Cardiac motion estimation in ultrasound images using a sparse representation and dictionary learning

Nora Leïla Ouzir

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *16/10/2018* **par :**

NORA OUZIR

Cardiac Motion Estimation In Ultrasound Images using a Sparse Representation and Dictionary Learning

---

JURY

ADRIAN BASARAB	Maître de Conférences	Directeur de Thèse
ISABELLE BERRY	Professeur des Universités	Examinateur
CHRISTINE FERNANDEZ	Professeur des Universités	Examinateur
GERSENDE FORT	Directrice de Recherche	Examinateur
DENIS KOUAME	Professeur des Universités	Examinateur
ESA OLLILA	Professeur Associé	Examinateur
GABRIEL PEYRE	Directeur de Recherche	Rapporteur
JEAN-YVES TOURNERET	Professeur des Universités	Directeur de Thèse
DIDIER VRAY	Professeur des Universités	Rapporteur

École doctorale et spécialité :
MITT : Signal, Image, Acoustique et Optimisation

Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (UMR CNRS 5505)

Directeur(s) de Thèse :
Adrian BASARAB et *Jean-Yves TOURNERET*

Rapporteurs :
Gabriel PEYRE et *Didier VRAY*

Abstract

Cardiovascular diseases have become a major healthcare issue. Improving the diagnosis and analysis of these diseases have thus become a primary concern in cardiology. The heart is a moving organ that undergoes complex deformations. Therefore, the quantification of cardiac motion from medical images, particularly ultrasound, is a key part of the techniques used for diagnosis in clinical practice. Thus, significant research efforts have been directed toward developing new cardiac motion estimation methods. These methods aim at improving the quality and accuracy of the estimated motions. However, they are still facing many challenges due to the complexity of cardiac motion and the quality of ultrasound images. Recently, learning-based techniques have received a growing interest in the field of image processing. More specifically, sparse representations and dictionary learning strategies have shown their efficiency in regularizing different ill-posed inverse problems. This thesis investigates the benefits that such sparsity and learning-based techniques can bring to cardiac motion estimation. Three main contributions are presented, investigating different aspects and challenges that arise in echocardiography.

Firstly, a method for cardiac motion estimation using a sparsity-based regularization is introduced. The motion estimation problem is formulated as an energy minimization, whose data fidelity term is built using the assumption that the images are corrupted by multiplicative Rayleigh noise. In addition to a classical spatial smoothness constraint, the proposed method exploits the sparse properties of the cardiac motion to regularize the solution via an appropriate dictionary learning step. Secondly, a fully robust optical flow method is proposed. The aim of this work is to take into account the limitations of ultrasound imaging and the violations of the regularization constraints. In this work, two regularization terms imposing spatial smoothness and sparsity of the motion field in an appropriate cardiac motion dictionary are also exploited. In order to ensure robustness to outliers, an iteratively re-weighted minimization strategy is proposed using weighting functions based on M-estimators. As a last contribution, we investigate a cardiac motion estimation method using a combination of sparse, spatial and temporal regularizations. The problem is formulated within a general optical flow framework. The proposed temporal regularization enforces smoothness of the motion trajectories between consecutive images. Furthermore, an iterative groupewise motion estimation allows us to incorporate the three regularization terms, while enabling the processing of the image sequence as a whole. Throughout this thesis, the proposed contributions are validated using synthetic and realistic simulated cardiac ultrasound images. These

datasets with available groundtruth are used to evaluate the accuracy of the proposed approaches and show their competitiveness with state-of-the-art algorithms. In order to demonstrate clinical feasibility, *in vivo* sequences of healthy and pathological subjects are considered for the first two methods. A preliminary investigation is conducted for the last contribution, *i.e.*, exploiting temporal smoothness, using simulated data.

Résumé

Les maladies cardiovasculaires sont de nos jours un problème de santé majeur. L'amélioration des méthodes liées au diagnostic de ces maladies représente donc un réel enjeu en cardiologie. Le cœur étant un organe en perpétuel mouvement, l'analyse du mouvement cardiaque est un élément clé pour le diagnostic. Par conséquent, les méthodes dédiées à l'estimation du mouvement cardiaque à partir d'images médicales, plus particulièrement en échocardiographie, font l'objet de nombreux travaux de recherches. Cependant, plusieurs difficultés liées à la complexité du mouvement du cœur ainsi qu'à la qualité des images échographiques restent à surmonter afin d'améliorer la qualité et la précision des estimations. Dans le domaine du traitement d'images, les méthodes basées sur l'apprentissage suscitent de plus en plus d'intérêt. Plus particulièrement, les représentations parcimonieuses et l'apprentissage de dictionnaires ont démontré leur efficacité pour la régularisation de divers problèmes inverses. Cette thèse a ainsi pour but d'explorer l'apport de ces méthodes, qui allient parcimonie et apprentissage, pour l'estimation du mouvement cardiaque. Trois principales contributions sont présentées, chacune traitant différents aspects et problématiques rencontrées dans le cadre de l'estimation du mouvement en échocardiographie.

Dans un premier temps, une méthode d'estimation du mouvement cardiaque se basant sur une régularisation parcimonieuse est proposée. Le problème d'estimation du mouvement est formulé dans le cadre d'une minimisation d'énergie, dont le terme d'attache aux données est construit avec l'hypothèse d'un bruit de Rayleigh multiplicatif. Une étape d'apprentissage de dictionnaire permet une régularisation exploitant les propriétés parcimonieuses du mouvement cardiaque, combinée à un terme classique de lissage spatial. Dans un second temps, une méthode robuste de flux optique est présentée. L'objectif de cette approche est de robustifier la méthode d'estimation développée au premier chapitre de manière à la rendre moins sensible aux éléments aberrants. Deux régularisations sont mises en œuvre, imposant d'une part un lissage spatial et de l'autre la parcimonie des champs de mouvements dans un dictionnaire approprié. Afin d'assurer la robustesse de la méthode vis-à-vis des anomalies, une stratégie de minimisation récursivement pondérée est proposée. Plus précisément, les fonctions employées pour cette pondération sont basées sur la théorie des M-estimateurs. Le dernier travail présenté dans cette thèse, explore une méthode d'estimation du mouvement cardiaque exploitant une régularisation parcimonieuse combinée à un lissage à la fois dans les domaines spatial et temporel. Le problème est formulé dans un cadre général d'estimation de flux

optique. La régularisation temporelle proposée impose des trajectoires de mouvement lisses entre images consécutives. De plus, une méthode itérative d'estimation permet d'incorporer les trois termes de régularisations, tout en rendant possible le traitement simultané d'un ensemble d'images. Dans cette thèse, les contributions proposées sont validées en employant des images synthétiques et des simulations réalistes d'images ultrasonores. Ces données avec vérité terrain permettent d'évaluer la précision des approches considérées, et de souligner leur compétitivité par rapport à des méthodes de l'état-de-l'art. Pour démontrer la faisabilité clinique, des images *in vivo* de patients sains ou atteints de pathologies sont également considérées pour les deux premières méthodes. Pour la dernière contribution de cette thèse, *i.e.*, exploitant un lissage temporel, une étude préliminaire est menée en utilisant des données de simulation.

Contents

Chapter 1

Introduction

Contents

1.1 Cardiac motion estimation

Cardiovascular diseases are the top cause of death globally. They were responsible for 17.7 million deaths in 2015, *i.e.,* up to 31% of the total number of deaths worldwide [Organization 2017]. According to [Organization 2017], coronary heart disease alone has caused 7.4 million of these deaths. For coronary heart disease, as for many other cardiac disorders, early detection has a key role in preventing aggravation and providing appropriate early treatments. It is therefore of critical importance to improve the techniques of cardiac function assessment, thus facilitating the diagnosis and management of these diseases.

The heart undergoes a cyclic motion that alternates between two phases: systole and diastole. During systole, the contraction of the myocardium propels the blood into the circulatory system. This stage is followed by a relaxation of the tissues called diastole, during which the ventricles are filled with blood. In a healthy heart, each of these stages is characterized by normal patterns and amounts of contractions and expansions of the tissues. Cardiovascular diseases, however, can result in an alteration of the mechanical function of the heart. In particular, changes in elasticity and contractility, as in the case of ischemia, can affect the global or regional motions of the myocardium, *i.e.*, the muscular tissues around the ventricles (see Fig. 1.1). In this context, automatic cardiac motion estimation as well as the associated strain measurements have been proved to be efficient tools for the diagnosis of cardiovascular diseases, *e.g.*, [D'hooge 2002, Sutherland 2004, Abraham 2007, Cottrell 2010, Shah 2012]. These techniques seek to quantitatively characterize the tissue motion and deformation, and thus, facilitate the diagnosis.

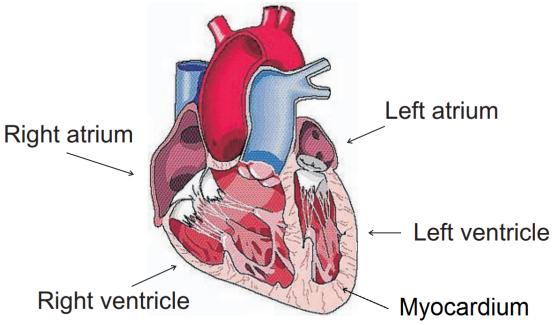
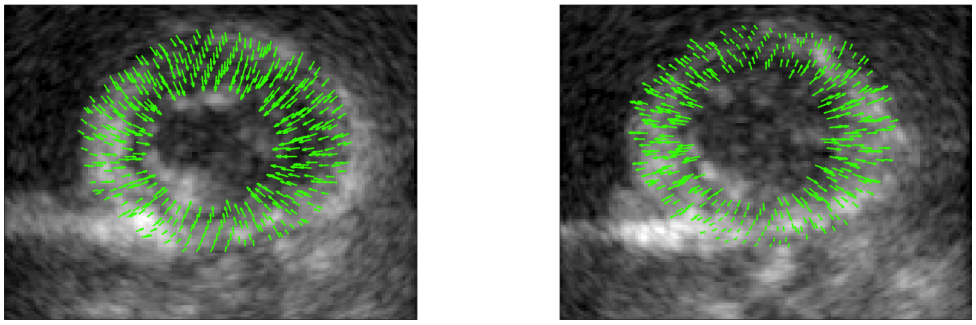


Figure 1.1: Illustration of the heart showing the left ventricle (LV) and the myocardium [Kindberg 2010].

There are a variety of methods used to evaluate the mechanical action of the heart. Invasive techniques, such as the implanting of radiopaque markers [Ingels 1975] or ultrasonic piezoelectric crystals [Villarreal 1988] are used to track marker positions during the cardiac cycle. Since these markers are physically moving with the tissue, the estimated deformations are close to the ground-truth motions of the heart. Non-invasive techniques rely principally on medical imaging. The acquired images of the heart pro-

vide essential information about the motions of the myocardial tissues. These images can be exploited either through direct visualization or using post-processing methods that extract qualitative and measurable features. As opposed to the invasive techniques (providing only sparse measurements), medical imaging allows dense motion fields to be computed, *i.e.*, a motion vector for each pixel (see an example in Fig. 1.2). Other features, such as the positions of the myocardial boundaries [Luo 2008], can also be extracted using post-processing techniques. In order to obtain dense motion fields (as shown in Fig. 1.2) cardiac motion estimation methods rely on regularization constraints that allow *a priori* knowledge about the motion, *e.g.*, spatial smoothness, to be incorporated. The resulting motions can then be used to compute associated indicators, *e.g.*, strain or torsion [Mirea 2016]. Different myocardial motion estimation methods, regularization strategies and quantitative indicators used to evaluate cardiac function will be overviewed in Chapter 2.



(a) (b)

Figure 1.2: Example of dense motion fields corresponding to (a) systolic motions and (b) diastolic motions of the heart.

1.2 Ultrasound imaging

Among the non-invasive techniques of cardiac function evaluation, medical imaging is used to assess its mechanical action by means of various modalities, such as magnetic resonance imaging (MRI), computed tomography (CT) and ultrasound imaging (UI). The most common MRI techniques for cardiac motion estimation include tagged-MRI [Prince 1992, Liu 2009] and cine-MRI [Spottiswoode 2007]. These modalities have the advantage of providing detailed morphological information and good tissue contrast. Similarly, CT is characterized by a high spatial resolution. However, because of its relatively high temporal resolution (typically from $\sim$ 10 to 1000 frames per second depending on the acquisition method), UI is more adapted to the rapid motion of the heart ($\sim$ 4-15 cm.s^{-1}). In contrast with MRI and CT, UI also presents important advantages such as portability, low budget requirements and reduced discomfort for the patient. This makes UI, particularly echocardiography, the most widely used modality in cardiology.

Ultrasound image acquisition

The acquisition of ultrasound (US) images is based on the interactions between emitted acoustic waves and the tissues of organs. The so-called echoes are generated and received by a transducer, which consists of an array of piezoelectric elements capable of producing US waves. Differences in the structure of the tissues, *i.e.*, their acoustic impedance, result in the phenomena of reflection and scattering. Reflections of the echoes back to the transducer occur typically at anatomical boundaries, while scattering results from rough or small structures causing the so-called speckle noise. The time delays between the received echoes provide measurements of the locations of structural changes in the tissues. In medical US, the transmitted waves usually have frequencies that vary between 2 MHz and 10 MHz [Angelsen 2000], and are around 4 MHz in echochardiography. Fig. 1.3 illustrates the above-mentioned principles of UI.

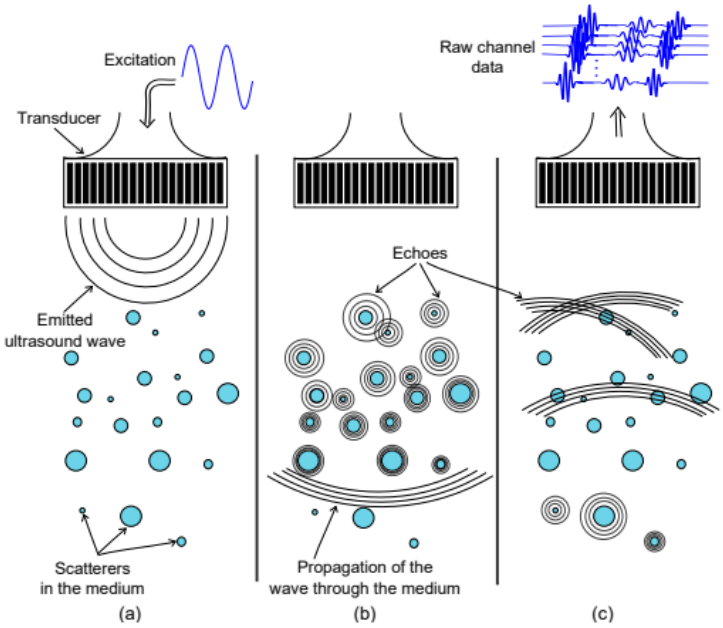


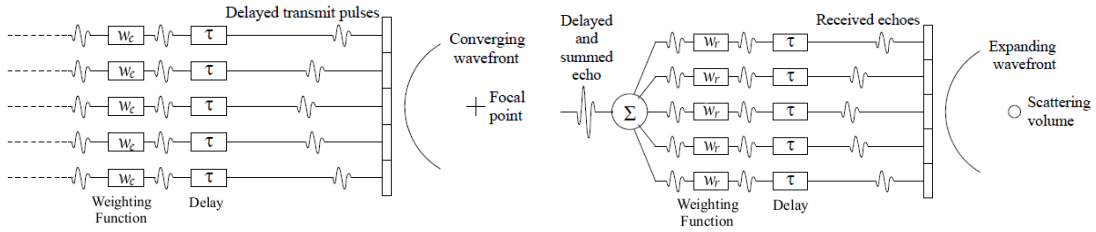
Figure 1.3: The principles of UI starting with (a) the generation and transmission of the US waves by the transducer, followed by (b) the interactions with anatomical structures and finally (c) the reception of the reflected and scattered US waves forming the raw data [SZASZ 2016].

Three main configurations of transducers can be used to acquire US images, *i.e.*, linear, convex and phased array transducers. In the former, the piezoelectric elements are arranged on a line, producing rectangular scans. In contrast, phased and convex array transducers use different acquisition angles, providing sector scans with larger fields of view. Phased array transducers are characterized by their small footprint (*i.e.*, aperture), wide coverage and depth. These transducers are therefore the tools of choice in echocardiography, which requires deep images to be acquired from a restricted area

(*i.e.*, between the ribs). In standard US imaging, the transducer is used to form image lines by sequentially scanning portions of the global imaged area. When imaging moving organs, such as the heart, one can increase the frame rate by simultaneously using all the elements in the transducer array. This strategy is called plane or diverging wave imaging and results in so-called ultrafast frame rates ($>$ 1000 frames per second). However, ultrafast UI can cause a significant decrease in the spatial resolution of the images.

Beamforming

The waves received by the transducer are initially in the form of raw radio frequency (RF) data. In a second step, the received echoes are shaped into a so-called beam. This process of beamforming aims at concentrating the energy into a narrow beam, allowing local information to be captured. Beamforming can be achieved mechanically, for example, by using a concave lens that focalizes the beam in a given area or direction. For array transducers, electronic beamforming is used to adjust the off-sets of the different received signals. This technique also allows a specific zone or direction of interest to be emphasized. The most commonly used beamforming technique is based on the delay and sum method. In the latter, one seeks to sum the US waves according to their corresponding time delays. A weighting process allows the desired focus to be controlled. Fig. 1.4 illustrates the principles of the delay and sum beamforming method. Adaptive beamforming strategies have also been investigated in order to enhance the spatial resolution and contrast of US images. As opposed to the delay and sum method, these techniques estimate the beamforming weights from the raw RF data itself, *e.g.*, based on the Minimum Variance method [Capon 1969] or maximum likelihood approaches [Krim 1996]. Note, however, that these adaptive methods are rarely used in clinical practice. In this thesis, we use US images acquired using classical beamforming techniques.



(a) Emission (b) Reception

Figure 1.4: The delay and sum beamforming method [Lorintiu 2015].

B-mode images

As explained above, the beamforming of the received raw signals is used to obtain RF lines. Each line corresponds to a column in the obtained US image. However, due to the difficult interpretation of the beamformed RF data, several transformations are

performed to obtain the final so-called B-mode images that are usually visualized in clinical practice. More specifically, envelop detection is performed by the demodulation of RF signals. In a second step, log-compression is used to reduce the image dynamics, followed by a scan conversion. The latter step allows the images to be visualized in the cartesian coordinate system when using a sector scanning as in echocardiography. Note that scan-conversion is not required for other UI applications using linear transducers. Fig. 1.5 illustrates these main steps, and shows an example of the obtained scan converted B-mode images. Aside from the above-mentioned steps, several other transformations are involved in the formation of US images, such as interpolation, filtering and time gain compensation.

The final B-mode images are contaminated by the speckle noise, which is an important characteristic of UI. The interactions between the emitted waves and anatomical structures that have small dimensions compared to the US wavelength, result in the scattering phenomenon. Scattering causes destructive or constructive interference patterns, *i.e.*, speckles. The speckle noise can be viewed as an undesirable property affecting the quality of US images. Several despeckling techniques have been investigated to mitigate this noise, thereby facilitating the interpretation of the corrupted images [Achim 2001, Loizou 2005, Michailovich 2006]. However, the speckle noise can also be exploited to analyse and extract valuable information from US images, *e.g.*, using its statistical properties for the purpose of myocardial segmentation [Mignotte 2001]. Another useful characteristic of speckle is that it follows the motion of the underlying tissues. The latter property makes it a reliable spatial marker for tracking moving organs, *e.g.*, the heart. The methods relying on the speckle characteristics for cardiac motion estimation will be discussed in Section 2.2.1.

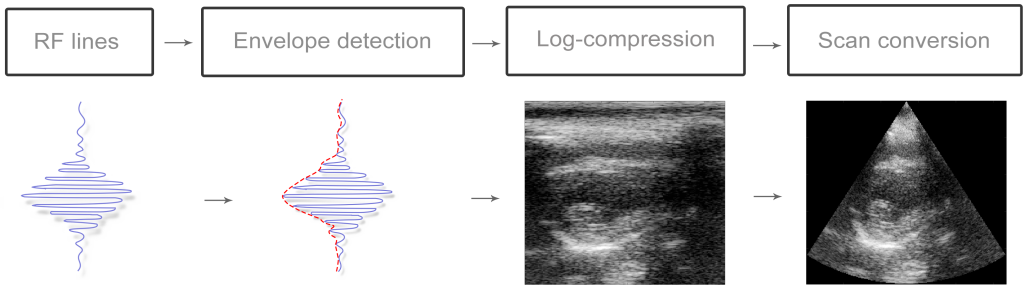


Figure 1.5: Transformations used to obtain B-mode images from the original RF data.

