regions result from inherent inaccuracies of the surface to surface registration employed in [40], as well as by tissue deformation. It can be clearly seen that naively planned volumes exhibit high intensity variations along the line within homogeneous tissue, while ours show a continuously high visibility, only containing speckle variations.

Slices of two acquisitions are shown in Fig. 5.8 together with the target lines inside of the volume, comparing naive planning to our method. As it can be observed, the naively planned trajectories cross bones and the resulting images are subject to strong shadowing artifacts in these locations. In contrast to this, the plans created with our method allow for imaging of the selected structures without shadowing throughout the sweep.

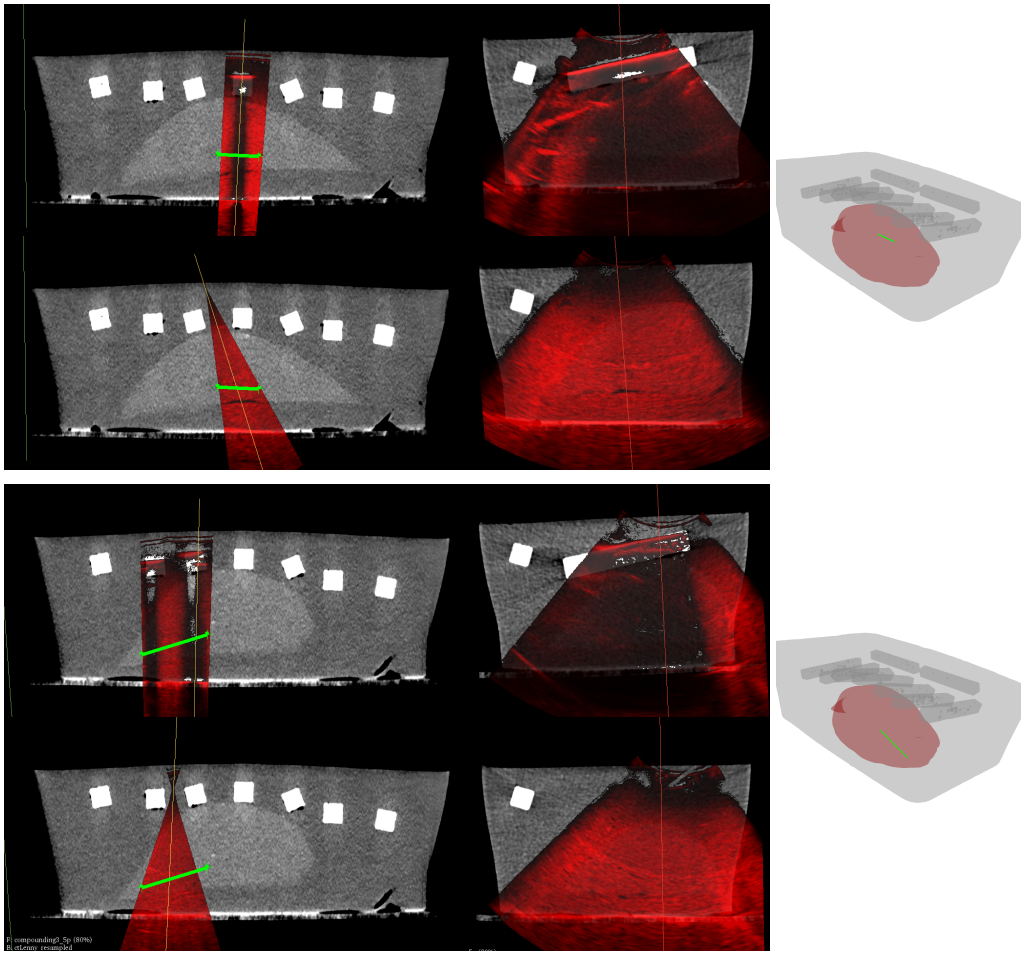


Fig. 5.8. **Phantom US acquisitions (red) for two trajectories** using the CT for planning. Rows 1 and 3 show the naive sweeps, rows 2 and 4 sweeps according to our planning, each in different planes (first and second column). The positioning of the target points is shown as green line in the rendering and the slices.

5.4 Discussion

On the foundation of our analysis and presented work, the results of both the synthetic trajectory planning as well as the phantom acquisitions show the need for trajectory planning approaches, considering the inner anatomy for optimizing the image quality. Based on the physics of the US image formation process, the introduced maximization of the expected image quality by acoustic transmission allows for the planning of single and multi view acquisitions not affected by shadowing artifacts, as we have shown. With this work, we aim at providing the basis for truly autonomous US acquisitions of a variety of anatomies, that could not be imaged previously due to restricted acoustic windows. Beyond that, the proposed method could also be used to train US technicians, providing trainees with feedback on the transducer positioning.

While our method provides optimal acquisition plans, their execution requires an accurate patient registration, as acoustic windows can be of limited size. For the example of cardiac US, a planning in the intercostal space needs to be precise w.r.t. the patient registration, as otherwise images would be distorted significantly by the ribs. This becomes even more

important, as respiratory motion could potentially require a continuous update of the patient registration in view of imaging in-vivo.

With this basis, our future work will include a more precise patient registration, possibly adding image based registration refinement as proposed in [40] and incorporating methods to detect degraded image quality during the acquisition [52]. This would allow the system to exclude affected images from the 3D compounding as well as the extension of our method for the planning of longer US trajectories with varying base points. Excluding degraded images based on their actual quality can reduce the impact of changes which inherently could not be accounted for in the planning, such as the presence of bowel gas or lesions resulting from the current treatment. Those changes could cause significant artifacts and attenuation.

The observed computation times do not allow for an interactive execution, but since the trajectories are meant to be obtained pre-operatively, this is not critical. Nevertheless, we plan to reduce the computational cost in the future.

Finally, our future work includes the generalization of the proposed method to use other tomographic modalities such as MRI as basis for the transmission estimation [97] and ultimately develop it towards an atlas-based approach as in [113]. We also aim at integrating further factors into the optimization, such as the volumetric coverage and anatomical context provided by the trajectory. To this end, a proof-of-concept study involving human acquisitions will be necessary to show the validity of the approach in a clinical setting, where CT data is a prerequisite for a clinical trial.

Part III

Conclusions and Outlook

6 Conclusions and Outlook

In this dissertation, we presented an overview of the field of medical ultrasound imaging and, based on identified challenges, presented novel approaches to further the field.

We introduced an open-source software framework for US processing aimed at making fundamental US imaging research more accessible. The platform, called SUPRA, is based on software rather than hardware, which makes it easier to improve and customize, and it is designed to be flexible and modular so that specialized solutions can be used in place of baseline algorithms. SUPRA is designed to be used in real-time, and it has been evaluated to show that it produces image quality comparable to a clinical pipeline.

Following that, we presented a method for the generation of synthetic training data for de-speckling filters. We also proposed a novel method for deep learning based speckle removal using data generated in this fashion, including an augmented loss to steer the resulting image appearance. The loss function makes use of the precise knowledge of the presence of certain geometries in the image. In experiments with simulated and real data, we demonstrated the effectiveness of the approach. The combination of data generation and model training has the potential to significantly decrease the cost of training de-speckling filters, as well as open up new possibilities for image filtering applications. While in this work we only employed the extended loss for enhancement of object interfaces, it is easy to imagine how this principle could be extended, *e.g.* to include blood-clutter reduction in the same network.

Finally, we presented a method to improve the trajectory planning for robotic US systems, formalizing the image quality considerations human operators already take into account. In estimating the influence of probe placement on the expected image quality, we showed improved target visibility when usable acoustic windows are defined by the presence of bones. Based on real patient data we performed 20 virtual planning scenarios and 5 real robotic acquisitions on a realistic tissue-mimicking phantom. Compared to a naive planning approach the investigated method showed a higher acoustic window quality within the sweeps while still covering the target structures, suggesting overall improved image quality. The phantom experiments showed the improvement in image quality directly, as the shadowing from bones was eliminated. With our approach that utilizes ultrasound physics to inform imaging and planning thereof, we have demonstrated the potential of the inclusion of this information in approaches to automate acquisitions and provide a pathway for improved imaging and better diagnostics.

We have investigated a number of approaches concerning US imaging, from very low-level, concerning the image formation, to processing of the so generated images to achieve better quality, and finally an application of the principles underlying the imaging process which promises to reduce operator dependence in US imaging.

Part IV

Appendix

A

List of Authored and Co-authored Publications

2019

[60] Fernanda Langsch, Salvatore Virga, Javier Esteban, **Rüdiger Göbl**, and Nassir Navab. “Robotic ultrasound for catheter navigation in endovascular procedures”. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 5404–5410.

[84] Magdalini Paschali, Walter Simson, Abhijit Guha Roy, **Rüdiger Göbl**, Christian Wachinger, and Nassir Navab. “Manifold exploring data augmentation with geometric transformations for increased performance and robustness”. *International Conference on Information Processing in Medical Imaging*. Springer. 2019, pp. 517–529.

[31] Marco Esposito, Christoph Hennersperger, **Rüdiger Göbl**, Laurent Demaret, Martin Storath, Nassir Navab, Maximilian Baust, Andreas Weinmann. “Total variation regularization of pose signals with an application to 3D freehand ultrasound”. *IEEE Transactions on Medical Imaging* 38.10(2019), pp. 2245–2258.

[89] Julia Rackerseder, Maximilian Baust, **Rüdiger Göbl**, Nassir Navab, and Christoph Hennersperger. “Landmark-Free Initialization of Multi-Modal Image Registration”. *Bildverarbeitung für die Medizin*. Springer, 2019, pp. 308–308.

2018

[88] Julia Rackerseder, Maximilian Baust, **Rüdiger Göbl**, Nassir Navab, and Christoph Hennersperger. “Initialize globally before acting locally: Enabling landmark-free 3D US to MRI registration”. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 827–835.

[32] **Rüdiger Göbl**, Nassir Navab, and Christoph Hennersperger. “SUPRA: open-source software-defined ultrasound processing for real-time applications”. *International Journal of Computer Assisted Radiology and Surgery (Mar. 2018)* 13, 759–767 (2018).

[26] Alper Denasi, Fouzia Khan, Klaas Jelmer Boskma, Mert Kaya, Christoph Hennersperger, **Rüdiger Göbl**, Maria Tirindelli, Nassir Navab, and Sarthak Misra. "An observer-based fusion method using multicore optical shape sensors and ultrasound images for magnetically-actuated catheters". *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 2018, pp. 50–57

[114] Salvatore Virga, **Rüdiger Göbl**, Maximilian Baust, Nassir Navab, and Christoph Hennersperger. "Use the force: Deformation correction in robotic 3D ultrasound". *International Journal of Computer Assisted Radiology and Surgery* 13.5 (2018), pp. 619–627.

2017

[33] **Rüdiger Göbl**, Salvatore Virga, Julia Rackerseder, Benjamin Frisch, Nassir Navab, and Christoph Hennersperger. "Acoustic window planning for ultrasound acquisition". *International Journal of Computer Assisted Radiology and Surgery* 12, 993–1001 (2017).

B Abstracts of Publications not Discussed in this Thesis

Robotic Ultrasound for Catheter Navigation in Endovascular Procedures

F. Langsch, S. Virga, J. Esteban, R. Göbl, N. Navab

Endovascular procedures require real time visual feedback on the location of inserted catheters. This is currently achieved using X-ray fluoroscopy, which causes exposure to radiation. This study describes an alternative method using a robotic ultrasound system for catheter tracking and navigation in endovascular interventions, focusing on endovascular aneurysm repair. This approach relies on the registration of pre-operative images to provide both a tracking trajectory and visual feedback of the real-time catheter position. The procedure was validated on healthy volunteers and on a phantom that included a realistic vessel structure, showing an average tracking error of the moving catheter tip of $1.78 \pm 1.02\,\mathrm{mm}$.

2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

Manifold Exploring Data Augmentation with Geometric Transformations for Increased Performance and Robustness

M. Paschali, W. Simson, A. G. Roy, R. Göbl, C. Wachinger, N. Navab

In this paper we propose a novel augmentation technique that improves not only the performance of deep neural networks on clean test data, but also significantly increases their robustness to random transformations, both affine and projective. Inspired by ManiFool, the augmentation is performed by a line-search manifold-exploration method that learns affine geometric transformations that lead to the misclassification on an image, while ensuring that it remains on the same manifold as the training data.

This augmentation method populates any training dataset with images that lie on the border of the manifolds between two-classes and maximizes the variance the network is exposed to during training. Our method was thoroughly evaluated on the challenging tasks of fine-grained skin lesion classification from limited data, and breast tumor classification of mammograms. Compared with traditional augmentation methods, and with images synthesized by Generative Adversarial Networks our method not only achieves state-of-the-art performance but also significantly improves the network's robustness.