2D and 3D imaging

Despite its many advantages, 2D UI suffers from out-of-plane motions and limited geometrical information. While 2D US requires a separate analysis of the different imaged planes, the emerging 3D imaging techniques enable a more complete visualization and assessment of the heart anatomy. Conversely, 3D data suffer from the problems of frame

rate and image quality in the azimuthal direction. Furthermore, an advantage of a fully 3D deformation analysis remains a hypothesis and is currently regarded as an experimental method [Mirea 2016, Alessandrini 2015]. Also, 2D echocardiography is still more routinely used than 3D in clinical practice. In this thesis, we will focus on cardiac motion estimation in 2D UI.

1.3 Open challenges

Due to the well established clinical feasibility and advantages of 2D echocardiography, and despite the arrival of new technologies such as 3D imagery [Compas 2014, Compas 2012, Wong 2013], the development of new 2D cardiac motion estimation methods is still an active area of research, *e.g.*, [Nagata 2015, D'hooge 2015, Khamis 2016, Aviles 2017]. In this context, many challenges have to be resolved in order to achieve the objective of reliable cardiac function assessment. For example, there is still a need for deriving accurate and reproducible quantitative measures of motion to overcome the current state of inter-vendor variability of left ventricular (LV) longitudinal strains. Furthermore, the assessment of the cardiac function is still limited to global measurements [Alessandrini 2015] and undergoes great amounts of smoothing, causing loss of clinically valuable local information [Mirea 2016]. An accurate local analysis of the cardiac deformation has a major impact on the diagnosis, treatment choice and timing of surgical interventions in many clinical cases, *e.g.*, ischemia, valvular heart disease and early detection of adverse cardiac effect of chemotherapy in oncology. Therefore, new motion estimation strategies that limit the loss of structural and local information are needed in the process of endorsing regional strains [Mirea 2016]. In particular, developing more adaptive alternatives to the purely geometrical regularizations used for cardiac motion estimation is still an open challenge. In the context of UI, 2D cardiac motion estimation is still a difficult task. In particular, US images are characterized by a poor signal-to-noise ratio caused by the speckle noise. Another drawback of UI is the presence of acquisition related artefacts that can affect cardiac motion estimation. Moreover, in the case of 2D UI, out-of-plane motions cause discrepancies in the speckle pattern, leading also to erroneous motion estimates. More generally, the smoothness assumptions typically used for cardiac motion estimation can be violated, *e.g.*, in the case of anatomical boundaries. These shortcomings still call for new robust motion estimation strategies that mitigate the effects of outliers in cardiac UI. Furthermore, several recent works have attempted to address the spatio-temporal nature of cardiac motion [De Craene 2012, Zhijun 2014, McLeod 2015]. However, many current methods either suffer from problems of large motions between distant frames, or do not process the image sequences as a whole. Therefore, it is still an open challenge to efficiently incorporate temporal aspects into the problem of cardiac motion estimation from US image sequences.

In Chapters 3, 4 and 5 of this thesis, we attempt to address the above-mentioned challenges by proposing three new cardiac motion estimation methods for echocardiography. Further details about the contributions of this thesis are provided in the following

section.

1.4 Contributions and outline

Contributions

The objective of this thesis is to address the challenges explained in Section 1.3, namely those related to regularization, robustness to outliers and spatio-temporal motion estimation. These contributions are the following:

1. We introduce a new method for 2D cardiac motion estimation in US images. Motivated by the success of sparse representations in a variety of problems [Elad 2006, Dong 2011, Zhang 2010], we promote the use of a sparsity-based regularization for cardiac motion estimation. The proposed method combines a US specific similarity measure with spatial smoothness and sparse regularizations, exploiting jointly the statistical nature of B-mode images, the smoothness and the sparse properties of cardiac motion [J2,C3]. More specifically, we promote the use of a regularization exploiting a sparse representation of cardiac motion based on dictionary learning. In [C4], the proposed method is validated using clinical data of a specific cardiac pathology. Furthermore, the proposed regularization strategy has been investigated within a more general OF framework in [C2].

2. A fully robust cardiac motion estimation strategy is then investigated [J1]. This method addresses the limitations of UI and motion estimation constraints by mitigating the effect of outliers. The problem is formulated within a general OF-based framework with spatial smoothness and sparsity-based regularizations. Robustness to outliers, such as imaging artefacts and anatomical motion boundaries, is introduced using robust weighting functions for the data fidelity term as well as for the two regularization terms. Furthermore, the use of a robust sparse coding step has been investigated in [C1].

3. A new method for 2D spatio-temporal cardiac motion estimation is proposed. In this work, the US image sequence is treated as a whole, *i.e.*, groupewise, allowing us to exploit the temporal consistency of the motions. The simultaneous estimation of all the frames in the sequence is performed using an efficient optimization strategy based on the C-SALSA algorithm. Moreover, piecewise temporal smoothness is incorporated using a regularization in the time domain.

Thesis outline

This thesis is organized as follows. In Chapter 2, the main categories of cardiac motion estimation methods are presented. In particular, details about the different data fidelity terms and regularization strategies are provided. An overview of the theory related to

sparse representations and dictionary learning is also provided. In Chapter 3, we introduce the proposed cardiac motion estimation strategy based on a sparse regularization [J2]. The proposed method is evaluated using synthetic data, simulated realistic cardiac images, as well as healthy and pathological sequences of *in vivo* data. Chapter 4 presents the proposed robust motion estimation method [J1]. The proposed approach is tested using synthetic images and realistic simulated cardiac sequences with available ground-truth. The feasibility of the method for real data is also demonstrated using *in vivo* images. In Chapter 5, the proposed groupewise cardiac motion estimation method is introduced. A preliminary evaluation using a simulated cardiac sequence is presented. Finally, conclusions and possible directions for future work are reported in Chapter 6.

List of publications

International Journal Papers

[J1] N. Ouzir, A. Basarab, O. Lairez, and J.-Y. Tourneret, "Robust Optical Flow Estimation in Cardiac Ultrasound images Using a Sparse Representation", *submitted to IEEE Transactions on Medical Imaging*, March 2018.

[J2] N. Ouzir, A. Basarab, H. Liebgott, B. Harbaoui, and J.-Y. Tourneret, "Motion estimation in echocardiography using sparse representation and dictionary learning", *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 64–77, Jan 2018.

International Conference Papers

[C1] N. Ouzir, P. Chiril, A. Basarab, and J.-Y. Tourneret, "Cardiac Motion Estimation with Dictionary Learning and Robust Sparse Coding in Ultrasound Imaging", *IEEE International Ultrasonics Symposium (IUS)*, Kobe, Japan, October, 2018.

[C2] N. Ouzir, A. Basarab, and J.-Y. Tourneret, "Optical Flow Estimation in Ultrasound Images Using a Sparse Representation", in Proc. *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curaçao, Dutch Antilles, Dec. 2017.

[C3] N. Ouzir, A. Basarab, and J.-Y. Tourneret, "Cardiac motion estimation in ultrasound images using spatial and sparse regularizations", in Proc. *IEEE International Conference on Image Processing (ICIP)*, Beijing, China, pp. 780 -784, Sept. 2017.

[C4] N. Ouzir, O. Lairez, A. Basarab, J-Y. Tourneret, "Tissue motion estimation using dictionary learning: application to cardiac amyloidosis", in *IEEE International Ultrasonics Symposium*, Washington, pp. 1-4, Sept. 2017.

Chapter 2

State-of-the-art

Contents

2.1 Introduction

This thesis investigates the problem of cardiac motion estimation in echocardiography using sparse representations and dictionary learning. Thus, the present chapter will be divided into two parts. The first section overviews the state-of-the-art methods in cardiac motion estimation, especially, in UI. In particular, the different data fidelity terms and regularization strategies investigated for cardiac motion estimation are presented. The second section reviews the basic theory related to sparse representations and dictionary learning, which will be an important part of the methods presented in this thesis. The algorithms used to solve these two problems are also explained.

2.2 Cardiac motion estimation

Motion estimation aims at recovering the motion of one or multiple objects of interest in a sequence of images. Motion estimation is used in a large variety of applications such as computer vision [Brox 2004], remote sensing [Bentoutou 2005] or medical imaging [Compas 2014, Huang 2014, Luo 2008]. These applications either seek to analyse the motion of the moving objects themselves (*e.g.*, organs in medical imaging) or use motion estimation for a number of other image processing tasks, such as video compression [Furht 2012] or super-resolution [Shen 2007]. In medical imaging, motion estimation has been exploited for different modalities including MRI [Oksuz 2015, Rueckert 1999b], US [Cohen 2002a] or CT [Onofrey 2015]. Estimating the motion of organs and tissues is useful in many medical imaging applications, for example, compensation of respiratory motions for surgerical interventions [Dawood 2006], change detection in cancer treatements through tumor motion estimation [Lu 2012] or motion estimation-based segmentation of MR brain scans [Vemuri 2003]. In cardiology, the tracking of the myocardium, or myocardial boundaries, is a key tool for the analysis and diagnosis of cardiovascular diseases [Abraham 2007, Cottrell 2010].

The objective of cardiac motion estimation is to quantify the deformation of the heart tissues, in particular, the myocardial wall in the LV. The estimated motions are used to asses the mechanical function of the heart and thereby aid in the diagnosis of cardiovascular diseases. Cardiac motion estimation can consist in tracking selected features, essentially anatomical boundaries, or recovering dense motion fields. In feature-based motion estimation, an initial set of features is identified prior to the tracking step. Thus, the tracking of anatomical boundaries requires a preprocessing step during which the initial contours are located, for example, the LV endocardium. After the tracking, a sequential segmentation of the heart is obtained. While boundary tracking approaches are computationally inexpensive due to the small number of tracked pixels, they do not provide information about the motion of the entire organ, *i.e.*, the heart. In contrast with the feature-based approach, dense motion estimation methods do not require preprocessing steps, but rather use the intensity structure of the images to estimate the motions of all pixels. This pixel-wise approach has the advantage of better describing the local motions of the tissues, allowing for a more detailed analysis of the undergone

deformation. This type of estimation comes, however, at the cost of higher computational requirements. Dense cardiac motion estimation strategies include the methods relying on block-matching [Suhling 2005, Alessandrini 2013a], optical flow [Baraldi 1996] or elastic image registration [Ledesma-Carbayo 2005]. Subsection 2.2.1 overviews these three categories of cardiac motion estimation methods, with a particular focus on the strategies developed in the context of cardiac US.

Two essential components in motion estimation are the data fidelity and regularization constraints. The former seeks to find correspondences between objects based on the observed characteristics of the images, *e.g.*, the pixel intensities. It can be formulated as a matching criterion to be minimized, for example, based of optical-flow or a similarity measure. Subsection 2.2.2 provides further details about different data fidelity terms employed for cardiac motion estimation. The second important component is the regularization strategy. Motion estimation from images is an ill-posed problem, in the sense that the corresponding inverse problem does not have a unique solution. Additional constraints are therefore necessary to solve motion estimation problems. These constraints often come in the form of spatial or temporal regularizations. They can also be formulated as additional explicit terms or can be implicit, *e.g.*, using parametric transformations. Subsection 2.2.3, overviews the regularization strategies used in the context of cardiac motion estimation..

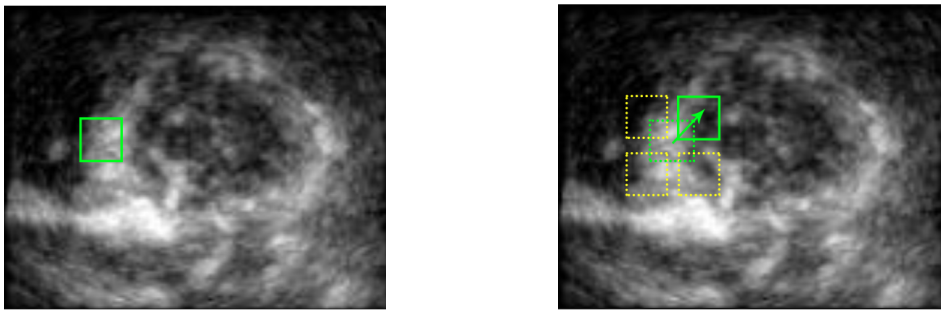
Following motion estimation, the interpretation of the resulting motion fields is usually achieved using quantitative indicators, such as strain, that provide insight into the deformations undergone by the myocardium. This section ends with a presentation of the quantitative indicators of cardiac motion in Subsection 2.2.4.

2.2.1 Motion estimation methods

Most of the techniques used for cardiac motion estimation fall into three main categories: block-matching (BM), optical flow (OF) and elastic image registration. BM algorithms consist in matching blocks in consecutive images using a predefined criterion. In the context of UI, BM-based algorithms are often referred to as speckle tracking methods [Kaluzynski 2001, Byram 2013]. The second category of cardiac motion estimation methods is OF. Gradient-based OF combines differential methods with the over time pixel intensity constancy assumption [Horn 1981]. OF methods have been very widely used in medical imaging [Dawood 2006], and more specifically for US cardiac motion estimation [Suhling 2005, Duan 2007, Tavakoli 2014]. Finally, in elastic registration, the images are related through a non-rigid geometric transformation. The deformations investigated for cardiac motion estimation are usually parametric [Ledesma-Carbayo 2005, Myronenko 2009a]. In elastic registration methods, the image characteristics can be easily incorporated using different similarity measures. Further details about BM, OF and registration methods used for cardiac motion estimation are provided in the following subsections.

2.2.1.1 Block-matching

Block-matching algorithms (BM) are among the most widely used techniques for cardiac motion estimation [Boukerroui 2003, Kaluzynski 2001, Suhling 2005, Alessandrini 2013b] and are still considered the standard implementation in current commercial systems [Alessandrini 2016a]. These methods consist in matching blocks between two consecutive images using a matching criterion. In order to reduce the computation time, the search is usually limited to a small region in the neighbourhood of the original block. This process is then repeated for all the blocks in the region of interest (*e.g.,* the myocardium) in order to obtain a global motion field. Due to the local nature of the matching process, BM is particularly appropriate for applications that only require motion estimates at a specific location in the image. Several BM methods use the image noise as a spatial marker for motion and rely on its statistical distribution to build a similarity metric [Strintzis 1997, Viola 2003]. Typical similarity measures used for BM include the sum the squared differences [Yeung 1998] and the sum of absolute differences [Kontogeorgakis 1994, Behar 2004b]. An overview of commonly used similarity measures will be provided in Subsection 2.2.2.1. In the context of UI, BM methods are usually referred to as speckle tracking algorithms. Fig. 2.1 illustrates the matching process employed in BM-based motion estimation.



(a) (b)

Figure 2.1: (a) Location of a block in the reference image. (b) Searching process in the second image: different blocks in the search region (yellow) and the matched block (green). The arrow shows the resulting motion between the two blocks.

Different BM or speckle tracking approaches have been studied for cardiac motion estimation. In [Behar 2004b, Suhling 2005], BM is paired with a local affine motion model for cardiac US. In [Suhling 2005], the problem of choosing the ideal block size is bypassed using an automatic multiscale window size strategy. A speckle tracking method for cardiac UI is also studied in [Boukerroui 2003]. This method uses a specific similarity measure for UI, based on the assumption of multiplicative Rayleigh noise (see Subsection 2.2.2.2). In addition to the methods designed for B-mode images,

BM has recently been investigated for RF US data in [Lopata 2011, Parajuli 2016]. While the high frequency RF signals are better suited for the estimation of small displacements, they suffer from motion sensitivity, and thus, require higher frame rates [Alessandrini 2016b]. Speckle tracking results using B-mode and RF data have been compared in [Yu 2006, Lopata 2009].

2.2.1.2 Optical flow

Optical flow (OF) methods rely on brightness constancy, *i.e.*, the assumption that the intensity of a pixel remains constant over a short period of time. The motion is then estimated by matching image intensities across frames. The brightness constancy assumption between two close time instants t and $t + \delta t$ can be formulated as follows

$$\boldsymbol{I}(\boldsymbol{x}, t) = \boldsymbol{I}(\boldsymbol{x} + \boldsymbol{U}, t + \delta t) \tag{2.1}$$

where $\boldsymbol{U}$ is the flow field, $\boldsymbol{I}$ contains the image intensities, and $\boldsymbol{x}$ denotes the pixel coordinates. OF provides a dense estimation of the motion field by means of the entire image [Mailloux 1989] or by a local window analysis [Suhling 2005]. In differential OF methods, also known as gradient-based OF, the motion is estimated by linking the spatial and temporal image intensity variations. This is achieved by taking the 1st order Taylor series expansion of the displaced image, *i.e.,*

$$\boldsymbol{I}(\boldsymbol{x} + \boldsymbol{U}, t + \delta t) \approx \boldsymbol{I}(\boldsymbol{x}, t) + \partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T \boldsymbol{U} \tag{2.2}$$

where $\nabla \boldsymbol{I}$ is the spatial intensity gradient and $\partial_t \boldsymbol{I}$ is the temporal derivative of $\boldsymbol{I}$ at time t. Equations (2.1) and (2.2) lead to the so-called OF constraint equations

$$\partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T \boldsymbol{U} = 0. \tag{2.3}$$

In the original formulation of Horn and Schunck (HS) [Horn 1981], the ℓ_2-norm was used to formulate a data fidelity term to be minimized based on (2.3), *i.e.,*

$$E_{\text{HS}} = \|\partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T \boldsymbol{U}\|_2^2. \tag{2.4}$$

OF methods have been used successfully in a large variety of applications ranging from computer vision [Brox 2004] to more specific ones such as atmospheric motion estimation in meteorology [Héas 2008]. These methods have also been investigated for various medical imaging modalities (including UI [Alessandrini 2013c, Duan 2007, Suhling 2005], MRI [Prince 1992] and CT [Dawood 2006]) and for different clinical applications requiring cardiac motion estimation. In [Suhling 2005], myocardial OF estimation is performed using both a local affine motion model and a linear model in time. In [Duan 2007], OF is investigated for simulated real-time 3D US images. A variant of OF, exploiting the monogenic signal characteristics and a local affine motion model, is also considered

for cardiac US in [Alessandrini 2013c]. However, because of the intensity preservation assumption and their local nature, the OF-based methods are generally sensitive to noise and can fail at estimating large displacements. Consequently, the use of multi-resolution schemes or BM initializations is common for the methods relying on OF [Behar 2004a, Suhling 2005, Alessandrini 2013b].

2.2.1.3 Elastic registration

Elastic registration aims at aligning two images using a non-rigid geometric transformation. The deformation model can be parametric, *e.g.*, represented on a B-spline basis [Myronenko 2009b], or discrete [Woo 2009]. The relation between two images at consecutive time instants t and $t + \delta t$ can be formulated as follows

$$\boldsymbol{I}(\boldsymbol{x}, t) = \boldsymbol{I}(T(\boldsymbol{x}), t + \delta t) \tag{2.5}$$

where T denotes the non-rigid transformation, $\boldsymbol{x}$ contains the coordinates of the pixel of interest and $\boldsymbol{I}$ is the image intensity vector. The motion between the two images is obtained when a correspondence is found according to a predefined criterion. In contrast with OF-based algorithms, elastic image registration allows a similarity measure to be specified based on the image characteristics. The transformation parameters that minimize the similarity measure provide the final motion between the images. Commonly used similarity measures are presented in Subsection 2.2.2.1.

The non-rigid parametric transformation models fall into two main categories: physical and basis functions-based models. In the context of cardiac registration, typical physiological models account for the elasticity and incompressibility properties of the myocardial tissue [McLeod 2012]. In the second category of models (*i.e.,* based on function representations) the deformation is usually parametric. For example, in the B-spline parametrization, a free form transformation model (FFD) [Sederberg 1986] allows a mesh of control points to be deformed. The deformation at each pixel is then recovered using a tensor product of the B-spline functions. This model allows the transformation to be controlled locally (around the control points), while enabling a fast computation with the use of a reduced number of control points. Fig. 2.2 shows an example of a registered pair of cardiac US images and the corresponding deformed B-spline mesh.

Non-rigid transformations are well suited to medical imaging problems, since the motions of soft tissues are inherently non-rigid. For example, an elastic registration method combining B-mode with tissue Doppler images is studied in [Porras 2016]. In [Myronenko 2009b], a US specific similarity measure is used to register 3D echocardiographic images with a B-spline parametrization. A spatiotemporal B-spline model is used for time consistent elastic registration in [Ledesma-Carbayo 2005] and [De Craene 2012]. All of the above-mentioned methods employ the B-spline parametrization, which is a very widely used model for cardiac motion estimation. Other transformation models investigated in the context of cardiac registration include transformations based on radial basis functions (RBF) [Parajuli 2015] or piecewise affine models that account for typi-

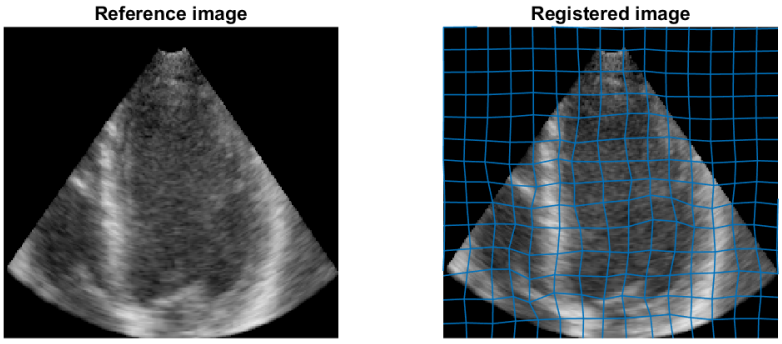


Figure 2.2: Resulting B-spline mesh for a pair of simulated US cardiac images using the method of [Myronenko 2009b].

cal motions of the heart (*i.e.*, translation, rotation, expansion, compression, and shear) [Suhling 2005, Alessandrini 2013b].

2.2.2 Data fidelity

One key element in all categories of motion estimation problems is the data fidelity term. This term formulates the criterion that has to be satisfied in order to find a correspondence between the reference image and the displaced one. The data fidelity term is based entirely on the observations, *i.e.*, on the information provided by the images, for example, the pixel intensities. In OF-based methods, the data fidelity term relies on the brightness constancy assumption and incorporates the spatial and temporal image intensity variations (see Subsection 2.2.1.2). Other types of data fidelity terms can be referred to as similarity measures. These measures are pixel-wise and are therefore typically used by BM and elastic registration methods to recover dense motion fields. Similarity measures can be based on the statistical characteristics of the image noise [Cohen 2002b], or on other features derived from the intensity structure of the images, *e.g.*, phase information [Alessandrini 2013b].

2.2.2.1 Similarity measures

One of the most common and simple similarity measures is the sum of squared differences (SSD) [Yeung 1998]. This measure computes the square of the difference between each pixel x in the reference image and its counterpart in the displaced one, *i.e.*,

$$\mathrm{SSD}(\boldsymbol{I}, t) = \sum_{x} [\boldsymbol{I}(x, t) - \boldsymbol{I}(x, t + \delta t)]^2 \tag{2.6}$$

where t and $t + \delta t$ stand for two consecutive time instants and $\boldsymbol{I}$ contains the image intensities. Due to the squared differences, higher values are assigned to pixels providing large distances. This is why the main drawback of SSD is its sensitivity to outliers.

Instead of using a quadratic formulation as in SSD, the sum of absolute differences (SAD) [Kontogeorgakis 1994] uses the absolute value of the distances. This approach allows the effect of pixels with large differences to be reduced. The SAD measure can be formulated as follows

$$\mathrm{SAD}(\boldsymbol{I}, \boldsymbol{U}) = \sum_{x} |\boldsymbol{I}(x,t) - \boldsymbol{I}(x+U, t+\delta t)|. \tag{2.7}$$

Other common similarity measures include the cross-correlation (CC) [Behar 2004b, Duan 2007] and normalized cross-correlation (NCC) measures. These measures seek to match pixels by maximizing the amount of correlation between the images. The CC measure can be formulated as follows

$$\mathrm{CC}(\boldsymbol{I}, t) = \frac{[\sum_x (\boldsymbol{I}_t - \bar{\boldsymbol{I}}_t)(\boldsymbol{I}_{t+\delta t}(\boldsymbol{U}) - \bar{\boldsymbol{I}}_{t+\delta t}(\boldsymbol{U}))]^2}{\sum_x (\boldsymbol{I}_t - \bar{\boldsymbol{I}}_t)^2 \sum_x (\boldsymbol{I}_{t+\delta t}(\boldsymbol{U}) - \bar{\boldsymbol{I}}_{t+\delta t}(\boldsymbol{U}))^2} \tag{2.8}$$

where $\boldsymbol{I}_t = \boldsymbol{I}(x,t)$, $\boldsymbol{I}_{t+\delta t} = \boldsymbol{I}(x+U, t+\delta t)$ and $\bar{\boldsymbol{I}}_t$, $\bar{\boldsymbol{I}}_{t+\delta t}$ denote the mean of the corresponding intensities. Mutual information (MI), has also been investigated for motion estimation in medical imaging [Elen 2008]. In MI, the images are statistically related, and one seeks to maximize the low joint entropy and high marginal entropies of the images. This allows one to assess how well the first image predicts the second one. MI has been particularly used in the context of multi-modal motion estimation problems [Onofrey 2015].

Most of the similarity measures used for motion estimation can be related to the statistical distribution of the image noise [Strintzis 1997]. More specifically, the minimization of the similarity measure leads to the maximum likelihood (ML) estimator of the motion for the corresponding statistical distribution. For example, the SSD measure makes the assumption of an additive Gaussian noise, while the SAD measure implies a Laplacian noise distribution. In UI, the ML approach has been used to incorporate specific assumptions about the speckle noise distribution [Cohen 2002a] (see Subsection 2.2.2.2).

2.2.2.2 Data fidelity terms in UI

In addition to the above mentioned similarity measures, other data fidelity terms based on the characteristics of US images have been investigated for echocardiography. For example, some methods have combined the B-mode data with Doppler imaging to construct a data fidelity term using, *e.g.*, single [Tavakoli 2014] or multiplane images [Porras 2016]. These methods benefit from the high temporal resolution of Doppler imaging and have been shown to be more resilient to image noise. Other methods based on statistical noise characteristics [Cohen 2002a] or phase information [Alessandrini 2013a] have also been investigated. The main works based on these two approaches are overviewed below.

Statistical methods

In [Strintzis 1997], the statistical properties of the speckle noise are incorporated into the motion estimation problem using the ML approach. This method is based on the assumption that US images are contaminated by a multiplicative Rayleigh distributed noise. However, it is assumed in [Strintzis 1997] that only one of the two images used in the motion estimation problem is corrupted by noise. In [Cohen 2002a], a Rayleigh multiplicative noise distribution is considered for both images, leading to an improved accuracy. The proposed similarity measure is particularly adapted to B-mode US images due to the use of the natural logarithm accounting for log-compression. In [Myronenko 2009b], this measure is extended in order to account for speckle correlation using a generalized Rayleigh distribution. In this thesis, we will make use of the Rayleigh multiplicative noise assumption for motion estimation in echocardiography. Further details about the computation of this data fidelity term will be provided in Chapter 3.

Phase-based data fidelity

As an alternative to the intensity-based data fidelity terms, which are subject to speckle decorrelation and noise, phase-based methods have also been investigated for UI. The instantaneous phase information can be obtained using complex or hypercomplex image representations. A typical example of such representations is the 2D analytic signal, which has been employed mainly for cardiac motion estimation using RF data [Alessandrini 2014, Salles 2015]. Since phase information is absent in the lateral direction, these methods generate transverse oscillations using, *e.g.*, filtering [Salles 2015] or specific beamforming techniques [Alessandrini 2014]. Another approach is based on the monogenic signal theory, which allows local amplitude, phase and orientation information to be captured from the original intensity images. In [Alessandrini 2013a], the classical intensity-based OF formulation is replaced with a phase conservation term based on the monogenic signal. The amplitude, phase and orientation features are provided by the responses to 2D spherical quadrature filters. The phase and orientation allow the sensitivity to intensity changes and noise to be bypassed, which are the main drawbacks in many OF-based methods. In this work, the monogenic phase-based data fidelity term is employed in a framework similar to the one of [Suhling 2005], with an affine transformation model and an automatic selection of the window size. Due to its general formulation, this method is well-suited to various medical imaging modalities and particularly echocardiography [Alessandrini 2013a].

2.2.3 Motion regularization

Another key element in motion estimation is regularization. Because motion estimation problems are generally formulated as ill-posed inverse problems, it is necessary to introduce some *a priori* information about the motion, *e.g.*, regarding the way it is expected to vary spatially. Common priors used for motion fields include spatial or temporal smoothness [Elen 2008, McLeod 2012]. These smoothness assumptions can be explicitly

incorporated into the motion estimation problem by means of regularization constraints [Horn 1981], but also using parametric motion models, *e.g.*, based on basis functions or physiological properties of the heart.

2.2.3.1 Explicit spatial smoothness

One way of overcoming the ill-posed nature of motion estimation is to combine the data fidelity term with an explicit regularization term. Following this approach, the motion estimation problem can be formulated as an energy minimization of these two terms as follows

$$\min_{\boldsymbol{U}} E_{\text{data}}(\boldsymbol{U}) + \lambda E_{\text{reg}}(\boldsymbol{U}) \tag{2.9}$$

where $\boldsymbol{U}$ is the motion field to be estimated and $\lambda \in \mathbb{R}^+$ is a parameter that controls the influence of the regularization term. A common way of enforcing spatial smoothness is through the penalization of the gradient [Baraldi 1996] or the Laplacian of the motion field [Ledesma-Carbayo 2008], *i.e.*,

$$E_{\text{reg}}(\boldsymbol{U}) = \phi(\boldsymbol{G}\boldsymbol{U}) \tag{2.10}$$

where ϕ is an appropriate penalty function and $\boldsymbol{G}$ stands for the gradient or Laplacian operators. The regularization term (2.10) enforces small pairwise differences between the displacements of each pixel and its local neighbourhood, leading to spatially smooth motions.

A typical choice for the function ϕ is the squared ℓ_2-norm that promotes smoothness but also penalizes motion discontinuities [Horn 1981]. The ℓ_2 regularization has been widely used for cardiac motion estimation, in particular, for OF-based methods [Baraldi 1996, Tavakoli 2014], but also in cardiac US registration [Parajuli 2015, Zhijun 2014, Myronenko 2009b]. Another choice for the spatial regularization function is the ℓ_1-norm, which also enforces smoothness while penalizing motion discontinuities less strictly. The ℓ_1-based spatial regularization is usually referred to as the ℓ_1 total variation (TV-ℓ_1).

Bending energy

In cardiac motion estimation, a commonly used spatial regularization is the so-called bending energy term [Rueckert 1999a]. This smoothness term is related to the bending energy of a 2D thin-plate of metal, and has been typically used in registration methods to further regularize parametric motion models such as B-spline, that are prone to unrealistic deformations [Elen 2008, Heyde 2013] (see Subsection 2.2.3.2). This regularization term can be formulated as follows

$$E_{\text{reg}}(\boldsymbol{r}, \boldsymbol{\mu}) = \left\| \frac{\partial^2 \boldsymbol{T}(\boldsymbol{r})}{\partial \boldsymbol{r} \partial \boldsymbol{r}^T} \right\|_F^2 \tag{2.11}$$

where T denotes the parametric transformation with parameters $\boldsymbol{\mu}$ at the points $\boldsymbol{r} = [x, y, z]$ and $\|.\|_F$ is the Frobenius norm.

Interpolation and filtering

Another approach used to explicitly enforce spatial smoothness is to use Gaussian filtering. More specifically, the motion is assumed to be a convolution of an auxiliary field with a Gaussian kernel [Somphone 2013]. Typically, Gaussian filtering, as well as interpolation techniques, have been used as post-processing steps for BM-based algorithms [Suhling 2005], which usually result in noisy motion fields.

2.2.3.2 Regularization in parametric transformations

Regularization can also be achieved using a parametric motion model. This type of regularization is classical in the methods relying on elastic registration [Ledesma-Carbayo 2005], but have also been used in other motion estimation approaches such as BM [Suhling 2005]. By narrowing down the number of possible transformations, parametric models ensure uniqueness of the solution. This is mainly achieved by employing a small set of parameters to be estimated. For example, by imposing an affine transformation, one only allows for specific types of motions, *i.e.*, translation, rotation, expansion, compression, and shear [Suhling 2005, Alessandrini 2013b]. The set of possible transformations is also reduced by using a low dimensional set of control points, as in the B-spline or RBF-based transformations. Another source of implicit regularization in basis functions-based parametrizations, such as B-spline and RBFs, is the interpolation used to recover the motion between the control points. Due to the use of a reduced set of basis functions, the resulting global motion field is spatially smooth. Note, however, that such implicit smoothness does not prevent nonphysical deformations, *e.g.*, folding or shrinking [Elen 2008]. Therefore, several cardiac motion estimation methods using parametric transformations, still rely on explicit smoothness terms or physiological assumptions to further constrain the motion [Myronenko 2009b, Elen 2008].

Physiological constraints

Regularizations using non-rigid transformations based on physical properties of the heart, *e.g.*, tissue elasticity or incompressibility of the myocardium, have also been investigated. The incompressibility property is based on the myocardial volume conservation during the cardiac cycle. One way of ensuring incompressibility is by constraining the motion fields to be divergence-free. The incompressibility property is used to recover dense cardiac motion fields from tagged-MRI in [Liu 2009], 2D echocardiography in [Parajuli 2015] and 3D echocardiography in [Elen 2008, De Craene 2012]. Elasticity has also been investigated for 3D cardiac motion estimation from MRI and echocardiography in [McLeod 2012]. In addition to incompressibility and elasticity, other physiological regularizations have been studied for cardiac motion estimation in UI. For example,

a topology preserving regularization allowing expansions and contractions to be controlled is studied in [Aviles 2017] and an invariant shape-based constraint is proposed in [Myronenko 2010].

2.2.3.3 Temporal regularization

Several dense motion estimation methods have been proposed to exploit the temporal properties of cardiac motion. A common approach consists in imposing smoothness in the time domain. More specifically, temporal smoothness is based on the assumption that physical points inside the myocardium follow smooth trajectories. The two main strategies investigated for temporal regularization of cardiac motion are the groupewise [Ledesma-Carbayo 2005] and sequential tracking methods [Zhang 2011, Zhijun 2014]. Other approaches have incorporated temporal smoothness as a post-processing step of pairwise cardiac motion estimation, *e.g.*, using interpolation [Parajuli 2015] or Kalman filtering [Myronenko 2010].

Groupwise estimation

The groupwise approaches use the image sequence as a whole, thus, allowing temporal coherence to be incorporated, *i.e.*, the estimation is carried out simultaneously for all the frames. These approaches usually estimate all the motions in the sequence with respect to a single reference frame. The main drawback of the reference frame approach is that it requires the estimation of large displacements between distant frames. Typically, these methods rely on a spatio-temporal B-spline parametrization [Ledesma-Carbayo 2005, Elen 2008, De Craene 2012]. These approaches based on a B-splines parametrization also allow temporal smoothness to be incorporatedn, leading to smooth motion trajectories. However, they do not take into account the correlation of motions between adjacent frames. In [De Craene 2012], this point has been addressed using smoothness of velocity over time instead of the displacements themselves. The idea behind this approach is that changes in velocity at a time instant, should modify the entire trajectory, thus enforcing dependency between small time steps.

Sequential registration

In contrast with the groupwise methods, the sequential strategies estimate the motions between consecutive frames in a pairwise approach. Thus, temporal regularization is incorporated sequentially by using the motions of the previous frame at each time instant [Zhang 2011]. These methods have the advantage of regularizing the motions of physical points between consecutive time instants. This also allows the problem of large inter-frame displacements to be bypassed. In [Zhijun 2014], multiple consecutive frames are used in a compromise strategy between the reference frame and sequential registration approaches. However, the sequential methods suffer from frame-to-frame accumulation errors, in addition to requiring a sequential registration of the whole sequence. Moreover,

the pairwise nature of this approach makes it difficult to incorporate time consistency in the sequence as a whole.

2.2.4 Quantitative indicators

In the previous sections, we have described cardiac motion estimation methods that result in dense motion fields. Such motion fields can be challenging to interpret, as they do not directly provide insight into the mechanical function of the heart (see Fig. 1.2). Therefore, other quantitative indicators are usually computed using the motions provided by dense cardiac motion estimation. These indicators allow the myocardial deformation to be interpreted by describing tissue characteristics such as elasticity and contractility. Commonly used quantitative indicators for cardiac function diagnosis include strain and torsion. The former describes changes in tissue length while the latter measures the rotational angle between the base and apex of the myocardium (see Fig. 2.3).

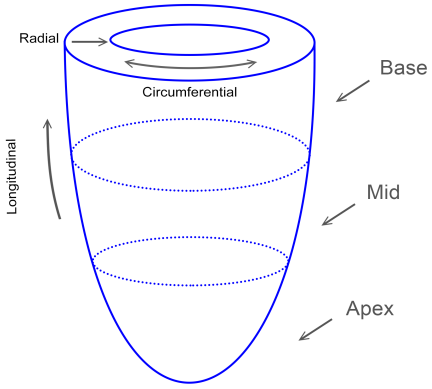


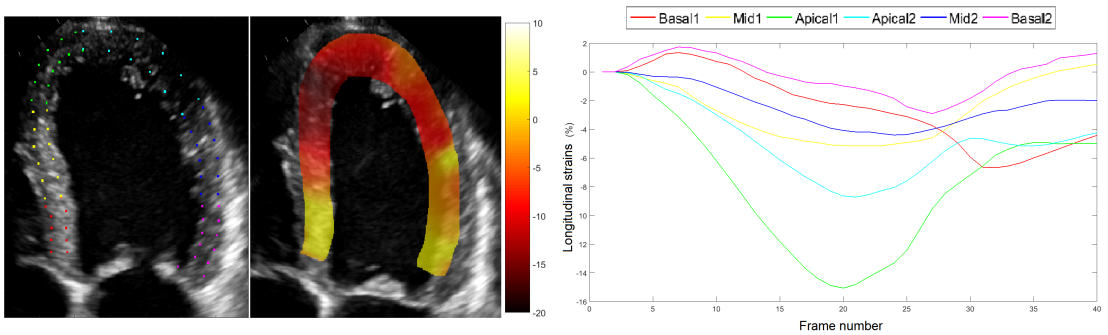
Figure 2.3: Longitudinal, radial and circumferential strain axes, and the basal, mid and apical regions of the heart.

Strain is very widely used to assess the amount of deformation of the myocardium throughout the cardiac cycle, *i.e.*, in the systole and diastole phases. One way to express strain is using the so-called Lagrangian strain. This type of strain describes the deformation of the myocardium with respect to its original shape, and can be computed along the circumferential, longitudinal or radial directions as shown in Fig. 2.3. One way of computing the Lagrangian strain consists in placing initial markers on the first frame of an image sequence and then measuring the distance d_k between adjacent points in each frame k of the cardiac cycle. Strain values s_k are obtained relatively to the first

frame as follows

$$s_k = \frac{d_k}{d_0} - 1 \tag{2.12}$$

where d_0 is the distance between points in the first frame. The strain is a dimensionless indicator that can be expressed in percentage, or visualized, *e.g.*, in the form of strain curves or strain maps. Fig. 2.4 shows an example of initial points used for strain computation, the obtained strain map for one frame of the sequence and the resulting strain curves. As shown in this figure, the myocardium is usually segmented into different anatomical regions of interest, *e.g.*, the base, mid-region and apex. These segments can then be used to compute average segmental strain values. In Fig. 2.4b, each curve corresponds to the average longitudinal strain for the corresponding region, *i.e.*, Basal1, Basal2 for the base segments, Mid1, Mid2 for the mid-region and Apical, Apical2 for the apex of the myocardium. Other types of indicators related to strain include strain rate, which describes the strain change with respect to time, and natural strain, which provides an instantaneous version of strain.