International Conference on Information Processing in Medical Imaging, IPMI 2019

Total Variation Regularization of Pose Signals with an Application to 3D Freehand Ultrasound

M. Esposito, C. Hennersperger, R. Göbl, L. Demaret, M. Storath,
N. Navab, M. Baust, A. Weinmann

Three-dimensional freehand imaging techniques are gaining wider adoption due to their flexibility and cost efficiency. Typical examples for such a combination of a tracking system with an imaging device are freehand SPECT or freehand 3D ultrasound. However, the quality of the resulting image data is heavily dependent on the skill of the human operator and on the level of noise of the tracking data. The latter aspect can introduce blur or strong artifacts, which can significantly hamper the interpretation of image data. Unfortunately, the most commonly used tracking systems to date, i.e., optical and electromagnetic, present a trade-off between invading the surgeon's workspace (due to line-of-sight requirements) and higher levels of noise and sensitivity due to the interference of surrounding metallic objects. In this paper, we propose a novel approach for total variation regularization of data from tracking systems (which we term pose signals) based on a variational formulation in the manifold of Euclidean transformations. The performance of the proposed approach was evaluated using synthetic data as well as real ultrasound sweeps executed on both a Lego phantom and human anatomy, showing significant improvement in terms of tracking data quality and compounded ultrasound images. Source code can be found at https://github.com/IFL-CAMP/pose_regularization.

IEEE Transactions on Medical Imaging, vol. 38, no. 10

Landmark-Free Initialization of Multi-Modal Image Registration

J. Rackerseder, M. Baust, R. Göbl, N. Navab, C. Hennersperger

To achieve convergence, nonlinear deformable image registration tasks of partial-view 3D ultrasound and MRI, as often seen in US guided interventions or retrospective studies thereof, need to be initialized. In clinical practice corresponding 3D landmarks are selected in both images. Performing this depends on the geometrical understanding of the targeted anatomy and the modality-specific appearance and is thus prone to error. Therefore, we propose a novel landmark-free initialization procedure that is robust in terms of target area overlap (pixels where target area and US volume are superimposed before initialization) as well as image overlap (pixels where MRI and US are superimposed). The method only requires N low-resolution coarse segmentations as input, which in most cases can be obtained automatically or with minimal user interaction, such as few pre-labeled pixels. A euclidean distance transform is applied to these N label maps, creating a multi-class distance map for both images. This leads to a minimization problem, where these maps are registered by optimizing our proposed similarity measure via a modified gradient descent scheme, which prevents unstable behaviour. The proposed method was evaluated, showing a success rate of 100% for registration tasks

with initial target area overlap over 10%. It also converges for all cases with image overlap of 30% or more.

Bildverarbeitung für die Medizin 2019

Initialize Globally Before Acting Locally: Enabling Landmark-Free 3D US to MRI Registration

J. Rackerseder, M. Baust, R. Göbl, N. Navab, C. Hennersperger

Registration of partial-view 3D US volumes with MRI data is influenced by initialization. The standard of practice is using extrinsic or intrinsic landmarks, which can be very tedious to obtain. To overcome the limitations of registration initialization, we present a novel approach that is based on Euclidean distance maps derived from easily obtainable coarse segmentations. We evaluate our approach on a publicly available brain tumor dataset (RESECT) and show that it is robust regarding minimal to no overlap of target area and varying initial position. We demonstrate that our method provides initializations that greatly increase the capture range of state-of-the-art nonlinear registration algorithms.

International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2018

An Observer-Based Fusion Method Using Multicore Optical Shape Sensors and Ultrasound Images for Magnetically-Actuated Catheters

A. Denasi, F. Khan, K. J. Boskma, M. Kaya, C. Hennersperger, R. Göbl, M. Tirindelli, N. Navab, S. Misra

Minimally invasive surgery involves using flexible medical instruments such as endoscopes and catheters. Magnetically actuated catheters can provide improved steering precision over conventional catheters. However, besides the actuation method, an accurate tip position is required for precise control of the medical instruments. In this study, the tip position obtained from transverse 2D ultrasound images and multicore optical shape sensors are combined using a robust sensor fusion algorithm. The tip position is tracked in the ultrasound images using a template-based tracker and a convolutional neural network based tracker, respectively. Experimental results for a rhombus path are presented, where data obtained from both tracking sources are fused using Luenberger and Kalman state estimators. The mean and standard deviation of the Euclidean error for the Luenberger observer is $0.20 \pm 0.11\,[\mathrm{mm}]$ whereas for the Kalman filter it is $0.18 \pm 0.13\,[\mathrm{mm}]$, respectively.

2018 IEEE International Conference on Robotics and Automation (ICRA)

Use the Force: Deformation Correction in Robotic 3D Ultrasound

S. Virga, R. Göbl, M. Baust, N. Navab, C. Hennersperger

Purpose. Ultrasound acquisitions are typically affected by deformations due to the pressure applied onto the contact surface. While a certain amount of pressure is necessary to ensure good acoustic coupling and visibility of the anatomy under examination, the caused deformations hinder accurate localization and geometric analysis of anatomical structures. These complications have even greater impact in case of 3D ultrasound scans as they limit the correct reconstruction of acquired volumes. *Methods.* In this work, we propose a method to estimate and correct the induced deformation based solely on the tracked ultrasound images and information about the applied force. This is achieved by modeling estimated displacement fields of individual image sequences using the measured force information. By representing the computed displacement fields using a graph-based approach, we are able to recover a deformation-less 3D volume. *Results.* Validation is performed on 30 in vivo human datasets acquired using a robotic ultrasound framework. Compared to ground truth, the presented deformation correction shows errors of $3.39 \pm 1.86\,\mathrm{mm}$ for an applied force of $5\,\mathrm{N}$ at a penetration depth of $55\,\mathrm{mm}$. *Conclusion.* The proposed technique allows for the correction of deformations induced by the transducer pressure in entire 3D ultrasound volumes. Our technique does not require biomechanical models, patient-specific assumptions or information about the tissue properties; it can be employed based on the information from readily available robotic ultrasound platforms.

International Journal of Computer Assisted Radiology and Surgery 13.5 (2018)

Bibliography

[1] K. Ando, R. Nagaoka, and H. Hasegawa. “Speckle reduction of medical ultrasound images using deep learning with fully convolutional network”. In: *Japanese Journal of Applied Physics* 59.SK (2020), SKKE06 (cit. on p. 48).

[2] J. P. Asen, J. I. Buskenes, C.-I. C. Nilsen, A. Austeng, and S. Holm. “Implementing capon beamforming on a GPU for real-time cardiac ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.1 (Jan. 2014), pp. 76–85 (cit. on p. 22).

[3] B. M. Asl and A. M. Deylami. “A low complexity minimum variance beamformer for ultrasound imaging using dominant mode rejection”. In: *Ultrasonics* 85 (2018), pp. 49–60 (cit. on pp. 20, 23).

[4] B. M. Asl and A. Mahloojifar. “Eigenspace-based minimum variance beamforming applied to medical ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* (2010) (cit. on pp. 10, 23, 27).

[5] B. M. Asl and A. Mahloojifar. “Minimum variance beamforming combined with adaptive coherence weighting applied to medical ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 56.9 (Sept. 2009), pp. 1923–1931 (cit. on p. 23).

[6] M. A. Averkiou. “Tissue harmonic imaging”. In: *2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium*. Vol. 2. Oct. 2000, pp. 1563–1572 (cit. on p. 40).

[7] M. Bae, S. B. Park, and S. J. Kwon. “Fast minimum variance beamforming based on Legendre polynomials”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 63.9 (Sept. 2016), pp. 1422–1431 (cit. on p. 23).

[8] N. Bilaniuk and G. S. K. Wong. “Speed of sound in pure water as a function of temperature”. In: *The Journal of the Acoustical Society of America* 93.3 (1993), pp. 1609–1612 (cit. on p. 41).

[9] C. Bluemel, G. Safak, A. Cramer, et al. “Fusion of freehand SPECT and ultrasound: First experience in preoperative localization of sentinel lymph nodes”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 43.13 (Dec. 2016), pp. 2304–2312 (cit. on p. 36).

[10] E. Boni, L. Bassi, A. Dallai, et al. “ULA-OP 256: A 256-channel open scanner for development and real-time implementation of new ultrasound methods”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 63.10 (2016), pp. 1488–1495 (cit. on p. 37).

[11] N. Bottenus, B. C. Byram, J. J. Dahl, and G. E. Trahey. “Synthetic aperture focusing for short-lag spatial coherence imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 60.9 (Sept. 2013), pp. 1816–1826 (cit. on pp. 10, 15).

[12] A. Buades, B. Coll, and J.-M. Morel. “A non-local algorithm for image denoising”. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. Vol. 2. IEEE. 2005, pp. 60–65 (cit. on pp. 48, 56).

[13] J. Camacho, M. Parrilla, and C. Fritsch. “Phase Coherence Imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 56.5 (May 2009), pp. 958–974 (cit. on pp. 10, 13, 27).

[14] S. Cammarasana, P. Nicolardi, and G. Patanè. “A Universal Deep Learning Framework for Real-Time Denoising of Ultrasound Images”. In: *arXiv preprint arXiv:2101.09122* (2021) (cit. on pp. 48, 49).

[15] J. Capon. “High-resolution frequency-wavenumber spectrum analysis”. In: *Proceedings of the IEEE* 57.8 (1969), pp. 1408–1418 (cit. on p. 17).

[16] T. Chernyakova, D. Cohen, M. Shoham, and Y. C. Eldar. “iMAP beamforming for high-quality high frame rate imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 66.12 (2019), pp. 1830–1844 (cit. on pp. 10, 25, 27).

[17] C. C. P. Cheung, C. H. Alfred, N. Salimi, et al. “Multi-channel pre-beamformed data acquisition system for research on advanced ultrasound imaging methods”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 59.2 (2012), pp. 243–253 (cit. on p. 37).

[18] K. Christensen-Jeffries, R. J. Browning, M.-X. Tang, C. Dunsby, and R. J. Eckersley. “In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles”. In: *IEEE transactions on medical imaging* 34.2 (2015), pp. 433–440 (cit. on p. 63).

[19] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot. “Nonlocal means-based speckle filtering for ultrasound images”. In: *IEEE Transactions on Image Processing* 18.10 (2009), pp. 2221–2229 (cit. on pp. 48, 56).

[20] J. J. Dahl, D. Hyun, M. Lediju, and G. E. Trahey. “Lesion detectability in diagnostic ultrasound with short-lag spatial coherence imaging”. In: *Ultrasonic Imaging* 33.2 (Apr. 2011), pp. 119–133 (cit. on p. 15).

[21] J. J. Dahl, M. Jakovljevic, G. F. Pinton, and G. E. Trahey. “Harmonic spatial coherence imaging: an ultrasonic imaging method based on backscatter coherence”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 59.4 (Apr. 2012), pp. 648–659 (cit. on p. 15).

[22] J. C. Dai, B. Dunmire, K. M. Sternberg, et al. “Retrospective comparison of measured stone size and posterior acoustic shadow width in clinical ultrasound images”. In: *World journal of urology* 36.5 (2018), pp. 727–732 (cit. on p. 49).

[23] J. Dang, B. Frisch, P. Lasaygues, et al. “Development of an anthropomorphic breast phantom for combined PET, b-mode ultrasound and elastographic imaging”. In: *IEEE Transactions on Nuclear Science* 58.3 (2011), pp. 660–667 (cit. on p. 71).

[24] D. E. Dausch, K. H. Gilchrist, J. B. Carlson, S. D. Hall, J. B. Castellucci, and O. T. von Ramm. “In vivo real-time 3-D intracardiac echo using PMUT arrays”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.10 (2014), pp. 1754–1764 (cit. on p. 4).

[25] L. Demi. “Practical guide to ultrasound beam forming: Beam pattern and image reconstruction analysis”. In: *Applied Sciences* 8.9 (Sept. 2018), p. 1544 (cit. on p. 9).

[26] A. Denasi, F. Khan, K. J. Boskma, et al. “An observer-based fusion method using multicore optical shape sensors and ultrasound images for magnetically-actuated catheters”. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2018, pp. 50–57 (cit. on p. 82).

[27] A. M. Deylami and B. M. Asl. “A fast and robust beamspace adaptive beamformer for medical ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 64.6 (June 2017), pp. 947–958 (cit. on pp. 10, 22, 27).

[28] A. M. Deylami and B. M. Asl. “Iterative minimum variance beamformer with low complexity for medical ultrasound imaging”. In: *Ultrasound in Medicine & Biology* 44.8 (Aug. 2018), pp. 1882–1890 (cit. on pp. 10, 20).

[29] L. Du, T. Yardibi, J. Li, and P. Stoica. “Review of user parameter-free robust adaptive beamforming algorithms”. In: *Digital Signal Processing* 19.4 (July 2009), pp. 567–582 (cit. on p. 18).