(a) (b)

Figure 2.4: (a) Example of initial points used for strain computation from an *in vivo* sequence and the resulting strain map for one frame. (b) The corresponding segmental strain curves for the entire sequence. The different colors indicate different segments of the myocardium.

2.3 Sparse representations and dictionary learning

By exploiting the sparse nature of signals or images, many recent works have proposed appropriate regularizations for solving different inverse problems [Elad 2006, Dong 2011, Mairal 2008a, Mairal 2008b]. For example, several methods have employed sparse representations for image denoising [Elad 2006, Dong 2011, Mairal 2008b], but also for inpainting [Mairal 2008b] or face recognition [Zhang 2010]. In particular, the above-cited methods exploit sparsity by expressing the unknown signal or image as a linear combination of a few elements of a dictionary. Furthermore, in these works, sparsity is introduced using adaptive dictionaries, *i.e.*, learned from the data itself, instead of predefined dictionaries. Aside from various signal processing applications, a few recent attempts to use sparse representations and dictionary learning for motion estimation have been investigated in the literature [Shen 2010, Jia 2011a]. In cardiac UI, appearance dictionaries were learned from two distinct cardiac regions for contour tracking [Huang 2014]. A similar strategy was considered in [Onofrey 2015] for brain segmentation in CT images. Finally, a dictionary of features was learned in [Oksuz 2015] and an SSD measure was used for the registration of cardiac MRI. In this thesis, we investigate the use of sparse representations and dictionary learning for motion estimation in cardiac US. The following subsections provide further details about the related theory, which are useful for the methods presented in this thesis.

2.3.1 Sparse representations

The goal of a sparse representation is to express a signal as a linear combination of only a few elements chosen from a collection of training signals, *i.e*, the dictionary. The underlying assumption is the redundancy and self-similarity properties of the signal of interest.

2.3.1.1 Problem formulation

Given a *dictionary* $\boldsymbol{D} \in \mathbb{R}^{n \times q}$, , *i.e.*, a set of elements called *atoms*, a vector $\boldsymbol{u} \in \mathbb{R}^n$ is represented by a weighted linear combination of few elements of $\boldsymbol{D}$, *i.e.*,

$$\boldsymbol{u} = \boldsymbol{D}\boldsymbol{\alpha} \tag{2.13}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^q$ is a sparse coefficient vector with few non-zero entries. When $n < q$, the dictionary is *overcomplete*. In this case, the sparse vector $\boldsymbol{\alpha}$ satisfying (2.15) is not unique, therefore calling for the definition of additional constraints. A classical way of exploiting the sparsity of $\boldsymbol{u}$ in the dictionary $\boldsymbol{D}$ is to look for the unknown vector $\boldsymbol{\alpha}$ with the minimum number of non-zero coefficients [Tosic 2011], *i.e.*, the solution of the following optimization problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ subject to } \boldsymbol{u} = \boldsymbol{D}\boldsymbol{\alpha} \tag{2.14}$$

where $\|.\|_0$ is the ℓ_0 *pseudo-norm*, which counts the number of non-zero elements of a vector. Since real data is generally affected by noise, equation (2.13) can be reformulated as follows

$$\boldsymbol{u} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{e} \tag{2.15}$$

where $\boldsymbol{e} \in \mathbb{R}^n$ is the additive noise. The minimization problem (2.14) leads to

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ subject to } \boldsymbol{u} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{e} \text{ and } \|\boldsymbol{e}\|_2^2 < \epsilon \tag{2.16}$$

where the constraint $\|\boldsymbol{e}\|_2^2 < \epsilon$ is motivated by the fact that the noise has bounded energy. An alternative to the error constrained formulation in (2.16), allowing the sparsity of $\boldsymbol{\alpha}$ to be controlled, can be written as follows

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{u} - \boldsymbol{D}\boldsymbol{\alpha}\|_2 \text{ subject to } \|\boldsymbol{\alpha}\|_0 \leq K. \tag{2.17}$$

In (2.17), one looks for the sparse coding vector $\boldsymbol{\alpha}$ that has a maximum of K non-zero elements and provides the best solution in terms of the residual norm of the approximation.

2.3.1.2 Approximate minimization

The solution of (2.17) is NP-hard. However, this problem can be solved using algorithms that provide good approximate solutions in polynomial time. The two main classes of algorithms that have been investigated in the literature are the greedy algorithms and convex relaxation methods. The former seek to find an approximate solution of (2.17) by selecting a single atom at each step. Examples of greedy algorithms used for sparse coding include the matching pursuit (MP) [Mallat 1993] and the orthogonal matching pursuit (OMP) [Pati 1993] algorithms. Convex relaxation methods relax the ℓ_0-minimization problem to an approximate l_1-minimization. The problem (2.14) is thus relaxed, *i.e.*, changed as follows

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \boldsymbol{u} = \boldsymbol{D}\boldsymbol{\alpha}. \tag{2.18}$$

The problem (2.18) is convex and can be solved in polynomial time. A known example of a convex relaxation method is the least absolute shrinkage and selection operator (LASSO) [Tibshirani 1996]. The quality of the approximations provided by these algorithms does not only depend on the signal itself but also on the chosen dictionary [Tosic 2011]. This means that the choice of the dictionary is crucial when using sparse representations. In other words, the dictionary should be adapted to the signal of interest.

2.3.2 Dictionary learning

Sparse representations consist in expressing a signal as a linear combination of a few atoms of a dictionary. Numerous predefined dictionaries, which use off-the-shelf bases, exist in the literature, *e.g.*, based on wavelets, discrete cosine transforms (DCT), or Fourier decompositions. However, it has been shown that adaptive dictionaries, *i.e.*, learned from the data itself, can outperform the predefined ones [Mairal 2009]. The methods used to learn such data-driven dictionaries are called dictionary learning (DL) methods. Details about the DL problem formulation as well as the different algorithms and strategies used for learning are provided in the following.

2.3.2.1 Problem formulation

In the context of 2D signals, particularly in image analysis, a common approach is to learn the dictionary using small blocks or patches extracted from the image. This choice is preferred in practice due to the large dimension of natural images, but also because it allows patch-wise self-similarity to be exploited by capturing local meaningful information [Elad 2006]. In this thesis, we adopt a similar approach for learning a dictionary, *i.e*, using patches extracted from sequences of 2D motion fields. In the following, the DL problem is thus formulated according to the patch-wise approach.

Learning a dictionary $\boldsymbol{D} \in \mathbb{R}^{n \times q}$ consists of a joint optimization problem with respect to the dictionary $\boldsymbol{D}$ and the sparse coefficient vectors $\boldsymbol{\alpha}$. The dictionary is constructed using all overlapping patches of the data in $\boldsymbol{U} \in \mathbb{R}^N$ representing, for example, an image or a motion field. Following the sparse coding problem (2.17) and assuming that the noise $\boldsymbol{e} \in \mathbb{R}^n$ is Gaussian zero-mean (see [Tosic 2011, Michal Aharon 2006, Mairal 2008a] for motivations), the DL problem can be formulated as follows

$$\min_{\boldsymbol{D},\boldsymbol{\alpha}} \sum\nolimits_p \|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 \text{ subject to } \forall p,\ \|\boldsymbol{\alpha}_p\|_0 \leq K \tag{2.19}$$

where $\boldsymbol{P}_p \in \mathbb{R}^{n \times N}$ is a binary operator that extracts the pth patch of size n from $\boldsymbol{U}$ and $\boldsymbol{\alpha}_p$ is the corresponding sparse code, with K its maximum number of non-zero coefficients. The optimization problem (2.19) is usually solved by iterating between two steps. The first one is a sparse coding step, where the dictionary $\boldsymbol{D}$ is fixed and the optimization is performed with respect to the coefficients $\boldsymbol{\alpha}_p$. The second step is the dictionary update step (or learning step), where the sparse coefficient vectors $\boldsymbol{\alpha}_p$ are fixed and the minimization is conducted with respect to the dictionary $\boldsymbol{D}$.

2.3.2.2 Dictionary learning algorithms

Numerous algorithms designed for solving the DL problem (2.19) have been investigated in the literature. These algorithms include, for example, the method of optimal directions (MOD) [Engan 1999], K-SVD [Michal Aharon 2006] and the online DL (ODL) [Mairal 2009]. In MOD, the sparse coding problem can be solved using any one of above-

mentioned strategies, *e.g.*, OMP or LASSO, and the dictionary is updated at each step using the pseudo-inverse of the sparse representation. K-SVD solves the sparse coding step with OMP and the dictionary update is performed column-wise using a singular value decomposition (SVD). Note that K-SVD is one of the most popular algorithms, as it has been widely used for image denoising [Elad 2006]. Unlike K-SVD, the ODL algorithm has been proposed to cope with large or dynamically changing training sets. It uses LASSO for the sparse coding step and an alternate optimization scheme on gradually augmented subsets of the training data.

2.3.2.3 Learning strategies

The dictionary can be either fixed in advance, *i.e.*, learned offline from a set of training data, or learned in an adaptive way from the current estimation, *i.e.*, using an online scheme. More details about these two strategies are provided below.

Offline dictionary learning

Given a set of ground-truth data, the dictionary is learned and fixed before the estimation step. This strategy was used for image denoising in [Elad 2006], face recognition in [Zhang 2010] and texture segmentation in [Mairal 2008a]. The offline learning supposes that the atoms of the dictionary encode sufficiently well a particular kind of signal. This implies that the training set is adapted to the problem and is extensive enough to span the different types of signals that might occur during the estimation. Since the learning process is done only once, the offline strategy has the advantage of being less time consuming.

Adaptive dictionary learning

This strategy, also known as online learning, is commonly used in the area of natural image denoising [Dong 2011, Mairal 2008b]. Online learning consists in extracting training patches from the noisy data itself. More specifically, an initial dictionary $\boldsymbol{D}_0$ is chosen and updated at each estimation step based on the current value of the signal of interest. The outcome of each iteration serves as an initialization for the next one. Note that the initial dictionary can either be an offline learned dictionary, a predefined dictionary (*e.g.*, DCT or wavelet dictionary), or simply a set of random patches extracted from an initial estimate of the signal, *e.g.*, the result of a first rough estimation. In the latter case, the quality of the initial dictionary $\boldsymbol{D}_0$ depends on the performance of the employed estimation method. Since the dictionary is jointly estimated with the signal, *e.g.*, to be denoised, the adaptive learning strategy is more time consuming. However, it remains the most appropriate approach in the absence of a suitable or sufficiently rich training set. In Chapters 3, 4 and 5 of this thesis, the offline learning is employed to train a dictionary encoding typical patterns of cardiac motions. In Chapter 3, the adaptive learning strategy is also investigated.

2.4 Conclusions

The first section of this chapter overviewed the main cardiac motion estimation strategies, particularly those used in the context of ultrasound imaging. More specifically, three categories of dense motion estimation methods were presented, *i.e.*, block-matching (or speckle tracking), optical flow and registration-based methods. Common data fidelity and regularization terms employed for cardiac motion estimation and ultrasound imaging were also overviewed. The data fidelity terms included general formulations, *e.g.*, optical flow or the sum of squared differences, and similarity measures specifically designed for ultrasound imaging, *e.g.*, based on the statistical properties of the speckle noise. The principal regularization strategies enforced spatial or temporal smoothness, and were either explicit or implicit, *e.g.*, using parametric transformations. Other regularizations based on physiological characteristics of the heart were also discussed. Finally, quantitative indicators derived from the estimated cardiac motions, in particular the Lagrangian strain, were explained. In the second section, the principles of sparse representations and the algorithms used to solve the resulting sparse coding problem were presented. The problem of dictionary learning, as well as the related learning strategies, were also introduced.

Chapter 3

Motion Regularization based on a Sparse Representation

Contents

3.1 Introduction

The previous chapter overviewed the different categories of cardiac motion estimation methods, as well as their most common regularization strategies. These regularizations were principally based on geometrical assumptions, *e.g.*, spatial smoothness or physiological properties of the heart. As explained in Section 1.3, one major drawback of these conventional approaches is the loss of structural information due to, for example, over-smoothing. As a main contribution of this chapter, we investigate other types of regularizations for the problem of cardiac motion estimation. These regularizations are based on sparse representations, which have gained an increasing interest over the last years.

Sparse representations have been shown to be very well suited to natural images and have thus been used as an effective tool for several image processing tasks [Elad 2006, Dong 2011, Zhang 2010, Mairal 2008a, Mairal 2008b]. Motivated by this success, it is argued in [Shen 2010] that the sparsity assumption also holds for motion fields. In fact, motion fields can be seen as images with generally well structured and overall simpler patches than those associated with natural images. In [Shen 2010], the authors added a sparsity prior to an OF estimation problem and used the wavelet basis for the sparse coding step. This approach was also considered in [Jia 2011a], where the wavelet basis was replaced by a learned motion dictionary and a multi-resolution scheme was adopted. The results of this work tend to favor the use of motion models learned from a training set tuned to the application at hand, contrary to the strategy studied in [Shen 2010]. In the case of cardiac motion, sparsity still holds as the cyclic motion of the heart usually alternates between two major phases called diastole and systole (often modeled locally by simple parametric models such as affine models [**?**, Alessandrini 2013b]).

In this chapter, we promote the use of a regularization exploiting sparse representations and DL for cardiac motion estimation in 2D US images. The proposed method combines a specific similarity measure with spatial smoothness and sparse regularizations, exploiting jointly the statistical nature of B-mode images, the smoothness and the sparse properties of cardiac motion. More specifically, the data fidelity term is based on the multiplicative Rayleigh noise model [Cohen 2002b] and the spatial smoothness is ensured by a regularization based on the gradient of the motion vector. Moreover, we introduce a sparse regularization based on DL using patterns of cardiac motion. For the sparse coding step associated with motion estimation, the dictionaries are learned using the ground-truth displacements of realistic simulations.

This chapter is organized as follows. Sections 3.2 and 3.3 formulate the cardiac motion estimation problem and introduce the proposed estimation strategy based on a sparse regularization. Some further implementation details are provided in Section 3.4. Simulation results are presented and discussed in Section 3.5 . We compare the performance of the proposed method to results obtained with three state-of-the-art algorithms: (i) the conventional BM method [Kaluzynski 2001] using the NCC similarity measure, (ii) the monogenic signal method using an affine motion model [Alessandrini 2013b] and (iii) an elastic registration method with a similarity measure based on the Rayleigh noise

assumption and a B-spline interpolation [Myronenko 2009a]. Furthermore, strain curves resulting from healthy and pathological subjects are analysed using *in vivo* images. Concluding remarks are finally reported in Section 3.6.

3.2 Motion estimation using the ML method

3.2.1 Problem formulation

We consider the estimation of a 2D displacement field $\boldsymbol{U} = (\boldsymbol{u}^T, \boldsymbol{v}^T)^T \in \mathbb{R}^{2N}$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the horizontal and vertical displacement vectors, between a pair of consecutive frames $\boldsymbol{I}_k \in \mathbb{R}^N$ and $\boldsymbol{I}_{k+1} \in \mathbb{R}^N$. The ML approach is a common framework for the motion estimation problem [Cohen 2002b, Myronenko 2009a]. It allows us to incorporate knowledge about the image formation model, or the acquisition process, into the formulation of a similarity measure. According to the ML method, the estimation of the motion $\boldsymbol{U}$ between two consecutive images is achieved by maximizing the conditional probability of the observations in $\boldsymbol{I}_{k+1}$ given the previous image $\boldsymbol{I}_k$ and $\boldsymbol{U}$, *i.e.,*

$$\max_{\boldsymbol{U}} \; p(\boldsymbol{I}_{k+1}|\boldsymbol{I}_k, \boldsymbol{U}). \tag{3.1}$$

However, the problem (3.1) is usually reformulated in the negative log-domain, where the maximization of the likelihood term is equivalent to the following minimization problem

$$\min_{\boldsymbol{U}} \; -\ln[p(\boldsymbol{I}_{k+1}|\boldsymbol{I}_k, \boldsymbol{U})]. \tag{3.2}$$

The negative log-likelihood in (3.2) is the data fidelity term of our motion estimation problem, which expresses the similarity between the displaced image $\boldsymbol{I}_{k+1}$ and the reference image $\boldsymbol{I}_k$. Details about the observation model used to construct the likelihood, and thus about the data fidelity term, are provided in the following.

3.2.2 Observation model

US B-mode images are obtained by performing a sequence of transformations on the original RF data. More specifically, envelop detection and log-compression are used to obtain the final B-mode version of the images (see Section 1.2 for more details). Similarly, the observation model can be first expressed based on known properties of US envelop images, then derived for B-mode data using an appropriate transformation.

Envelop images

One of the most widely accepted noise models in US image processing is the Rayleigh multiplicative noise [Goodman 2007, Kotropoulos 1994]. This model is based on the

initial assumption of Gaussian RF signals, leading to a Rayleigh distributed signal envelope. Following the Rayleigh noise assumption, motion can be introduced for all pairs of consecutive frames in a given image sequence using the following observation model

$$\begin{cases} \boldsymbol{r}_k(n) &= \boldsymbol{a}_k(n)\boldsymbol{s}_k(n) \\ \boldsymbol{r}_{k+1}(n) &= \boldsymbol{a}_{k+1}(n)\boldsymbol{s}_{k+1}(n-\boldsymbol{u}_k(n)) \end{cases} \tag{3.3}$$

for $k = 1, ..., M-1$ and $n = 1, ..., N$, where M is the number of frames, N is the image size and

- $\boldsymbol{r}_k(n)$ and $\boldsymbol{r}_{k+1}(n)$ are the envelop image amplitudes at pixel n for the frames k and $k+1$,
- $\boldsymbol{s}_k(n)$ and $\boldsymbol{s}_{k+1}(n)$ are the unknown noise-free signals
- $\boldsymbol{a}_k$ and $\boldsymbol{a}_{k+1}(n)$ are the Rayleigh multiplicative noises corrupting the nth pixel of the frames k and $k+1$
- $\boldsymbol{u}_k(n)$ is the unknown displacement value at pixel n between the frames k and $k+1$.

Note that, for conciseness, the above obsevation model (3.3) is written for one motion component only (*i.e.*, horizontal), the extension to the 2D case being straightforward. Note also that in the above formulation, motion estimation is viewed as a passive time delay estimation problem [Byram 2013], in which two consecutive frames are considered as two realizations of the same original signal. The sampling of the continuous formulation yields the discrete version used in (3.3).

B-mode images

In the case of B-mode images, the pixel intensities can be expressed as the log-compressed values of the envelope image, *i.e.,*

$$\begin{cases} \boldsymbol{I}_k &= b\log(\boldsymbol{r}_k) + g \\ \boldsymbol{I}_{k+1} &= b\log(\boldsymbol{r}_{k+1}) + g \end{cases} \tag{3.4}$$

where b and g stand for the dynamic range and linear gain constants [Myronenko 2009a]. Using (3.4), the observation model for the B-mode images can be formulated as follows

$$\begin{cases} \boldsymbol{I}_k(n) &= b[\boldsymbol{a}_{b,k}(n) + \boldsymbol{s}_b(n)] + g \\ \boldsymbol{I}_{k+1}(n) &= b[\boldsymbol{a}_{b,k+1}(n) + \boldsymbol{s}_b(n-\boldsymbol{u}_k(n))] + g \end{cases} \tag{3.5}$$

where $\boldsymbol{a}_{b,k}(n) = \log[\boldsymbol{a}_k(n)]$ and $\boldsymbol{s}_b(n) = \log[\boldsymbol{s}(n)]$. In the following, the observation model (3.5) is used to construct the data fidelity term.

3.2.3 Data fidelity term

The data fidelity term measures the similarity between two consecutive frames and should be designed according to the nature of the images. In the case of UI, classical intensity-based similarity measures, such as the SSD, CC, or NCC, suffer from the presence of speckle noise [Viola 2003]. This is due to the fact that these measures were built with the underlying assumption of Gaussian noise. One solution to this problem is to consider more appropriate noise models, such as the Rayleigh noise model introduced in 3.2.2. The probability density function (pdf) of a Rayleigh distributed noise a is defined as

$$p(a) = \frac{a}{\sigma^2} \exp\left(\frac{-a^2}{2\sigma^2}\right) \mathbb{1}_{\mathbb{R}^+}(a) \tag{3.6}$$

where $\sigma \in \mathbb{R}^+$ is a scale parameter and $\mathbb{1}_{\mathbb{R}^+}(.)$ denotes the indicator function on $\mathbb{R}^+$.

Using the observation model (3.5), a likelihood term specific to log-compressed B-mode US images was developed in [Cohen 2002b]. More specifically, one can write (for one component of motion $\boldsymbol{u}_k$)

$$\begin{aligned} \boldsymbol{I}_{k+1}(n+\boldsymbol{u}_k(n)) &= \boldsymbol{I}_k(n) + b[\boldsymbol{a}_{b,k+1}(n+\boldsymbol{u}_k(n)) - \boldsymbol{a}_{b,k}(n)] \\ &= \boldsymbol{I}_k(n) + b\log(\eta_k) \end{aligned} \tag{3.7}$$

where $\eta_k = \boldsymbol{a}_{k+1}(n+\boldsymbol{u}_k(n))/\boldsymbol{a}_k(n)$ is the ratio between the multiplicative Rayleigh noises. The likelihood term can then be formulated using (3.6) and (3.7) (see Appendix A for more details) leading to

$$p[\boldsymbol{I}_{k+1}(n+\boldsymbol{u}_k(n))|\boldsymbol{I}_k(n), \boldsymbol{u}_k(n)] = \frac{2\sigma^4\eta_k^2}{b(\eta_k^2+1)^2} \tag{3.8}$$

Finally, after applying the negative log-transform, the data fidelity term corresponding to the minimization problem (3.2) can be obtained for the displacement vector $\boldsymbol{u}_k$. Straightforward computations lead to the following data fidelity term

$$E_{\mathrm{CD}_2}(\boldsymbol{u}) = -2d_k(\boldsymbol{u}) + 2\log[e^{2d_k(\boldsymbol{u})}+1] + \mathrm{cst} \tag{3.9}$$

where $d_k(\boldsymbol{u}) = \frac{1}{b}\sum_{n=1}^{N}[\boldsymbol{I}_{b,k+1}(n+\boldsymbol{u}_k(n)) - \boldsymbol{I}_{b,k}(n)]$ and $\mathrm{cst} = -\log\left(2\sigma^4/b\right)$ is a constant.

Other similarity measures

In this work, we consider the simple case of mutually uncorrelated Rayleigh noises with equal variances, which corresponds to the CD_2 similarity measure proposed in [Cohen 2002b]. However, the use of more sophisticated metrics that consider the temporal correlation between image speckles would be also possible. Note that this case has been addressed in [Myronenko 2009a], where a similarity measure MS_2 has been pro-

posed as a more realistic extension of CD_2. Note also that different noise models, *e.g.*, Rician, Nakagami and generalized gamma, or other similarity measures, *e.g.*, based on OF (see Section 2.2.1.2), could be employed in the proposed motion estimation framework. However, the choice of the Rayleigh noise model is motivated by the fact that it is widely accepted in US image processing [Goodman 2007, Kotropoulos 1994]

3.3 Sparse and spatial regularizations

A standard approach used to overcome the ill-posed nature of the motion estimation problem consists in introducing assumptions about the spatial behaviour of the motion. However, the reliance on such global and purely geometrical priors, which do not hold in many cases, such as motion boundaries, makes it inadequate for the estimation of complex or multiple motions. In the following, a sparsity-based prior is introduced in order to bypass these shortcomings. As seen in Section 2.2.3, this *a priori* information can be incorporated explicitly into the motion estimation problem through an appropriate energy term also referred to as the regularization term. More precisely, the proposed energy term combines two different types of regularizations. The first one expresses the sparsity of the motion field when decomposed on a dictionary of typical cardiac motions, while the second one exploits a more traditional spatial coherence. The motion estimation problem is thus formulated as the minimization of an appropriate energy function defined as

$$E(\boldsymbol{U},\boldsymbol{I}) = E_{\mathrm{CD}_2}(\boldsymbol{U},\boldsymbol{I}) + \lambda_{\mathrm{P}} E_{\mathrm{P}}(\boldsymbol{U}) + \lambda_{\mathrm{TV}} E_{\mathrm{TV}}(\boldsymbol{U}) \tag{3.10}$$

where E_{CD_2} was defined in (3.9), E_{P} is the sparsity-based regularization term, E_{TV} stands for the spatial constraint promoting the smoothness of the motion and $(\lambda_{\mathrm{P}}, \lambda_{\mathrm{TV}}) \in \mathbb{R}^2$ are two regularization parameters that balance the effects of the data fidelity and regularization terms.

3.3.1 Sparse regularization based on DL

Recent advances have made possible to simulate realistic cardiac US image sequences with ground-truth [Alessandrini 2016b], enabling the use of learning-based methods for cardiac motion estimation. In this context, we propose to exploit the sparsity of the motion field using a learned dictionary that captures typical patterns of myocardial motion.

As explained in Section 2.3, the dictionary $\boldsymbol{D}$ is overcomplete, leading to a sparse representation of the motion when decomposed on $\boldsymbol{D}$. In this work, the resulting sparse coding problem is performed separately for the horizontal and vertical motion components $\boldsymbol{u}$ and $\boldsymbol{v}$. We also use separate dictionaries $\boldsymbol{D}_u \in \mathbb{R}^{n\times q}$ and $\boldsymbol{D}_v \in \mathbb{R}^{n\times q}$ associated with the horizontal and vertical motion components $\boldsymbol{u}$ and $\boldsymbol{v}$ as in [Jia 2011b], where n is the patch size and q is the number of atoms in the dictionaries $\boldsymbol{D}_u$ and $\boldsymbol{D}_v$. We exploit the sparse properties of the motion by extracting overlapping patches from the global motion field $\boldsymbol{U}$. This patch-wise approach is motivated by the fact that it allows

meaningful local cardiac motion patterns to be captured. Each motion patch is then expressed as a weighted linear combination of a few elements of the dictionary $\boldsymbol{D}$, *i.e.*,

$$\boldsymbol{P}_p\boldsymbol{U} \approx \boldsymbol{D}\boldsymbol{\alpha}_p, \forall p \tag{3.11}$$

where $p = 1, ..., N_p$ with N_p the total number of patches and

- $\boldsymbol{\alpha} \in \mathbb{R}^{2q \times N_p}$ is a matrix whose columns are the sparse coding vectors $\boldsymbol{\alpha}_p = (\boldsymbol{\alpha}_{u,p}^T, \boldsymbol{\alpha}_{v,p}^T)^T$ (containing only a few non-zero elements),
- $\boldsymbol{D} \in \mathbb{R}^{2n \times 2q}$ is a block diagonal matrix whose blocks are $\boldsymbol{D}_u$ and $\boldsymbol{D}_v$, *i.e.*,

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_u & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_v \end{bmatrix}$$

- $\boldsymbol{P}_p \in \mathbb{R}^{2n \times 2N}$ is an operator that extracts the pth patch in the horizontal and vertical directions from $\boldsymbol{U}$.

The proposed patch-wise sparse regularization term E_{P} is constructed using the ℓ_2-norm as follows

$$E_{\mathrm{P}}(\boldsymbol{U}, \boldsymbol{\alpha}) = \sum_{p=1}^{N_p} \|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2. \tag{3.12}$$

The sparse regularization term (3.12) constrains each motion patch $\boldsymbol{P}_p\boldsymbol{U}$ to be sparsely represented in the learned dictionary $\boldsymbol{D}$ of typical motion patterns. In contrast with the classical spatial smoothness assumptions, the learned motion patterns encode more complex and general behaviors of the cardiac motion (including motion discontinuities inside the myocardium), leading to a spatially more flexible and cardiac motion-specific prior. This property will be outlined in Section 3.5 devoted to simulation results. Note that the use of overlapping patches in (3.12) introduces an implicit inter-patch regularization in accordance with the *expected patch log-likelihood* framework [Sulam 2015]. This regularization is due to the fact that a single pixel is counted multiple times.

3.3.2 Spatial regularization

The sparse regularization term (3.12) can result in patch artefacts and does not guarantee smoothness inside each patch of motion. In order to overcome these drawbacks, we propose to make use of a spatial regularization enforcing a spatially smooth motion field. For this purpose, we use a regularization based on the ℓ_2-norm, which imposes weak spatial gradients on the two motion components, thereby insuring a smooth variation of the motion field. This is a common choice that leads to the following spatial regularization term [Horn 1981]

$$E_{\mathrm{TV}}(\boldsymbol{U}) = \|\nabla\boldsymbol{U}\|_2^2. \tag{3.13}$$

The term (3.13) is also referred to as the ℓ_2-TV and has the advantage of being differentiable.

Finally, after combining the spatial constraint in (3.13) with the data fidelity term (3.9) and the sparse regularization term (3.12), the global energy function is

$$E(\boldsymbol{U}, \boldsymbol{I}, \boldsymbol{\alpha}) = E_{\mathrm{CD}_2}(\boldsymbol{U}, \boldsymbol{I}) + \lambda_{\mathrm{P}} \sum\nolimits_p \|\boldsymbol{P}_p \boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 + \lambda_{\mathrm{TV}} \|\nabla \boldsymbol{U}\|_2^2. \tag{3.14}$$

The combination of (3.9), (3.13) and (3.12) results in an original motion estimation problem exploiting jointly the statistical properties of the speckle noise, using a Rayleigh noise model, and the smooth and sparse properties of cardiac motion.

3.4 Motion estimation

In order to exploit the sparse properties of cardiac motion, the dictionary has to be learnt either offline as a first step, or in an adaptive way during the motion estimation. In the former case, the motions are estimated in a separate step by solving the minimization problem (3.14). Conversely, when using the adaptive dictionary leaning strategy, the dictionary is constructed based on the estimated motions themselves. These different steps are described in the following.

3.4.1 Offline dictionary learning

The proposed method requires a training step during which the dictionary $\boldsymbol{D}$ is learnt using a set of ground-truth cardiac motion fields denoted as $\boldsymbol{U_t}$. The dictionary $\boldsymbol{D}$ is learned from patches of the training set $\boldsymbol{U_t}$ by solving

$$\min_{\boldsymbol{D}, \boldsymbol{\alpha}} \sum\nolimits_p \|\boldsymbol{P}_p \boldsymbol{U_t} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 \tag{3.15}$$

$$\text{subject to } \|\boldsymbol{\alpha}_{u,p}\|_0 \leq K \text{ and } \|\boldsymbol{\alpha}_{v,p}\|_0 \leq K, \forall p$$

In this work, the DL problem (3.15) is solved using the ODL algorithm with OMP for the sparse coding step. This choice is motivated by the fact that it has empirically been shown to give more accurate estimation results than the K-SVD algorithm and to be more efficient from a computational point of view [Wei 2015]. Once the dictionary $\boldsymbol{D}$ has been learned, it is fixed and used for the motion estimation process described below.

3.4.2 Motion field estimation

Using the data fidelity and regularization terms detailed in the previous Sections 3.2 and 3.3, the cardiac motion estimation can be formulated as the following optimization problem

$$\min_{\boldsymbol{\alpha}, \boldsymbol{U}} \left\{ E_{\mathrm{CD}_2}(\boldsymbol{U}) + \lambda_{\mathrm{P}} \sum\nolimits_p \|\boldsymbol{P}_p \boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 + \lambda_{\mathrm{TV}} \|\nabla \boldsymbol{U}\|_2^2 \right\} \tag{3.16}$$

$$\text{subject to } \|\boldsymbol{\alpha}_{u,p}\|_0 \leq K \text{ and } \|\boldsymbol{\alpha}_{v,p}\|_0 \leq K, \forall p$$

where E_{CD_2} has been introduced in (3.9) and the dictionary has been determined using the method described in Section 3.4.1. Since (3.16) is hard to solve directly, we adopt an alternate minimization scheme, similar to the half quadratic splitting strategy employed for motion estimation in [Jia 2011a] and denoising in [Sulam 2015]. More precisely, for fixed values of λ_{P} and λ_{TV}, we alternate optimizations with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{U}$. This process is repeated during a few iterations (typically 4 or 5 [Sulam 2015]) after which the sparsity parameter λ_{P} is increased. Note that when increasing λ_{P}, more importance is attached to the distance $\|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2$, which forces the estimated patches to be close to the atoms of the dictionary. More details about these two steps are provided below.

1. **Sparse coding**
   For fixed values of $\boldsymbol{U}$ and $\boldsymbol{D}$, the sparse coding problem is solved using the OMP algorithm. For all patches, the corresponding sparse vectors are found by solving

$$\min_{\boldsymbol{\alpha}} \sum\nolimits_p \|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 \tag{3.17}$$

$$\text{subject to } \|\boldsymbol{\alpha}_{u,p}\|_0 \leq K \text{ and } \|\boldsymbol{\alpha}_{v,p}\|_0 \leq K, \forall p$$

   where $p = 1, ..., N_p$, with N_p the total number of patches. Recall that the parameter K indicates the maximum number of non-zero coefficients of $\boldsymbol{\alpha}_p$.

2. **Motion field estimation**
   Once the sparse codes and the dictionary have been determined, the motion field $\boldsymbol{U}$ is updated (starting from a first initialization, *e.g.*, $\boldsymbol{U}_0 = 0$) by solving the following minimization problem

$$\min_{\boldsymbol{U}} \left\{ E_{\mathrm{CD}_2}(\boldsymbol{U}) + \lambda_{\mathrm{P}} \sum\nolimits_p \|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 + \lambda_{\mathrm{TV}} \|\nabla \boldsymbol{U}\|_2^2 \right\} \tag{3.18}$$

   where $E_{\mathrm{CD}_2}(\boldsymbol{U})$ is given in (3.9). The solution to (3.18) can be found by equating the gradient to zero, leading to

$$\nabla E_{\mathrm{CD}_2}(\boldsymbol{U}) + \lambda_{\mathrm{P}} \nabla E_{\mathrm{P}}(\boldsymbol{U}) + \lambda_{\mathrm{TV}} \nabla E_{\mathrm{TV}}(\boldsymbol{U}) = 0. \tag{3.19}$$

   Following the optimization approach studied in [Myronenko 2009a], we use the implicit Euler time marching method to solve (3.19). More specifically, at each iteration $i+1$, the displacement $\boldsymbol{U}_{i+1}$ is estimated as follows

$$\boldsymbol{U}_{i+1} = (Id + 2\gamma\lambda_{\mathrm{TV}}\Delta)^{-1}[\boldsymbol{U}_i - \gamma(\nabla E_{\mathrm{CD}_2}(\boldsymbol{U}_i) + \lambda_{\mathrm{P}} \nabla E_{\mathrm{P}}(\boldsymbol{U}_i))] \tag{3.20}$$

   where $\gamma \in \mathbb{R}$ is a stepsize parameter and Δ denotes the Laplacian operator. Note that the gradient of the sparse regularization term is easy to compute and can be

expressed as

$$\nabla E_{\mathrm{P}}(\boldsymbol{U}) \propto \sum\nolimits_p \boldsymbol{P}_p^T \boldsymbol{P}_p \boldsymbol{U} - \sum\nolimits_p \boldsymbol{P}_p^T \boldsymbol{D} \boldsymbol{\alpha}_p \tag{3.21}$$

where the first term $\sum_p \boldsymbol{P}_p^T \boldsymbol{P}_p$ is a constant factor that counts the number of times each pixel is considered. The second term $\sum_p \boldsymbol{P}_p^T \boldsymbol{D} \boldsymbol{\alpha}_p$ represents the reconstruction of the motion field from the current sparse representation.

3.4.3 Adaptive dictionary learning

In the absence of a training set, or when the latter is not sufficiently rich, it might be advantageous to use an adaptive dictionary learning strategy. When the dictionary needs to be updated, an offline initialization of the dictionary denoted as $\boldsymbol{D_0}$ is required. Instead of performing a simple sparse coding step, the dictionary $\boldsymbol{D}$ is then updated at each iteration from patches of the current estimate of the motion field $\boldsymbol{U}$ using the ODL algorithm. The motion estimation problem with adaptive DL is formulated as the following optimization problem

$$\min_{\boldsymbol{D},\boldsymbol{\alpha},\boldsymbol{U}} \left\{ E_{\mathrm{CD}_2}(\boldsymbol{U}) + \lambda_{\mathrm{P}} \sum\nolimits_p \|\boldsymbol{P}_p \boldsymbol{U} - \boldsymbol{D} \boldsymbol{\alpha}_p\|_2^2 + \lambda_{\mathrm{TV}} \|\nabla \boldsymbol{U}\|_2^2 \right\}$$

$$\text{subject to } \ \|\boldsymbol{\alpha}_{u,p}\|_0 \leq K \text{ and } \|\boldsymbol{\alpha}_{v,p}\|_0 \leq K, \forall p \tag{3.22}$$

A full description of the sparse coding, dictionary update and motion estimation steps is provided in Algorithm 1.

3.5 Experimental Results

In order to evaluate the performance of the proposed method, experiments are first conducted using two datasets of synthetic and realistic simulations for which ground-truth measurements are available. Three sequences of *in vivo* US B-mode images, representing healthy and diseased patients, are then considered.

3.5.1 State-of-the-art methods

The proposed approach is compared with three state-of-the-art motion estimation methods. These three approaches have been presented in Sections 2.2.1.1, 2.2.1.3 and 2.2.2.2. In the following, we provide more details about the choice of the data fidelity term and the regularization strategy for each method.

- BM: for each patch, a full-grid search is conducted in a defined search window using the NCC similarity measure [Kaluzynski 2001]. This method does not use an explicit spatial regularization term. However, a spatial regularization is induced in post-processing by the cubic interpolation used to obtain the final sub-pixel displacements.

Algorithm 1: Motion field estimation for a pair of images using DL.

Input : $\boldsymbol{I}$, K, λ_{TV},$\boldsymbol{\lambda}_{\mathrm{P}}$, OuterSteps, InnerSteps

Initialization: $\boldsymbol{U}_0 = 0$, $\boldsymbol{D}_0 =$ offline dictionary

$\boldsymbol{U} = \boldsymbol{U_0}$;

for $i = 1$ **to** *OuterSteps* **do**

  for $j = 1$ **to** *InnerSteps* **do**

    `%Dictionary update (optional step)`

    $\boldsymbol{D}_j \leftarrow \mathrm{ODL}(\boldsymbol{U},\boldsymbol{D}_{j-1})$;

    `%If not updated:`

    $\boldsymbol{D}_j = \boldsymbol{D}_0$;

    `%Sparse coding`

    $\boldsymbol{\alpha}_j \leftarrow \mathrm{OMP}(\boldsymbol{D}_j,\boldsymbol{U},K)$;

    `%Motion estimation`

    $\boldsymbol{U} \leftarrow \min_{\boldsymbol{U}} E_{\mathrm{CD}_2}(\boldsymbol{I},\boldsymbol{U}) + \lambda_{\mathrm{TV}}\|\nabla \boldsymbol{U}\|_2^2 + \boldsymbol{\lambda}_{\mathrm{P}}(i)\sum_p \|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}_j\boldsymbol{\alpha}_{p,j}\|_2^2$;

  end

end

Output : The motion field $\boldsymbol{U}$ and the corresponding sparse codes $\boldsymbol{\alpha}$.

- B-spline: in order to evaluate the interest of the sparsity-based regularization term, we consider the B-spline-based method studied in [Myronenko 2009a]. This algorithm uses the same similarity measure (CD_2) and spatial regularization as the proposed method. Note however that an FFD model [Sederberg 1986, Rueckert 1999b] is considered to parametrize the motion. The displacements are finally estimated for a mesh of B-spline control points.

- Monogenic signal: this method uses the monogenic phase in order to construct the similarity measure and considers a local affine motion model, without any additional spatial regularization. It corresponds to the method of [**?**] for which the intensity-based similarity measure has been replaced by a spatial phase-based metric. Note that the monogenic signal method of [Alessandrini 2013b] has been shown to improve motion estimation with respect to [**?**].

Note that in order to cope with large displacements, a coarse-to-fine estimation scheme was employed for the monogenic signal and B-spline methods. However, this multi-resolution scheme was not used for the proposed and BM methods.