[30] B. Englmair, C. Leitner, and C. Baumgartner. “Improved tracking of muscle tendon junctions in ultrasound images using speckle reduction”. In: *dHealth*. Vol. 271. Studies in Health Technology and Informatics. June 2020, pp. 1–8 (cit. on p. 47).

[31] M. Esposito, C. Hennersperger, R. Göbl, et al. “Total variation regularization of pose signals with an application to 3D freehand ultrasound”. In: *IEEE transactions on medical imaging* 38.10 (2019), pp. 2245–2258 (cit. on p. 81).

[32] R. Göbl, N. Navab, and C. Hennersperger. “SUPRA: open-source software-defined ultrasound processing for real-time applications”. In: *International Journal of Computer Assisted Radiology and Surgery* (Mar. 2018) (cit. on pp. 35, 53, 81).

[33] R. Göbl, S. Virga, J. Rackerseder, B. Frisch, N. Navab, and C. Hennersperger. “Acoustic window planning for ultrasound acquisition”. In: *International Journal of Computer Assisted Radiology and Surgery* 12.6 (2017), pp. 993–1001 (cit. on pp. 63, 82).

[34] C. Graumann, B. Fuerst, C. Hennersperger, F. Bork, and N. Navab. “Robotic ultrasound trajectory planning for volume of interest coverage”. In: *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE. 2016, pp. 736–741 (cit. on pp. 64, 68).

[35] W. Hackenberger, P. Rehrig, T. Ritter, and T. Shrout. “Advanced piezoelectric materials for medical ultrasound transducers”. In: *2001 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No.01CH37263)*. Vol. 2. 2001, 1101–1104 vol.2 (cit. on p. 4).

[36] H. Hasegawa. “Advances in ultrasonography: Image formation and quality assessment”. In: *J Med Ultrasonics* 48.4 (2021), pp. 377–389 (cit. on p. 9).

[37] H. Hasegawa. “Enhancing effect of phase coherence factor for improvement of spatial resolution in ultrasonic imaging”. In: *J Med Ultrasonics* 43.1 (Jan. 2016), pp. 19–27 (cit. on p. 14).

[38] H. Hasegawa and H. Kanai. “Effect of subaperture beamforming on phase coherence imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.11 (Nov. 2014), pp. 1779–1790 (cit. on pp. 10, 14).

[39] T. Heimann, B. Van Ginneken, M. A. Styner, et al. “Comparison and evaluation of methods for liver segmentation from CT datasets”. In: *IEEE transactions on medical imaging* 28.8 (2009), pp. 1251–1265 (cit. on p. 69).

[40] C. Hennersperger, B. Fuerst, S. Virga, et al. “Towards MRI-based autonomous robotic US acquisitions: A first feasibility study”. In: *IEEE Transactions on Medical Imaging* 36.2 (2016), pp. 538–548 (cit. on pp. 63, 64, 68, 70, 72, 74).

[41] C. Hennersperger, D. Mateus, M. Baust, and N. Navab. “A quadratic energy minimization framework for signal loss estimation from arbitrarily sampled ultrasound data”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2014, pp. 373–380 (cit. on p. 64).

[42] K. Hollman, K. Rigby, and M. O’Donnell. “Coherence factor of speckle from a multi-row probe”. In: *1999 IEEE Ultrasonics Symposium. Proceedings. International Symposium (Cat. No.99CH37027)*. Vol. 2. 1999, 1257–1260 vol.2 (cit. on pp. 10, 12, 27).

[43] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl. “Beamforming and speckle reduction using neural networks”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 66.5 (May 2019), pp. 898–910 (cit. on pp. 10, 26, 48).

[44] D. Hyun, A. L. C. Crowley, and J. J. Dahl. “Efficient strategies for estimating the spatial coherence of backscatter”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 64.3 (Mar. 2017), pp. 500–513 (cit. on pp. 10, 16).

[45] D. Hyun, G. E. Trahey, M. Jakovljevic, and J. J. Dahl. “Short-lag spatial coherence imaging on matrix arrays, Part 1: Beamforming methods and simulation studies”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.7 (July 2014), pp. 1101–1112 (cit. on pp. 10, 15).

[46] M. Jakovljevic, B. C. Byram, D. Hyun, J. J. Dahl, and G. E. Trahey. “Short-lag spatial coherence imaging on matrix arrays, Part II: Phantom and in vivo experiments”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.7 (July 2014), pp. 1113–1122 (cit. on p. 16).

[47] M. Jakovljevic, G. E. Trahey, R. C. Nelson, and J. J. Dahl. “In vivo application of short-Lag spatial coherence imaging in human liver”. In: *Ultrasound in Medicine & Biology* 39.3 (Mar. 2013), pp. 534–542 (cit. on p. 15).

[48] J. A. Jensen. “Field: A program for simulating ultrasound systems”. In: *Proc. of the 6th Nordic Conf. on Human-Computer Interaction*. Medical & Biological Engineering & Computing. Springer, 1996, pp. 351–353 (cit. on p. 51).

[49] J. A. Jensen and N. B. Svendsen. “Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 39.2 (1992), pp. 262–267 (cit. on p. 51).

[50] M. K. Jeong and S. J. Kwon. “Estimation of side lobes in ultrasound imaging systems”. In: *Biomedical Engineering Letters* 5.3 (Sept. 2015), pp. 229–239 (cit. on p. 41).

[51] F. Q. Jin and M. L. Palmeri. “Does ultrasonic data format matter for deep neural networks?” In: *2020 IEEE International Ultrasonics Symposium (IUS)*. 2020 (cit. on p. 48).

[52] A. Karamalis, W. Wein, T. Klein, and N. Navab. “Ultrasound confidence maps using random walks”. In: *Medical image analysis* 16.6 (2012), pp. 1101–1112 (cit. on pp. 64, 74).

[53] H. G. Khor, G. Ning, X. Zhang, and H. Liao. “Ultrasound speckle reduction using wavelet-based generative adversarial network”. In: *IEEE Journal of Biomedical and Health Informatics* 26.7 (2022), pp. 3080–3091 (cit. on pp. 48, 61).

[54] K. Kim, S. Park, J. Kim, S.-B. Park, and M. Bae. “A fast minimum variance beamforming method using principal component analysis”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.6 (June 2014), pp. 930–945 (cit. on pp. 10, 22).

[55] T. Klein and W. M. Wells III. “RF ultrasound distribution-based confidence maps”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing. 2015, pp. 595–602 (cit. on p. 64).