3.5.2 Performance Measures

Endpoint error

The first performance measure is the endpoint error [Otte 1994, Alessandrini 2013b] that can be used for datasets with available ground-truth, *i.e.*, for the synthetic and realistic simulations. For each pixel $n = 1, ..., N$, the endpoint error is the ℓ_2 point distance between the estimates and the ground-truth and is defined as

$$\epsilon(n) = \sqrt{[\boldsymbol{u}(n) - \hat{\boldsymbol{u}}(n)]^2 + [\boldsymbol{v}(n) - \hat{\boldsymbol{v}}(n)]^2} \tag{3.23}$$

where $\boldsymbol{u}(n), \boldsymbol{v}(n)$ and $\hat{\boldsymbol{u}}(n), \hat{\boldsymbol{v}}(n)$ are the true and estimated horizontal and vertical displacements at pixel n.

Strain

Strain measurements are used to describe the deformation of the myocardium with respect to its original shape, allowing both quantitative and qualitative evaluations of motion estimates. When a ground-truth is not available, *i.e.*, for *in vivo* data, the strain is used for qualitative assessment only. Following the method described in Section 2.2.4, we compute the radial and longitudinal (respectively circumferential) strains for the apical four-chambers and short-axis views. The myocardium is then automatically segmented into 6 regions of interest for which an average strain value is computed [Alessandrini 2016b]. Note that the segmentation is not used for motion estimation and that its impact on the strain values is the same for all the presented methods. Fig. 3.1

displays the initial points and segmentation of the myocardium for an example of simulated data (see Section 3.5.5).

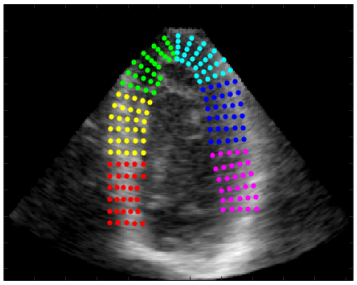


Figure 3.1: Initial points used for strain computation for the LADprox sequence considered in Section 3.5.5.

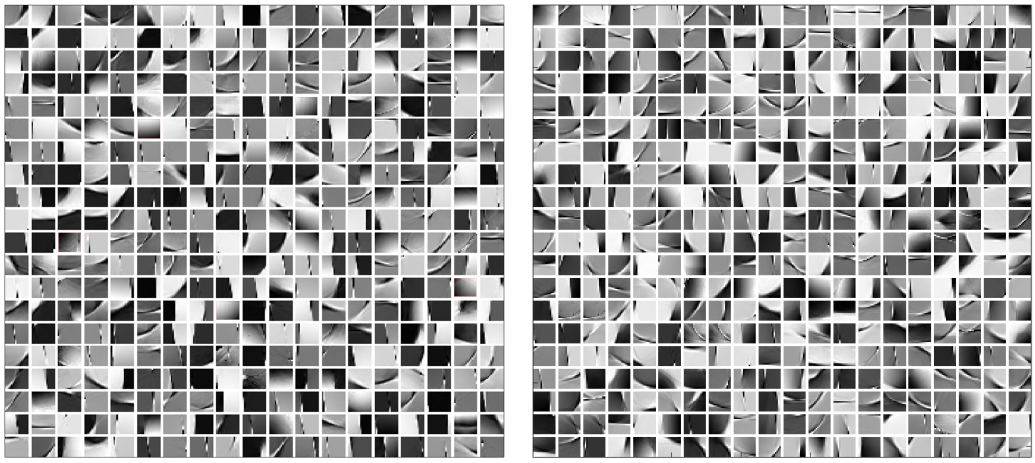
3.5.3 Dictionary learning and regularization parameters

The parameters used for DL and for regularizing the proposed motion estimation method are detailed in this section. The parameters used for the three state-of-the-art algorithms are also provided for each dataset.

Dictionary learning parameters

The DL parameters were selected empirically and fixed for all tests. For simplicity, the horizontal and vertical motion dictionaries $\boldsymbol{D_u}$ and $\boldsymbol{D_v}$ were learned separately on patches of size $w = 16 \times 16$ with 1.5 redundancy. The number of atoms was set to $n_a = 384$, leading to dictionaries of size $\boldsymbol{D_{u,v}} \in \mathbb{R}^{256\times384}$. Different patch sizes classically used in the literature [Michal Aharon 2006, Tosic 2011] were considered, *i.e.*, 8×8, 16×16 and 32×32. Since the errors for the three patch sizes had the same order of magnitude, the intermediate size 16×16 was selected to obtain a compromise between the number of patches and the size of the dictionary. Finally, the maximum number of non-zero coefficients used to represent one patch was fixed to $K = 5$ by cross-validation. Generally, the sparsity parameter K is much smaller than the number of atoms in the dictionary $K \ll n_a$ [Wei 2015] and is related to the noise level [Elad 2006]. The resulting horizontal and vertical dictionaries were finally used as offline dictionaries. Note that these estimated dictionaries could be used to initialize the adaptive DL strategy.

For all tests, the motion dictionaries were learnt using a realistic simulation sequence with available ground-truth. This sequence, referred to as LADdist, contains realistic motion fields generated according to [Alessandrini 2016b]. Note that this sequence represents a pathological case (see Section 3.5.5 for more details). This choice was motivated by the fact that cardiac diseases are typically localized, allowing the presence of normal and diseased motion patches in the training set. Fig. 3.2 shows the resulting atoms for the horizontal and vertical dictionaries $\boldsymbol{D_u}$ and $\boldsymbol{D_v}$ with a number of patches used for



$\boldsymbol{D}_u$ $\boldsymbol{D}_v$

Figure 3.2: Atoms of the horizontal and vertical dictionaries learned from the displacements of the LADdist sequence (Section 3.5.5) with $w = 16 \times 16$ and $K = 5$.

learning equal to $N_t = 517225$. This figure allows us to appreciate how the dictionaries capture the spatial properties of motion. More specifically, we can observe the presence of atoms containing two distinct regions separated by clear edges. These types of motion patches are well suited to the abrupt changes frequently observed in cardiac displacements, which usually occur in motion boundaries. Other atoms illustrate more or less varying motions that belong to the same region.

Regularization parameters

The first-order spatial regularization parameter λ_{TV} was manually varied such that $0.01 \leq \lambda_{\text{TV}} \leq 10$. The optimal value of λ_{TV} was selected as the value providing the smallest average error for the sequences with available ground-truth. Conversely, the value of λ_{TV} was chosen based on the visualization of the resulting motions for the *in vivo* data. The values of this regularization parameter are reported in Table 3.1 for all datasets. At this point, it is worth mentioning that it would be interesting to consider approaches based on Bayesian inference [Pereyra 2016] or on the Stein's unbiased risk estimate (SURE) [Ramani 2008] to estimate this regularization parameter. However, this automatic choice of the regularization hyperparameter from the observed data was not considered in this thesis. For each outer iteration of the proposed method, the sparse regularization parameter λ_{P} was logarithmically increased from 10^{-3} to 10^{2} (see Section 3.4) in 6 iterations [Sulam 2015].

The same procedure was applied for the parameters of the three state-of-the-art methods using the simulated sequences, *i.e*, the parameter values returning the smallest average error were selected for all the experiments. For the B-spline algorithm, the mesh window size between the B-spline control points was $w_{\text{B-spline}} = 15 \times 15$ and the

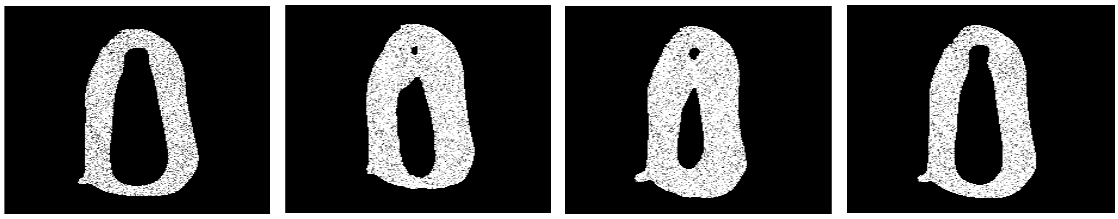
Dataset	Synthetic	LADprox	Normal	Rca	sync	LCX	LBBB	LBBBsmall	In vivo
λ_{TV}	0.75	0.75	1	0.25	0.1	0.1	0.1	0.1	0.1

Table 3.1: Values of λ_{TV} (spatial regularization) used for the synthetic, realistic simulations and *in vivo* sequences (see Sections 3.5.4, 3.5.5 and 3.5.6).

regularization parameter was set to $\lambda_{\text{B-spline}} = 3$ to avoid too much deformation. For the monogenic signal method, the initial wavelengths were set to $\lambda_0 = 0.25$ for the LADprox, Normal, LBBB, LBBBsmall and sync sequences and to $\lambda_0 = 0.1$ for the Rca and LCX sequences. The number of refinement steps was fixed to 4. The window size for the BM algorithm was set to $w_{\text{BM}} = 16 \times 16$. The parameters for the *in vivo* sequences were manually tuned to give the best visual aspect. The final parameters were $\lambda_{\text{B-spline}} = 0.1$ for the spatial regularization of the B-spline method whereas $\lambda_0 = 0.5$ for the initial wavelength of the monogenic signal algorithm. The block-size was fixed to $w_{\text{BM}} = 32 \times 32$ for the BM method. For all other parameters, we used the values from the tests conducted on the realistic simulations dataset.

3.5.4 Synthetic Data

In order to validate the data fidelity term, the proposed regularizations, and to show the behavior of the proposed method in an ideal setting, tests were first conducted on a sequence of synthetic images. The images were generated by moving an initial myocardium mask according to a set of ground-truth displacements. The mask and the known motions were taken from the LADprox sequence investigated in [Alessandrini 2016b, Alessandrini 2015] (see Section 3.5.5 for details). In order to consider an ideal scenario, the speckle noise was generated according to a Rayleigh distribution. Examples of images extracted from the synthetic sequence are displayed in Fig. 3.3.



(a) (b) (c) (d)

Figure 3.3: Myocardium area of the synthetic sequence: (a) Frame 1, (b) Frame 10, (c) Frame 22 and (d) Frame 33.

The average means and standard deviations (stds) of the endpoint errors are reported in Table 3.2. The evolution of the mean endpoint error along the sequence is also shown in Fig. 3.4. The proposed method provides the smallest estimation errors, while the second best algorithm is the B-spline method that uses the same similarity measure

CD_2. These results show the good performance of the proposed method and confirm that the similarity measure used for the data term (see Section 3.2.3) is adapted to this ideal scenario. The execution times associated with the different methods are also reported in Table 3.2. The proposed method is clearly computationally intensive with respect to the state-of-the-art methods. However, the current Matlab implementation could be optimized using, *e.g.*, a C implementation. Furthermore, a parallel computing strategy could be also considered for the horizontal and vertical dictionaries for the sparse coding step and for the computation of the sparse regularization term as well as in the learning phase. Note finally that real time methods are not always required in US imaging applications, *e.g.*, for computer-aided diagnosis.

	Proposed	Block-matching	B-spline	Monogenic
Error	**0.075±0.0671**	0.384±0.222	0.0958±0.0809	0.110±0.075
Time (s)	51.63	6.08	3.27	**0.37**

Table 3.2: Average means and stds of endpoint errors and average execution times per pair of images for the synthetic sequence.

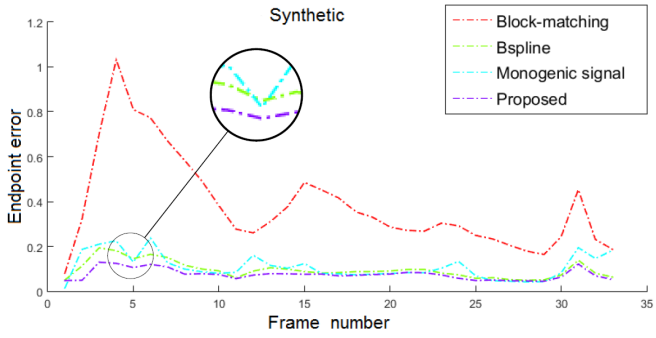


Figure 3.4: Mean endpoint errors for the synthetic sequence.

3.5.5 Realistic Simulations

This section evaluates the performance of the proposed method using a set of realistic simulations of B-mode US data [Alessandrini 2016b]. The images and motions were generated using real data combined with synthetic biomechanical and US models. More precisely, the images were generated by varying the parameters of a highly realistic E/M model, resulting in 8 sequences of different pathophysiological conditions. To the best of our knowledge, it is the most realistic simulation method available in the recent literature. The reader is invited to consult [Alessandrini 2016b, Alessandrini 2015] for more details about the data generation process. Note that a ground-truth was available for all measurements. Tests were conducted on sequences of 3D images (of size 224 $\times$ 176 $\times$ 208 voxel^3, with a voxel size 0.7 $\times$ 0.9 $\times$ 0.6 mm^3, a frame rate of 21-23 Hz [Alessandrini 2016b]) from which we extracted 2D slices (the middle slice) for apical four

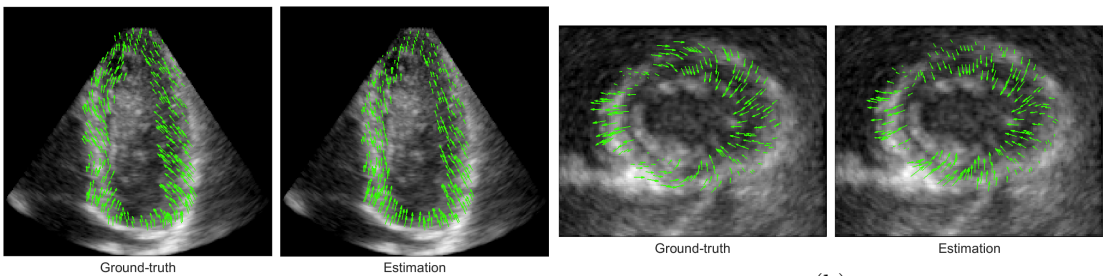
chambers and short-axis views. The sequences contained 34 or 40 images that span a full cardiac cycle and represent either healthy heart motions (*i.e.*, Normal), ischemic cases (*i.e.*, LADdist, LADprox, Rca and LCX) or dilated cardiomiopathy cases (*i.e.*, sync, LBBB and LBBBsmall) [Alessandrini 2016b][1]. Table 3.3 provides more details about the locations and characteristics of the pathologies related to each sequence.

Sequence	Pathology	Characteristics
Normal	none	healthy
LADdist	ischemia	distal occlusions of the left anterior descending artery
LADprox	ischemia	proximal occlusions of the left anterior descending artery
Rca	ischemia	occlusions of the right circumflex coronary artery
LCX	ischemia	occlusions of the left circumflex coronary artery
Sync	dilated cardiomiopathy	synchronous activation patterns
LBBB	dilated cardiomiopathy	dyssynchronous activation patterns
LBBBsmall	dilated cardiomiopathy	dyssynchronous activation patterns

Table 3.3: Description of the pathologies related to the sequences in the realistic simulations dataset [Alessandrini 2016b].

The true displacements of the ischemic sequence LADdist were used to learn the dictionaries. This choice allowed us to evaluate the method for different scenarios, *i.e.*, when the dictionary contains patterns of similar or different pathologies when compared to the test sequence. More specifically, the motion estimation accuracy was evaluated using the remaining sequences containing (i) one healthy sequence, (ii) one sequence with a pathology that is similar to the training sequence and (iii) five sequences with distinct pathologies.

Fig. 3.5 displays a typical example of estimated motion fields using the proposed method compared to the corresponding true displacement meshes. The estimated motion field is clearly qualitatively consistent with the ground-truth.



(a) (b)

Figure 3.5: Ground-truth and estimated meshes of the 5th frame of the LADprox sequence: (a) apical four chambers view, (b) short axis view.

[1]The data and related papers can be found at https://team.inria.fr/asclepios/data/straus/.

3.5.5.1 Sparse and spatial regularizations

In this section, we analyze the effects of the spatial and sparse regularization terms when used separately and combined as in (3.16). In order to have a visual interpretation of the

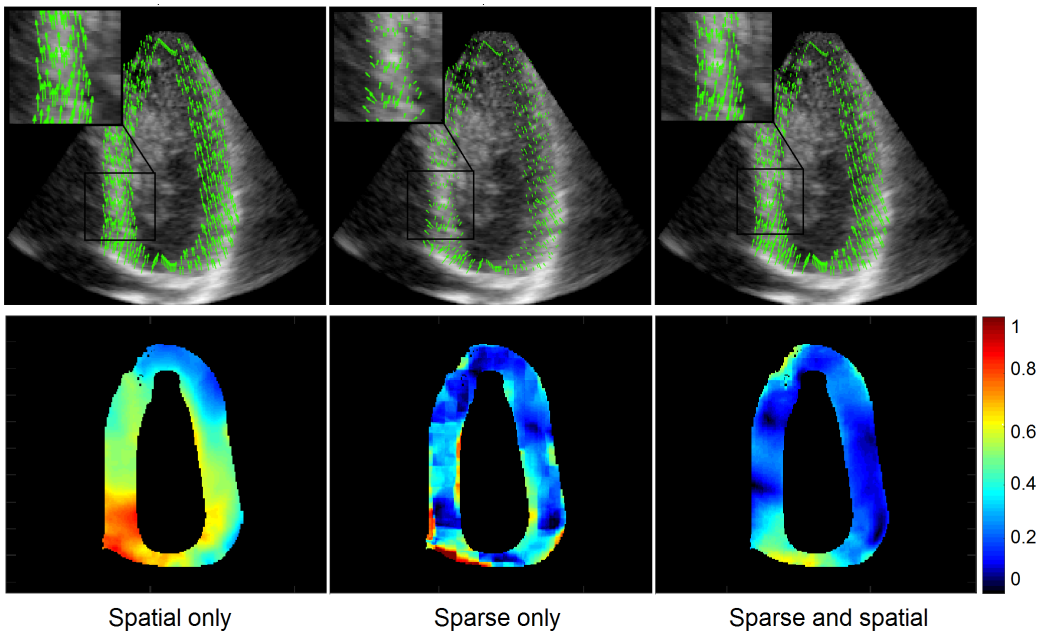


Figure 3.6: The estimated motion fields (top) and the corresponding error maps (bottom) for the 4th frame of the LADprox sequence with separated and combined sparse and spatial regularizations.

influence of each regularization term on the estimation process, the estimated motion fields between a pair of consecutive systolic frames (4th and 5th frames) of the LADprox sequence as well as the corresponding error maps are illustrated in Fig. 3.6. The error maps correspond to the endpoint error for the displacements of each pixel in the 4th frame. It is clear from the error maps (bottom) that the combined use of the spatial and sparse regularizations provides the smallest errors for this frame. The top row of this figure shows that the motion field estimated with the spatial term alone is over-smoothed and lacks structure, while the motions resulting from the sparse regularization alone lack smoothness. This is for example the case for some patch borders that create non-existent motion boundaries.

These conclusions are confirmed in terms of motion estimation accuracy for the entire sequence, detailed in Table 3.4. The global endpoint errors of the Normal sequence are also provided. Table 3.4 confirms that the combination of the two regularization terms provides the best performance in terms of average mean and std for all the frames of these sequences.

Sequence	Regularization strategy		
	Sparse and spatial	Spatial	Sparse
LADprox	**0.147±0.088**	0.205±**0.088**	0.1702±0.14
Normal	**0.251±0.09**	0.30±0.14	0.27±0.13

Table 3.4: Average error means and stds for the LADprox and Normal sequences using spatial, sparse and both regularizations.

3.5.5.2 Comparison with state-of-the-art methods

Global endpoint error

The results obtained for all the sequences in terms of mean endpoint error are summarized in Table 3.5. Note that the LADdist sequence was not considered for evaluation since it was used for the training of the motion dictionaries. The results show that the proposed method performs better than the three other algorithms in terms of average mean and std of the endpoint error calculated using the entire sequence.

A comparison with the errors studied in [Alessandrini 2016b] for 3D state-of-the-art motion estimation methods is provided in Table 3.6. The results show smaller errors for the proposed 2D method for the considered normal geometry dataset (*i.e.*, LADprox, Rca, LCX and Normal). The much better results obtained with the proposed method can be explained by the fact that it is not affected by the azimuthal spatial resolution contrary to the 3D methods of [Alessandrini 2016b]. At this point it is worth mentioning that there is no theoretical limitation for a 3D extension of the method, since the three terms used in the formulation of the motion estimation problem (3.14) are not limited to 2D.

Sequence	Method			
	Proposed	Block-matching	B-spline	Monogenic
LADprox	**0.147±0.088**	0.658±0.463	0.283±0.116	0.304±0.177
Normal	**0.251±0.099**	0.876±0.484	0.467±0.145	0.432±0.208
Rca	**0.165±0.102**	0.758±0.483	0.286±0.123	0.313±0.169
LCX	**0.153±0.108**	0.581±0.341	0.835±0.519	0.3014±0.174
sync	**0.199±0.140**	0.985±0.583	0.981±0.6163	0.352±0.223
LBBB	**0.171±0.132**	0.673±0.459	0.619±0.365	0.285±0.192
LBBBsmall	**0.186±0.140**	0.821±0.538	0.701±0.379	0.316±0.205

Table 3.5: Endpoint error for the realistic simulations dataset.

Algorithm	Normal Geometry
AAOF	[0.09; **0.38**; 1.35]
AFFD	[0.14; **0.48**; 1.30]
S-Demons	[0.15; **0.49**; 1.41]
BM	[0.26; **0.86**; 2.11]
RFBM	[0.10; **0.72**; 4.40]
Proposed 2D	[0.03; **0.19**; 0.63]

Table 3.6: Comparison with 3D state-of-the-art methods studied in [Alessandrini 2016b] in terms of global motion estimation error [5th percentile, median, 95th percentile].

Time evolution of the errors

In order to appreciate the performance of the methods during the cardiac cycle, Fig. 3.7 shows the time evolution of the estimates for all the sequences. The difference in the performance between the methods is less pronounced at the end of the sequences, where the displacements are small. It is also clear that the BM algorithm is the least accurate for all sequences. The proposed method outperforms the B-spline and monogenic signal algorithms for the LADprox sequence and for almost all the frames of the Rca sequence. However, for the Normal sequence, the proposed method is slightly outperformed by the B-spline method for the 4th and 6th frames and by the monogenic signal algorithm for the 2nd and 5th frames, whereas the proposed algorithm still provides better estimates in the middle and end of the sequence. Note that, contrary to the proposed method, the B-spline and monogenic signal algorithms employ a coarse-to-fine estimation scheme to deal with such large displacements. However, the differences in the estimation accuracy between the beginning (large displacements) and the end (small displacements) of the cardiac cycle are less pronounced, which is an interesting property of the proposed method. Similar results were obtained for the remaining sequences.

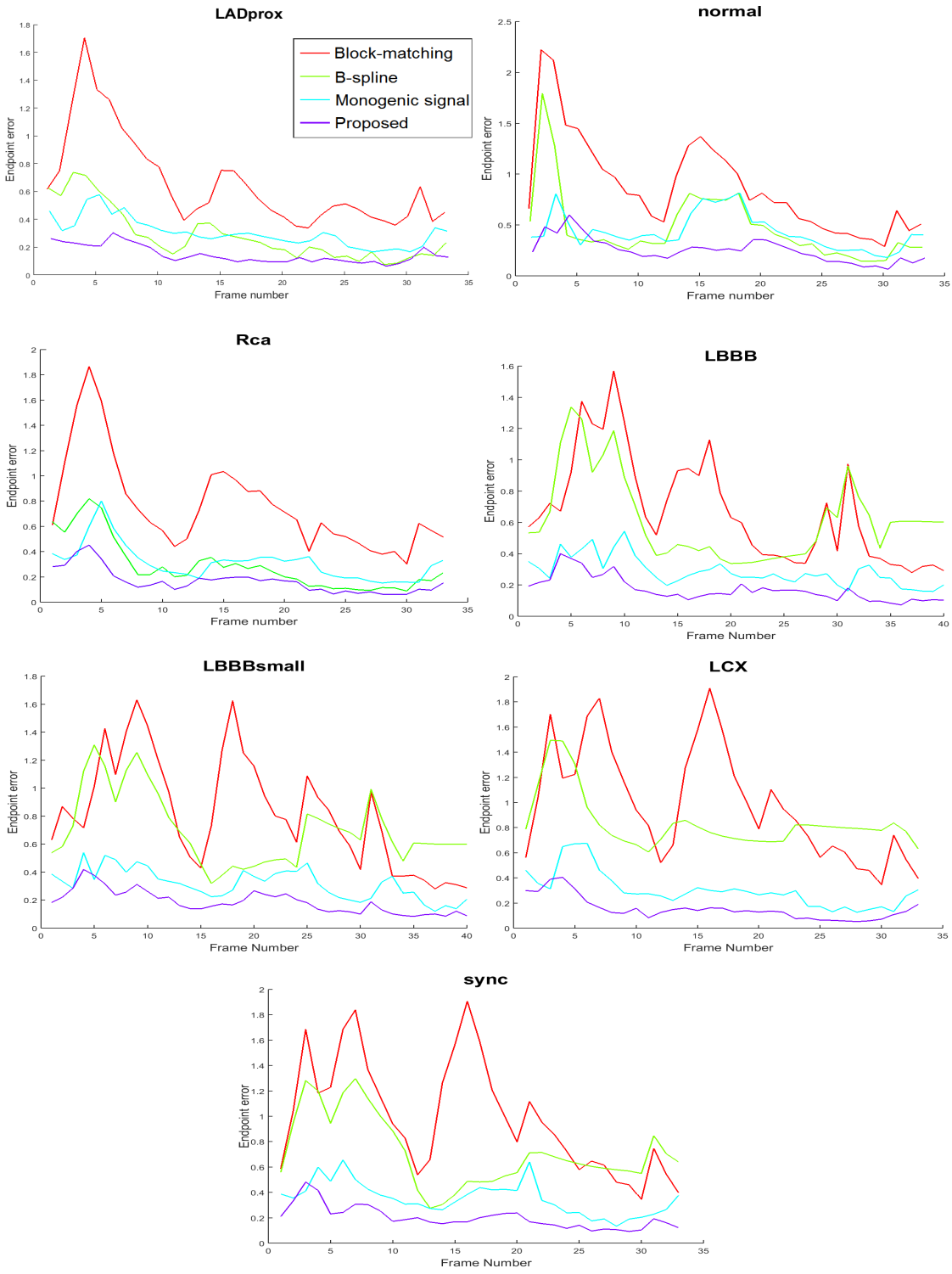
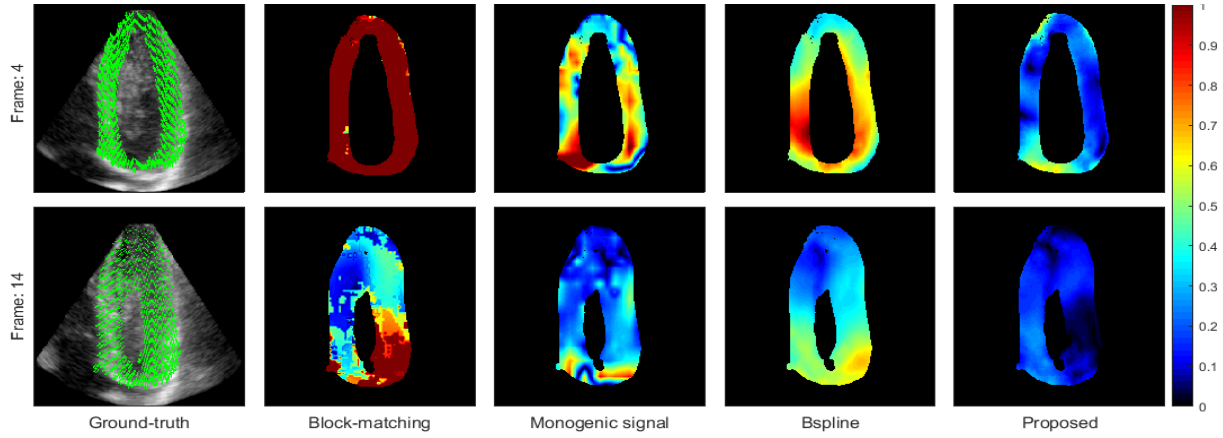


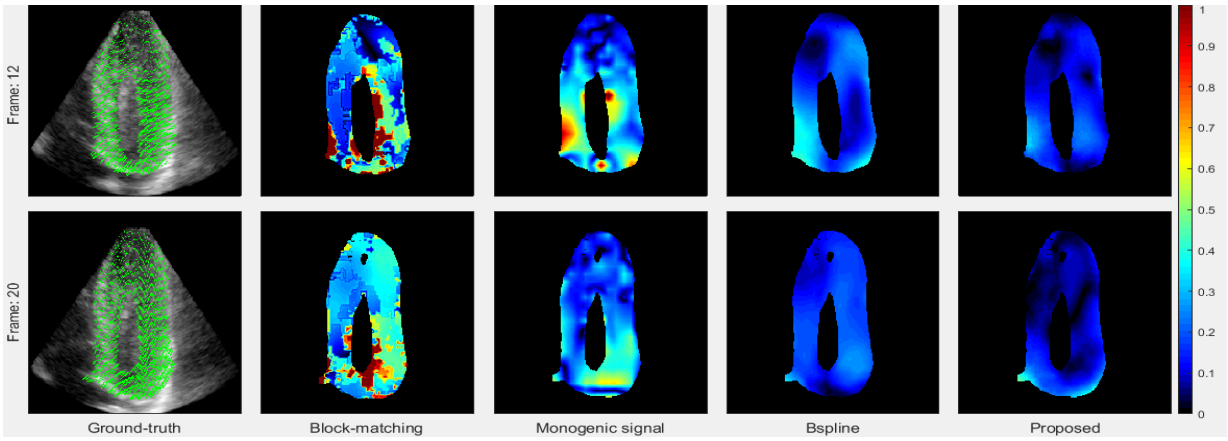
Figure 3.7: Mean endpoint errors for the sequences in the realistic simulations dataset.

Spatial analysis of the errors

In order to examine the local behaviour of the algorithms, error maps of the displacement estimates are displayed in Figs. 3.8a and 3.8b. The error maps of Fig. 3.8a match the 4th and 14th frames of the LADprox sequence, which respectively correspond to the maximum average displacement values in systole and diastole phases, whereas Fig. 3.8b shows the 12th and 20th frames of the sequence, corresponding respectively to the end systole and end diastole. Overall, all other considered sequences presented a similar behaviour.



(a)



(b)

Figure 3.8: Error maps and ground-truth meshes for the 4th, 12th , 14 and 20th frames of the LADprox sequence.

The behaviour observed for each considered algorithm can be commented as follows

- BM: as shown in previous results, the BM algorithm is outperformed by all methods. This can be explained by the resulting integer values of the displacements,

which are necessary to interpolate in order to achieve sub-pixel precision. Another issue with this method is the use of the NCC similarity measure, which is less adapted to UI than the E_{CD_2} data attachment term.

- B-spline: this method provides high errors in regions of rapidly varying motion. This is due to the fact that the B-spline model imposes too much smoothness on the motion field. When the regularization parameter is set to smaller values to limit over-regularization, very large errors appear in some frames of the sequence due to an excessive deformation of the B-spline mesh. This makes the parameter $\lambda_{\text{B-spline}}$ difficult to tune over an entire sequence. For the frames 12 and 20, the displacements are smaller and the B-spline method shows relatively lower errors.

- Monogenic signal: this algorithm does not suffer from over-regularization. However, it performs poorly for the 4th and 12 frames of the sequence. This is mainly due to the fact that the phase constancy assumption considered for this method holds less for large displacements.

- Proposed method: the proposed approach provides the smallest errors regardless of the displacement magnitudes. This shows that the algorithm is less sensitive to large variations and is adapted to more complex patterns of motion. Note that some errors are located at the border of the myocardium. These errors are possibly due to the influence of patches including estimates from outside this region.

Strain accuracy

The average means and stds of strain errors for the LADprox sequence are summarized in Table 3.7. The proposed method provides the most accurate estimates in terms of mean strain errors. Fig. 3.9 shows the mean longitudinal strain value for all the frames of the LADprox sequence. We can observe that the strain estimated with the proposed method is closer to the ground-truth.

Direction	Method		
	Proposed	B-spline	Monogenic
Radial	**0.085**±0.137	0.096±0.145	0.119±**0.114**
Longitudinal	**0.028±0.040**	0.031±0.060	0.041±0.060

Table 3.7: Average means and stds of the strain errors for the LADprox sequence.

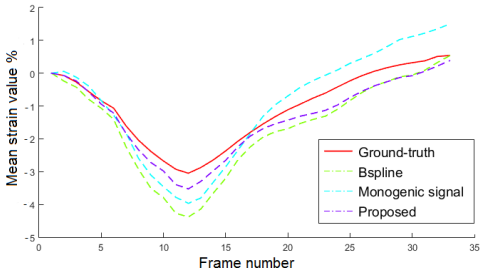


Figure 3.9: Mean longitudinal strain values (LADprox sequence).

Other data fidelity terms

Finally, we have investigated the possibility of using the proposed spatial and sparse regularizations jointly with other data fidelity terms. More precisely, Table 3.8 shows the average means and stds of the endpoint errors for three different similarity measures, *i.e.*, MI, SSD and the data fidelity term in (3.9) (CD_2), for the LADprox sequence. These results show that the CD_2 measure is slightly better than the MI and SSD measures for this sequence. However, the obtained results are clearly interesting when compared to those of Table 3.5. In order to have a more detailed view of the behaviour of the errors over the sequence, the mean endpoint errors are shown in Fig. 3.10 for all the frames. We can see that the differences between the errors is less pronounced when the displacements are large. These results confirm that the CD_2 measure is overall more adapted to the LADprox sequence.

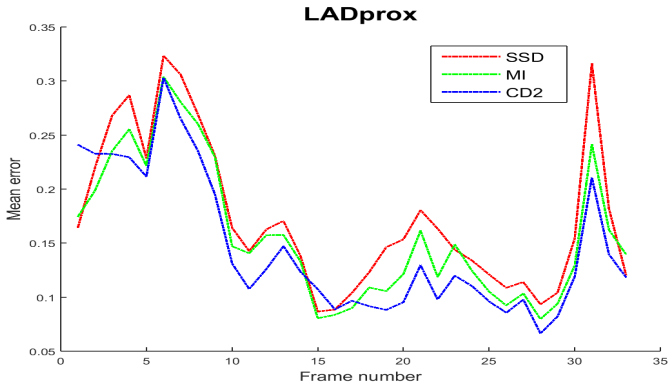


Figure 3.10: Comparison between SSD, MI and CD2 similarity measures for the LADprox sequence.

CD_2	MI	SSD
0.147±0.088	0.157±0.091	0.173±0.105

Table 3.8: Endpoint error for different similarity measures for the LADprox sequence.

3.5.6 In Vivo Data

This section evaluates the ability of the proposed approach to process *in vivo* US data acquired with a commercial scanner. In particular, we emphasize the coherence of the results obtained via a comparison between strain curves of healthy and pathological subjects. We also provide a visual analysis of the motion fields obtained with different methods on the same healthy cardiac sequence. Note that all the results presented below have been obtained without any post-processing and that the segmentation of the myocardium was performed manually by a cardiologist. In order to appreciate the wall

motion, all acquisitions were performed using the bidimensional mode. The data (image size: 248 × 267 pixels) was acquired from a short-axis view, using a GE Vivid S6 US machine equipped with a 3Sc-RS 1.5-4.0 MHz transducer array (center frequency 2.75 MHz).

The healthy reference sequence was acquired from a 37 years old male volunteer with no cardiovascular disease, whereas the second sequence (*i.e*, patient $n°1$) corresponds to an 88 years old female who has undergone non ST elevation myocardial infarction in the left descending artery territory. Finally, the third sequence (*i.e.*, patient $n°2$) is from a 40 years old male diagnosed with an acute myocarditis after an examination with coronary angiography and MRI. While the healthy subject did not have any wall motion trouble, the 88 years old woman had a severe hypokinesia in the antero-median segment. In the case of the 40 years old male (*i.e*, patient $n°2$), the initial echocardiography was deemed normal despite the inflammation of the cardiac muscle further diagnosed by MRI.

Visual analysis

First, a visual analysis of the motion estimates was performed for the healthy sequence. The results obtained for all previously introduced methods were then compared. Again, for all algorithms, the parameters were manually tuned to give the best visual aspect. The final parameters are provided in Section 3.5.3. The dictionaries $\boldsymbol{D_u}$ and $\boldsymbol{D_v}$ were learned using the LADdist sequence as explained in Section 3.5.3.

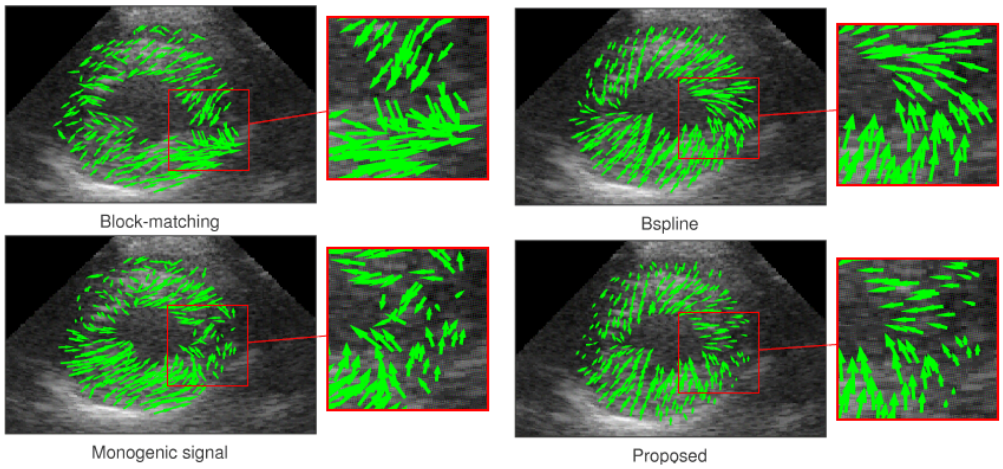


Figure 3.11: Comparison between the obtained motion fields for a systolic frame in the healthy *in vivo* sequence.

Fig. 3.11 shows the estimated motion vectors obtained for a systolic frame of the healthy sequence. Globally, all the methods provide a motion field that is coherent with the inward contraction (systole) of the cardiac muscle. However, the local behavior still presents differences in the amount of smoothness and spatial variation for the estimated fields. For instance, the B-spline method imposes too much regularization on the dis-

placements. In fact, the estimated vectors were not allowed to change rapidly enough to fit the spatial variations of the motion. On the other hand, the monogenic signal algorithm is adapted to more complex motions, but still lacks smoothness overall. For the proposed method, this smoothness was ensured while allowing more variation and complexity than the B-spline method. Overall, these observations confirm the outcome of the tests conducted on the realistic simulations dataset in Section 3.5.5.

Strain analysis

In order to gain additional insight into the obtained *in vivo* motion fields, the radial and circumferential strains were computed for the healthy and pathological sequences. Fig. 3.12 shows an example of segmentation for the healthy sequence. Fig. 3.13 provides the corresponding segmental strain values computed using the proposed method for the three sequences.

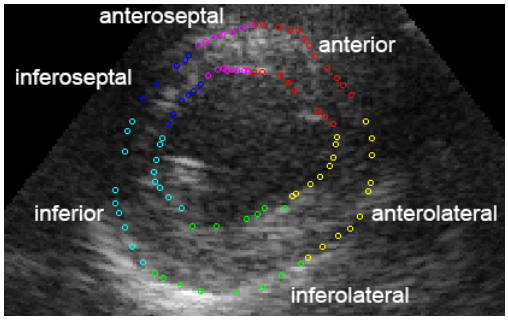


Figure 3.12: Segmentation of the myocardium for one frame of the healthy *in vivo* sequence.