[56] S. Kraft, S. Conjeti, P. B. Noël, S. Carlier, N. Navab, and A. Katouzian. “Full-wave intravascular ultrasound simulation from histology”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2014, pp. 627–634 (cit. on p. 64).

[57] K. Krissian, C. F. Westin, R. Kikinis, and K. G. Vosburgh. “Oriented speckle reducing anisotropic diffusion”. In: *IEEE Transactions on Image Processing* 16.5 (2007), pp. 1412–1424 (cit. on p. 40).

[58] A. Krull, T.-O. Buchholz, and F. Jug. “Noise2void-learning denoising from single noisy images”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2129–2137 (cit. on p. 49).

[59] A. Kurjak and S. Kupesic. “Clinical application of 3D sonography”. In: Taylor & Francis, 2000. Chap. Three-dimensional ultrasonography in hepatogastroenterology (cit. on p. 63).

[60] F. Langsch, S. Virga, J. Esteban, R. Göbl, and N. Navab. “Robotic ultrasound for catheter navigation in endovascular procedures”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 5404–5410 (cit. on p. 81).

[61] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger. “PLUS: Open-source toolkit for ultrasound-guided intervention systems”. In: *IEEE Transactions on Biomedical Engineering* 61.10 (2014), pp. 2527–2537 (cit. on p. 36).

[62] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl. “Short-lag spatial coherence of backscattered echoes: Imaging characteristics”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 58.7 (2011), pp. 1377–1388 (cit. on pp. 10, 14, 27).

[63] J. Lehtinen, J. Munkberg, J. Hasselgren, et al. “Noise2Noise: Learning image restoration without clean data”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by J. Dy and A. Krause. Vol. 80. Proceedings of Machine Learning Research. July 2018, pp. 2965–2974 (cit. on pp. 49, 50).

[64] M.-L. Li, W.-J. Guan, and P.-C. Li. “Improved synthetic aperture focusing technique with applications in high-frequency ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 51.1 (Jan. 2004), pp. 63–70 (cit. on p. 13).

[65] P.-C. Li and M.-L. Li. “Adaptive imaging using the generalized coherence factor”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 50.2 (Feb. 2003), pp. 128–141 (cit. on pp. 10, 13, 27).

[66] Y. L. Li and J. J. Dahl. “Angular coherence in ultrasound imaging: Theory and applications”. In: *The Journal of the Acoustical Society of America* 141.3 (Mar. 2017), pp. 1582–1594 (cit. on pp. 10, 15).

[67] H.-l. Liu, Z.-h. Zhang, and D.-q. Liu. “Adaptive diagonal loaded minimum variance beamforming applied to medical ultrasound imaging”. In: *J. Cent. South Univ.* 22.5 (2015), pp. 1826–1832 (cit. on p. 20).

[68] J.-Y. Lu, H. Zou, and J. F. Greenleaf. “Biomedical ultrasound beam forming”. In: *Ultrasound in Medicine & Biology* 20.5 (Jan. 1994), pp. 403–428 (cit. on p. 9).

[69] A. C. Luchies and B. C. Byram. “Training improvements for ultrasound beamforming with deep neural networks”. In: *Physics in Medicine & Biology* 64.4 (Feb. 2019), p. 045018 (cit. on p. 26).

[70] A. C. Luchies and B. C. Byram. “Deep neural networks for ultrasound beamforming”. In: *IEEE Trans. Med. Imag.* 37.9 (Sept. 2018), pp. 2010–2021 (cit. on pp. 10, 26).

[71] B. Luijten, R. Cohen, F. J. de Bruijn, et al. “Adaptive ultrasound beamforming using deep learning”. In: *IEEE Trans. Med. Imag.* 39.12 (2020), pp. 3967–3978 (cit. on pp. 10, 27).

[72] M. Makuuchi, H. Hasegawa, S. Yamazaki, K. Takayasu, and N. Moriyama. “The use of operative ultrasound as an aid to liver resection in patients with hepatocellular carcinoma”. In: *World journal of surgery* 11.5 (1987), pp. 615–621 (cit. on p. 63).

[73] G. Matrone, A. Ramalli, A. S. Savoia, P. Tortoli, and G. Magenes. “High frame-rate, high resolution ultrasound imaging with multi-line transmission and filtered-delay multiply and sum beamforming”. In: *IEEE Trans. Med. Imag.* 36.2 (Feb. 2017), pp. 478–486 (cit. on pp. 10, 17).

[74] G. Matrone, A. Ramalli, P. Tortoli, and G. Magenes. “Experimental evaluation of ultrasound higher-order harmonic imaging with Filtered-Delay Multiply And Sum (F-DMAS) non-linear beamforming”. In: *Ultrasonics* 86 (May 2018), pp. 59–68 (cit. on pp. 10, 16).

[75] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes. “The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging”. In: *IEEE Trans. Med. Imag.* 34.4 (Apr. 2015), pp. 940–949 (cit. on pp. 10, 16, 27).

[76] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes. “Depth-of-field enhancement in Filtered-Delay Multiply and Sum beamformed images using Synthetic Aperture Focusing”. In: *Ultrasonics* 75 (Mar. 2017), pp. 216–225 (cit. on pp. 10, 17).

[77] M. McCormick. “An open source, fast ultrasound B-Mode software implementation for commodity hardware”. In: *The Insight Journal* (January-June 2010) (cit. on p. 36).

[78] S. Mehdizadeh, A. Austeng, T. F. Johansen, and S. Holm. “Minimum variance beamforming applied to ultrasound imaging with a partially shaded aperture”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 59.4 (Apr. 2012), pp. 683–693 (cit. on p. 19).

[79] S. V. Mohd Sagheer and S. N. George. “A review on medical image denoising algorithms”. In: *Biomedical Signal Processing and Control* 61 (2020), p. 102036 (cit. on p. 48).

[80] W. K. Moon, C.-M. Lo, J. M. Chang, C.-S. Huang, J.-H. Chen, and R.-F. Chang. “Computer-aided classification of breast masses using speckle features of automated breast ultrasound images”. In: *Medical Physics* 39.10 (2012), pp. 6465–6473 (cit. on p. 47).

[81] C.-I. Nilsen and I. Hafizovic. “Beamspace adaptive beamforming for ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 56.10 (Oct. 2009), pp. 2187–2197 (cit. on pp. 10, 21, 22).

[82] O. Oralkan, A. Ergun, J. Johnson, et al. “Capacitive micromachined ultrasonic transducers: next-generation arrays for acoustic imaging?” In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 49.11 (2002), pp. 1596–1610 (cit. on p. 4).

[83] J. Park, S. M. Wi, and J. S. Lee. “Computationally efficient adaptive beamformer for ultrasound imaging based on QR decomposition”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* (2016) (cit. on p. 23).

[84] M. Paschali, W. Simson, A. G. Roy, R. Göbl, C. Wachinger, and N. Navab. “Manifold exploring data augmentation with geometric transformations for increased performance and robustness”. In: *International Conference on Information Processing in Medical Imaging*. Springer. 2019, pp. 517–529 (cit. on p. 81).

[85] M. Polichetti, F. Varray, J.-C. Béra, C. Cachard, and B. Nicolas. “A nonlinear beamformer based on p-th root compression—Application to plane wave ultrasound imaging”. In: *Applied Sciences* 8.4 (Apr. 2018), p. 599 (cit. on pp. 25, 27).

[86] F. Prieur, O. M. H. Rindal, and A. Austeng. “Signal coherence and image amplitude With the filtered delay multiply and sum beamformer”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 65.7 (July 2018), pp. 1133–1140 (cit. on p. 17).

[87] Y. Qi, Y. Wang, J. Yu, and Y. Guo. “Short-lag spatial coherence imaging using minimum variance beamforming on dual apertures”. In: *BioMedical Engineering OnLine* 18.1 (Dec. 2019), p. 48 (cit. on p. 15).

[88] J. Rackerseder, M. Baust, R. Göbl, N. Navab, and C. Hennersperger. “Initialize globally before acting locally: enabling landmark-free 3D US to MRI registration”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 827–835 (cit. on p. 81).

[89] J. Rackerseder, M. Baust, R. Göbl, N. Navab, and C. Hennersperger. “Landmark-free initialization of multi-modal image registration”. In: *Bildverarbeitung für die Medizin 2019*. Springer, 2019, pp. 308–308 (cit. on p. 81).

[90] R. Ribeiro and J. Sanches. “Fatty liver characterization and classification by ultrasound”. In: *Pattern Recognition and Image Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 354–361 (cit. on p. 47).

[91] M. Riva, C. Hennersperger, F. Milletari, et al. “3D intra-operative ultrasound and MR image guidance: pursuing an ultrasound-based management of brainshift to enhance neuronavigation”. In: *International Journal of Computer Assisted Radiology and Surgery* 12.10 (Oct. 2017), pp. 1711–1725 (cit. on p. 36).

[92] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, et al. “The UltraSound ToolBox”. In: *2017 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2017, pp. 1–4 (cit. on pp. 27, 36).

[93] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’15)*. Springer. 2015, pp. 234–241 (cit. on p. 50).

[94] G. Sakas and S. Walter. “Extracting surfaces from fuzzy 3D-ultrasound data”. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995, pp. 465–474 (cit. on p. 47).

[95] S. M. Sakhaei and S. E. Shamsian. “Twofold minimum variance beamforming for enhanced ultrasound imaging”. In: *J Med Ultrasonics* 45.1 (Jan. 2018), pp. 17–24 (cit. on p. 23).

[96] A. Salari and B. M. Asl. “User parameter-free minimum variance beamformer in medical ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 68.7 (2021), pp. 2397–2406 (cit. on pp. 10, 20).

[97] M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, and W. Wein. “Patient-specific 3D ultrasound simulation based on convolutional ray-tracing and appearance optimization”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 510–518 (cit. on pp. 64, 74).

[98] U. Schneider, E. Pedroni, and A. Lomax. “The calibration of CT Hounsfield units for radiotherapy treatment planning”. In: *Physics in medicine and biology* 41.1 (1996), p. 111 (cit. on pp. 66, 70).

[99] C. Schulte zu Berge. “Real-time processing for advanced ultrasound visualization”. PhD thesis. Technische Universität München, 2016 (cit. on p. 47).

[100] S. E. Shamsian and S. M. Sakhaei. “Fast adaptive beamforming through a cascade structure for ultrasound imaging”. In: *J Med Ultrasonics* 46.3 (July 2019), pp. 287–296 (cit. on p. 23).

[101] D. P. Shattuck, M. D. Weinshenker, S. W. Smith, and O. T. von Ramm. “Explososcan: a parallel processing technique for high speed ultrasound imaging with linear phased arrays.” In: *The Journal of the Acoustical Society of America* 75.4 (1984), pp. 1273–1282 (cit. on pp. 36, 38).

[102] C.-C. Shen and P.-Y. Hsieh. “Ultrasound baseband delay-multiply-and-sum (BB-DMAS) nonlinear beamforming”. In: *Ultrasonics* 96 (July 2019), pp. 165–174 (cit. on pp. 10, 17).

[103] T. Su, D. Li, and S. Zhang. “An efficient subarray average delay multiply and sum beamformer algorithm in ultrasound imaging”. In: *Ultrasonics* 84 (Mar. 2018), pp. 411–420 (cit. on p. 17).

[104] T. Su, S. Zhang, D. Li, and D. Yao. “Combined sign coherent factor and delay multiply and sum beamformer for plane wave imaging”. In: *Acoustical Physics* 64.3 (May 2018), pp. 379–386 (cit. on p. 24).

[105] J.-F. Synnevåg, A. Austeng, and S. Holm. “A low-complexity data-dependent beamformer”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 58.2 (Feb. 2011), pp. 281–289 (cit. on pp. 10, 24).

[106] J.-F. Synnevåg, A. Austeng, and S. Holm. “Benefits of minimum-variance beamforming in medical ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 56.9 (Sept. 2009), pp. 1868–1879 (cit. on pp. 10, 19, 27).

[107] J.-F. Synnevåg, A. Austeng, and S. Holm. “Adaptive beamforming applied to medical ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 54.8 (Aug. 2007), pp. 1606–1613 (cit. on pp. 10, 18).

[108] M. Tanter and M. Fink. “Ultrafast imaging in biomedical ultrasound”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.1 (Jan. 2014), pp. 102–119 (cit. on p. 40).

[109] H. Tchelepi, P. W. Ralls, R. Radin, and E. Grant. “Sonography of diffuse liver disease”. In: *Journal of Ultrasound in Medicine* 21.9 (2002), pp. 1023–1032 (cit. on p. 63).

[110] K. Thomenius. “Evolution of ultrasound beamformers”. In: *1996 IEEE Ultrasonics Symposium. Proceedings*. Vol. 2. 1996, 1615–1622 vol.2 (cit. on pp. 9, 11).

[111] C. Tomasi and R. Manduchi. “Bilateral filtering for gray and color images”. In: *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*. IEEE. 1998, pp. 839–846 (cit. on p. 55).

[112] F. Vignon and M. Burcher. “Capon beamforming in medical ultrasound imaging with focused beams”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 55.3 (Mar. 2008), pp. 619–628 (cit. on pp. 10, 19).

[113] S. Virga, O. Zettinig, M. Esposito, et al. “Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms”. In: *In. Proc. IROS 2016 – IEEE International Conference on Intelligent Robots and Systems* (2016) (cit. on pp. 64, 70, 74).

[114] S. Virga, R. Göbl, M. Baust, N. Navab, and C. Hennersperger. “Use the force: deformation correction in robotic 3D ultrasound”. In: *International journal of computer assisted radiology and surgery* 13.5 (2018), pp. 619–627 (cit. on p. 82).

[115] R. Wagner, S. Smith, J. Sandrik, and H. Lopez. “Statistics of speckle in ultrasound B-scans”. In: *IEEE Transactions on Sonics and Ultrasonics* 30.3 (1983), pp. 156–163 (cit. on p. 53).

[116] M. Walczak, M. Lewandowski, and N. Zolek. “A real-time streaming DAQ for Ultrasonix Research scanner”. In: *IEEE International Ultrasonics Symposium, IUS*. Sept. 2014, pp. 1257–1260 (cit. on p. 37).

[117] Y.-H. Wang and P.-C. Li. “SNR-dependent coherence-based adaptive imaging for high-frame-rate ultrasonic and photoacoustic imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.8 (Aug. 2014), pp. 1419–1432 (cit. on pp. 10, 13).

[118] S.-L. Wang and P.-C. Li. “Aperture-domain processing and its applications in ultrasound imaging: A review”. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 224.2 (Feb. 2010), pp. 143–154 (cit. on p. 9).

[119] S.-L. Wang, C.-H. Chang, H.-C. Yang, Y.-H. Chou, and P.-C. Li. “Performance evaluation of coherence-based adaptive imaging using clinical breast data”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 54.8 (Aug. 2007), pp. 1669–1679 (cit. on p. 13).

[120] S.-L. Wang and P.-C. Li. “MVDR-based coherence weighting for high-frame-rate adaptive imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 56.10 (Oct. 2009), pp. 2097–2110 (cit. on p. 24).

[121] Y. Wang, C. Zheng, H. Peng, and Q. Chen. “An adaptive beamforming method for ultrasound imaging based on the mean-to-standard-deviation factor”. In: *Ultrasonics* 90 (Nov. 2018), pp. 32–41 (cit. on p. 25).

[122] W. Wein, S. Brunke, A. Khamene, M. R. Callstrom, and N. Navab. “Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention”. In: *Medical image analysis* 12.5 (2008), pp. 577–585 (cit. on pp. 64, 66, 67).

[123] A. Wiacek, E. González, and M. A. L. Bell. “CohereNet: A deep learning architecture for ultrasound spatial correlation estimation and coherence-based beamforming”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 67.12 (2020), pp. 2574–2583 (cit. on pp. 10, 27).

[124] T. Wilson, J. Zagzebski, T. Varghese, Q. Chen, and M. Rao. “The ultrasonix 500RP: A commercial ultrasound research interface”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 53.10 (2006), pp. 1772–1781 (cit. on p. 37).

[125] M. Xu, X. Yang, M. Ding, and M. Yuchi. “Spatio-temporally smoothed coherence factor for ultrasound imaging”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 61.1 (Jan. 2014), pp. 182–190 (cit. on pp. 10, 14).

[126] T. Yang, W. Wang, G. Cheng, M. Wei, H. Xie, and F. L. Wang. “FDDL-Net: frequency domain decomposition learning for speckle reduction in ultrasound images”. In: *Multimedia Tools and Applications* (Aug. 2022) (cit. on pp. 48, 61).

[127] D. Yin, Z. Gu, Y. Zhang, et al. “Speckle noise reduction in coherent imaging based on deep learning without clean data”. In: *Optics and Lasers in Engineering* 133 (2020), p. 106151 (cit. on p. 49).

[128] Y. Yu and S. Acton. “Speckle reducing anisotropic diffusion”. In: *IEEE Transactions on Image Processing* 11.11 (2002), pp. 1260–1270 (cit. on pp. 48, 55).

[129] G. Zahnd, S. Salles, H. Liebgott, D. Vray, A. Sérusclat, and P. Moulin. “Real-time ultrasound-tagging to track the 2D motion of the common carotid artery wall in vivo”. In: *Medical Physics* 42.2 (2015), pp. 820–830 (cit. on p. 24).

[130] O. Zettinig, A. Shah, C. Hennersperger, et al. “Multimodal image-guided prostate fusion biopsy based on automatic deformable registration”. In: *International Journal of Computer Assisted Radiology and Surgery* 10.12 (Dec. 2015), pp. 1997–2007 (cit. on p. 36).

[131] J. Zhao, Y. Wang, J. Yu, W. Guo, S. Zhang, and S. Aliabadi. “Short-lag spatial coherence ultrasound imaging with adaptive synthetic transmit aperture focusing”. In: *Ultrasonic Imaging* 39.4 (July 2017), pp. 224–239 (cit. on pp. 10, 15).

[132] B. Zhuang, R. Rohling, and P. Abolmaesumi. “Accumulated angle factor-based beamforming to improve the visualization of spinal structures in ultrasound images”. In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 65.2 (Feb. 2018), pp. 210–222 (cit. on p. 24).

Online Resources

[133] P. Coupé. *Optimized Bayesian non-local means filter, Matlab package*. 2011. URL: https://web.archive.org/web/20201015011836/https://sites.google.com/site/pierrickcoupe/softwares/denoising-for-medical-imaging/speckle-reduction/obnlm-package (visited on Dec. 7, 2021) (cit. on p. 56).

[134] *Ultrasound Toolbox – MATLAB toolbox for processing ultrasonic signals*. URL: http://www.ultrasoundtoolbox.com/ (visited on Aug. 1, 2022) (cit. on pp. 27, 31, 32).

[135] Virginia Image and Video Analysis, School of Engineering and Applied Science, University of Virginia. *MATLAB and C implementation of speckle reducing anisotropic diffusion (SRAD)*. 2006. URL: https://web.archive.org/web/20171213062541/http://viva-lab.ece.virginia.edu/downloads.html (visited on Dec. 7, 2021) (cit. on p. 55).

List of Figures

List of Tables