In the case of the healthy subject, both the circumferential and radial directions present homogeneous strain curves for all segments. Typically, the circumferential strain exhibits negative strain values during the systolic contraction phase, followed by a plateau that continues until complete left ventricular (LV) relaxation at end diastole. On the other hand, the strain curves of the first pathological subject (patient $n°1$) indicate clearly that the segments corresponding to the left descending artery vascularization and its neighboring regions have been affected by the ischemia. Specifically, the anterior segment is completely hypokinetic (or akinetic), which also leads to abnormal myocardial wall motion in the neighboring segments (*i.e.*, anteroseptal and anterolateral segments). Radial strain curves are usually harder to interpret, but they are nonetheless coherent with the state of the two subjects. Note that these findings can be assessed through a simple visual inspection of the myocardial wall motion in the bidimensionnal mode. In the case of the patient $n°2$ we can clearly see abnormal deformations for the anterolateral, inferolateral and inferior segments in the radial direction. The anterolateral and inferior segments also present reduced circumferential strains. It is worth pointing out that these results suggest an abnormal heart condition, contrary to the

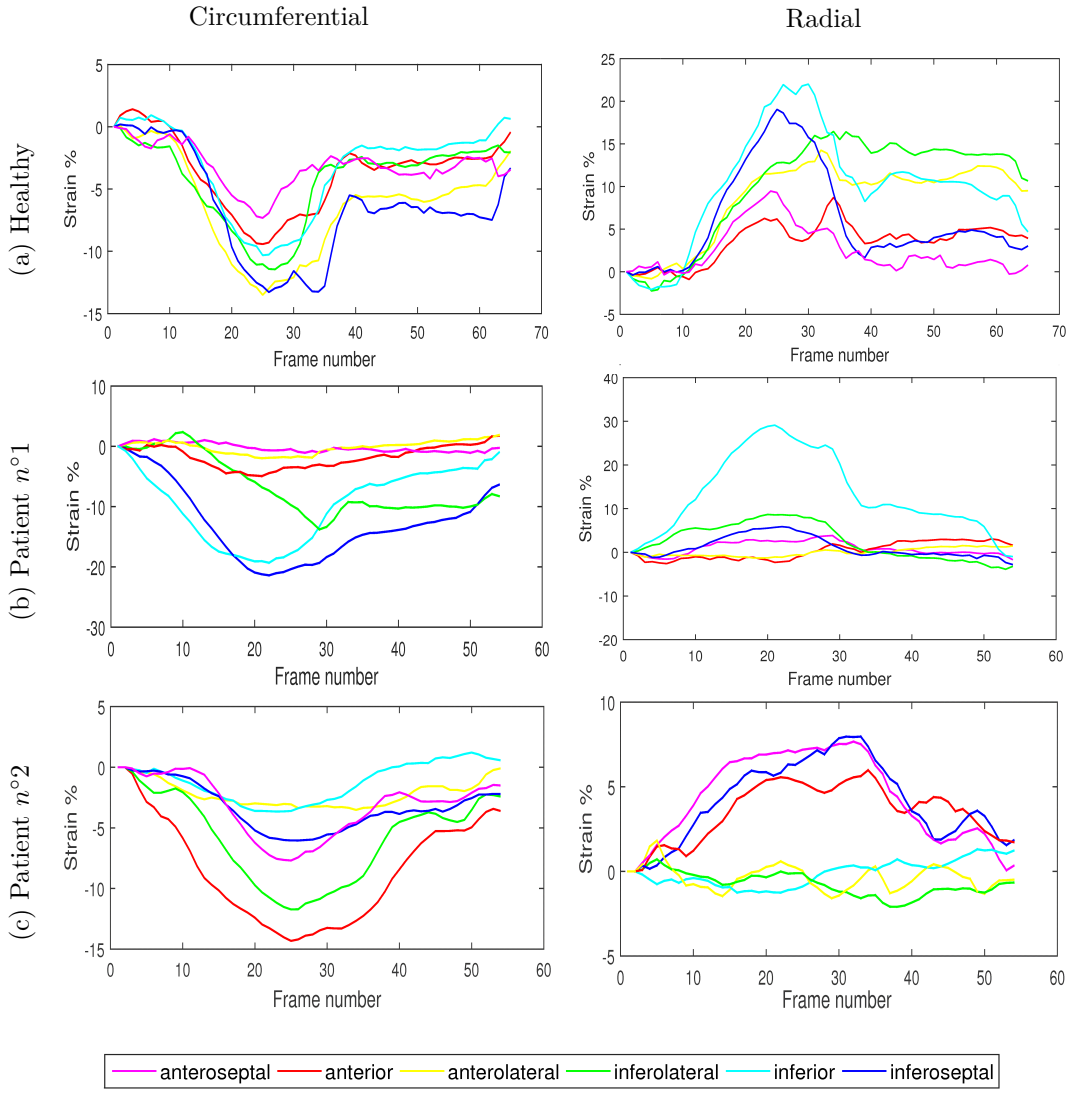


Figure 3.13: *In vivo* strain curves: circumferential (right) and radial (left) strains of (a) the healthy sequence, (b) the patient $n°1$ and (c) the patient $n°2$.

findings related to the visual examination of this sequence by the cardiologist. These final results support the suitability of the method regarding clinical assessment with real cardiac US data.

3.5.7 Adaptive dictionary learning

In this section, we briefly explore the results of cardiac motion estimation with adaptive DL, *i.e.*, updated dictionary (see Section 3.4.3 for more details) in the case of real data. The adaptive learning is enabled by running the optional 4th step of Algorithm

1 in Section 3.4. In order to illustrate the possible differences between the offline and adaptive DL strategies, we use the healthy *in vivo* sequence introduced above and the Normal sequence from the realistic simulations dataset. Note that in contrast with the offline strategy, the updated dictionary is initialized using random patches of the estimated motion field.

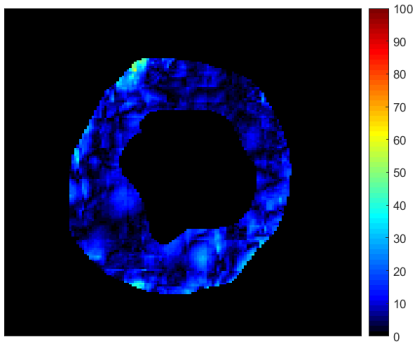


Figure 3.14: Relative difference in % between the motion fields of the same frame, estimated with the offline and adaptive learning methods.

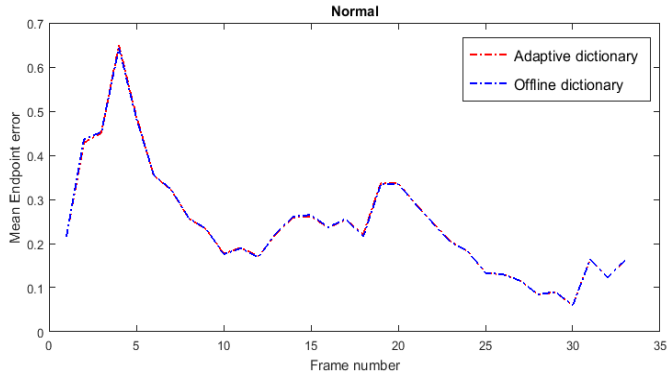


Figure 3.15: Mean endpoint errors for the Normal sequence using the offline and adaptive dictionary learning strategies.

Fig. 3.14 shows the relative difference between the estimated motion fields of a systolic frame in the healthy sequence (see Fig. 3.11) using both approaches. It is clear that the adaptive and offline DL strategies provide close estimates, with small relative differences (*i.e*, mean and std: $9.115\% \pm 6.757\%$). For the rest of the sequence, the relative differences remain overall negligible. Fig. 3.15 shows an example of results for the Normal sequence. We can see that the adaptive DL and fixed dictionary strategies provide similar performances for this dataset. Since the adaptive DL algorithm is more computationally intensive than its offline counterpart, we think that its use is not necessary when a rich training dataset is available. However, the adaptive DL strategy might be an appropriate choice when there are few training data.

3.6 Conclusions

This chapter introduced a method for estimating the cardiac motion of 2D US images with a regularization based on sparse representation and DL. The proposed approach combined two pertinent aspects of cardiac US. First, it incorporated the noise characteristics of US B-mode images via a similarity measure derived from the multiplicative Rayleigh noise assumption. The proposed method also regularized the motion by exploiting a sparse motion prior based on DL of typical cardiac motion patterns combined with a spatial smoothing term based on the gradient of the motion field. The experimental results showed the effectiveness of these regularizations for cardiac motion estimation. In terms of motion and strain accuracy, the results obtained with synthetic and realistic simulations demonstrated the competitiveness of this approach with respect to state-of-the-art methods. The results obtained on real data suggested that the method is consistent with the clinical interpretation related to images of healthy and pathological subjects. For future work, it would be necessary to investigate possible extensions of the algorithm to 3D UI. It is worth mentioning at this point that the data fidelity and regularization terms used in the actual formulation are not inherently limited to 2D and could be extended to 3D. Furthermore, the possible uses of the dictionary and the resulting sparse codes for other problems (*e.g.*, segmentation) would also deserve attention.

Chapter 4

Robust Optical Flow Using a Sparse Representation

Contents

4.1 Introduction

Despite its advantages, UI presents several shortcomings that make the interpretation of US images a difficult task. These limitations are principally related to the poor signal-to-noise-ratio caused by the multiplicative speckle noise but also to acquisition-related artefacts. In echocardiography, the interactions with artificial or anatomical highly reflective structures, *e.g.*, ribs, calcifications or prosthetic material, cause the so-called shadowing artefacts. Shadows are a primary cause of signal loss. Conversely, many other factors can produce brighter speckles or pseudoenhancement. Reverberations, ringdown and mirror artefacts are also common in UI. They are due to echoes bouncing multiple times before reaching the transducer, and thus, generating apparent false images or reflections. Several other types of UI artefacts are referred to as clutter [Turek 2015, Perperidis 2016]. Common consequences of clutter include low tissue-chamber contrast and overlaying stationary or moving structures that obscure the signal. The above-mentioned artefacts affect cardiac motion estimation. For example, regions with signal drop-outs or static reverberation clutter will seem akinetic. Motion artefacts can also affect the estimated displacements of adjacent tissues due to the smoothing commonly used in motion estimation problems. Anatomical boundaries, *e.g.*, between the myocardium and the background, can create motion discontinuities that violate these smoothness assumptions. Other factors affecting motion estimation in 2D cardiac US include out-of-plane motions resulting in discrepancies in the speckle pattern and erroneous estimations. Finally, random background motions (*e.g.*, in the blood) can affect the motion estimates inside the myocardium.

One way of overcoming the problem of image artefacts is to use detection or filtering strategies prior to motion estimation [Perperidis 2016, Turek 2015, Bjaerum 2002]. While this method can be efficient for specific types of artefacts, it does not solve the other problems affecting motion estimation, *i.e.*, motion discontinuities and the presence of atypical motions. For this purpose, more general techniques have been investigated to handle outliers in motion estimation problems. The most common approaches are based on robust M-estimators. The latter are defined according to the theory of robust statistics and provide the possibility of directly handling outliers. For example, M-estimators have been used for OF estimation in [Black 1996, Héas 2012] and an iteratively re-weighted approach has been considered in [Odobez 1995]. Many other strategies have been studied for robust motion estimation, *e.g.*, local OF methods based on the least median of squares [Ong 1999, Kim 2006], the use of multiple images to address the problem of drift between frames [Sariyanidi 2017] or feature extraction techniques to account for illumination variations [Sariyanidi 2017]. Robust approaches have also been investigated for medical imaging problems. For example, motion discontinuities have been taken into account in [McLeod 2015] for myocardial boundaries. In this work, a beforehand image segmentation is used to down-weight the epicardial motions. A robust registration is also applied to CT in [Aghajani 2016] by combining the l_1-norm for the observation model with a weighted edge preserving smoothness term. In the context of 3D US, an M-estimator-based OF method using a robust spatial smoothness term is

proposed in [Pratikakis 2001] for brain images. A specific similarity measure accounting for temporal speckle correlation is also considered in [Myronenko 2009a] for 3D cardiac US. In [Aviles 2017], a topology-preserving cardiac motion inference employing an M-estimator function for the data term is investigated for ultrafast US data. Robustness to data outliers has also been studied in the context of 2D elastography in [Rivaz 2011]. In this work, a robust weighting function is introduced within an iteratively re-weighted minimization strategy to deal with uncorrelated RF data. A similar approach has been employed for high-fame-rate echocardiography in [Porée 2018].

The objective of this chapter is to robustify the method introduced in Chapter 3 in order to mitigate the effect of outliers. The problem is formulated within a general OF-based energy minimization framework with two regularization terms. Following the strategy promoted in Chapter 3, these regularization terms are used to enforce spatial smoothness and sparsity of motion in a learnt cardiac motion dictionary. Moreover, robustness is introduced using weighting functions derived from M-estimators. In order to reduce the influence of imaging artefacts, motion discontinuities and background motions, weights are assigned to the data fidelity and the regularization terms. Finally, an iterative strategy is used to jointly estimate the motions, the sparse codes and the corresponding weights.

This chapter is organized as follows. Section 4.2 briefly reviews the theory of M-estimators and weight functions. Details about the problem formulation and the proposed robust motion estimation method are provided in Section 4.3. Section 4.4 describes the proposed implementation strategy. In Section 4.5, synthetic experiments are used to highlight the interest of robustness for the data fidelity and regularization terms. The proposed method is then compared to the non-robust cardiac motion estimation method introduced in Chapter 3 and to the robust OF method of [Black 1996], using realistic simulations of cardiac images. The feasibility of the method for real data is also demonstrated using *in vivo* images. Finally, discussions and concluding remarks are reported in Section 4.6.

4.2 Robust estimation

In the presence of imaging artefacts and native motion boundaries, the assumptions that are usually made about the nature of the images and motion are not always satisfied. The estimates that deviate from their expected behaviour are called outliers. They produce high residual errors and may affect the solution if not taken into account. For example, the usual choice of the squared ℓ_2-norm results in high costs for large errors (see Fig. 4.1a). A more appropriate cost function reduces the influence of estimates with large residuals, and is thus, *robust* to outliers. ML-type estimators, also called M-estimators, are robust functions that address the issue of outliers by reducing their impact on the estimates. For example, a typical M-estimator is obtained with the Huber cost function [Huber 1981], which combines the properties of the ℓ_2-norm for inliers and the robustness to outliers with the ℓ_1-norm (see Fig. 4.1a).

Different M-estimator functions provide different outlier rejection properties. In some

cases, the choice of an appropriate function can be motivated by some knowledge about the nature or amount of outliers. Other desired properties, such as differentiability, can also be taken into account in the context of optimization. In the case of the Huber function displayed in Fig. 4.1, the influence of outliers does not go completely to zero. In order to be more robust to outliers, one can rely on other types of robust functions, called redescending M-estimators. The latter allow the impact of outliers to be further reduced by controlling the decrease of the M-estimator function to zero.

4.2.1 Weighting functions

In this work, we use weight functions associated with redescending M-estimators. These functions depend on the first derivative of their corresponding M-estimator, *i.e.,*

$$w(\boldsymbol{e}_i) = \frac{r'(\boldsymbol{e}_i)}{\boldsymbol{e}_i} \tag{4.1}$$

where w denotes the weight function, r is an appropriate objective function, r' is the derivative of r and $\boldsymbol{e}_i$ is the residual error at pixel i. In order to ensure robustness, the weights of inliers tend to 1, while those of outliers tend to 0. In this work, two weight functions of redescending robust M-estimators are considered. The first one is the Lorentzian M-estimator, which is characterized by a differentiable weight function, with a gradual transition between the inliers and outliers defined as

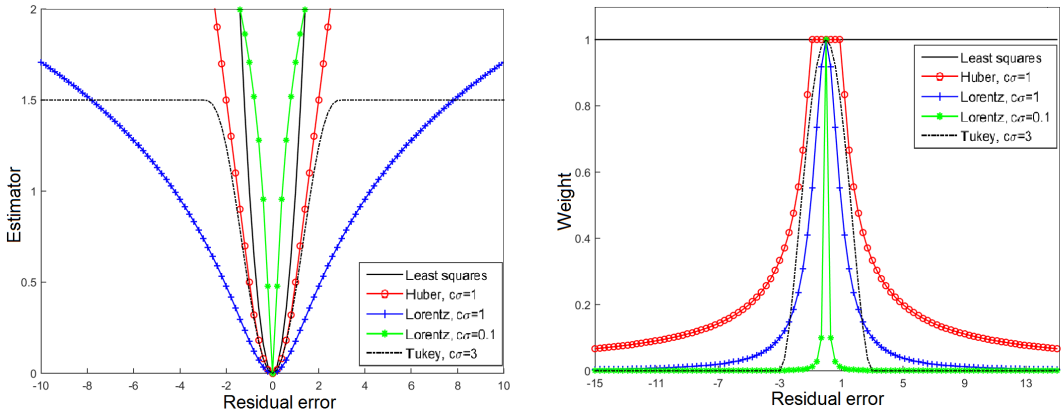
$$w_L(\boldsymbol{e}_i) = \frac{1}{1 + (\boldsymbol{e}_i / c\sigma)^2} \tag{4.2}$$

where $\sigma > 0$ is a scale parameter and $c > 0$ is a constant. The second example is the Tukey bisquare weight, referred to as Tukey Biweight and investigated for motion estimation in [Odobez 1995]. This function provides a hard rejection of outliers in comparison with the Lorentzian and is defined by

$$w_B(\boldsymbol{e}_i) = \begin{cases} [1 - (\boldsymbol{e}_i / c\sigma)^2]^2, & |\boldsymbol{e}_i| \leq c\sigma \\ 0, & |\boldsymbol{e}_i| > c\sigma \end{cases} \tag{4.3}$$

where $c > 0$ is also a constant and $\sigma > 0$ a scale parameter.

Fig. 4.1 shows the shapes of the weight functions (4.2) and (4.3) as a function of the residual error. The weights resulting from the least-squares and Huber formulations are also shown for comparison. In the least-squares case, all the weights have unit values and therefore all estimates contribute equally to the solution. This figure shows how observations with high residual errors result in considerably lower weights for the robust estimators. Note that for the Lorentzian function, the weights decrease gradually, whereas the transition from inliers to outliers is more abrupt for the Tukey Biweight function.



(a) (b)

Figure 4.1: (a) Different M-estimator functions and (b) the associated weights.

4.2.2 Parameter estimation

The two weight functions in (4.2) and (4.3) have tuning parameters c and σ that allow the outlier rejection threshold to be controlled, *i.e.*, the value of the residual error above which an estimate is considered as an outlier. In [Black 1996] outliers are gradually introduced by employing a continuation method where the value of the threshold is lowered at each iteration. Another approach for estimating c and σ consists in computing the outlier threshold jointly with the current estimate through the estimation of the scale parameter σ [Meer 1991]. This parameter represents the standard deviation of the residual errors for the inlier estimates. Due to the presence of outliers, the standard deviation of the errors is typically estimated using a robust estimator, for example, the median absolute deviation (MAD)[Meer 1991] defined as

$$\widehat{\sigma}_{\text{MAD}} = \sigma_0 \underset{i=1,\ldots,N}{\text{median}}[|\boldsymbol{e}_i - \text{median}(\boldsymbol{e})|] \tag{4.4}$$

with $\boldsymbol{e} = (e_1, ..., e_N)^T$. In order to be consistent, the MAD estimator of scale needs to be multiplied by a constant factor σ_0 (*e.g.*, $\sigma_0 = 1.4826$ for Gaussian errors [Odobez 1995]). Furthermore, the value of the parameter c is fixed *a priori*. This value depends on the considered M-estimator and allows the resilience to outliers to be controlled. The influence of the tuning parameters for the Lorentzian M-estimator is illustrated in Fig. 4.1 (with different parameter values, such as $c\sigma = 1$ and $c\sigma = 0.1$). In particular, smaller values of $c\sigma$ result in more outliers, *i.e.*, the weights tend to zero for relatively smaller residuals. Note that lowering the threshold value provides more robustness to outliers at the cost of lower efficiency for inliers. In this chapter, we make use of the robust weight functions introduced above in order to mitigate the effect of outliers.

4.3 Robust motion estimation

4.3.1 Problem formulation

This section formulates the motion estimation problem for a pair of consecutive frames in a US image sequence. The intensities of the two images (with N pixels) are concatenated in $\boldsymbol{I} \in \mathbb{R}^{2N}$ and the motion field between these two images is denoted as $\boldsymbol{U} = (\boldsymbol{u}^T, \boldsymbol{v}^T)^T \in \mathbb{R}^{2N}$, where $\boldsymbol{u} \in \mathbb{R}^N$ and $\boldsymbol{v} \in \mathbb{R}^N$ are the horizontal and vertical displacement vectors. The proposed robust motion estimation method is formulated as an energy minimization in an OF framework. The considered energy is defined as the sum of a data fidelity term denoted as E_{Q} and regularizations denoted as E_{S} and E_{W}. The first regularization E_{S} ensures a smooth spatial variation of the motion field, while the second one E_{W} exploits the patch-wise sparse properties of the motion vectors in $\boldsymbol{U}$, when decomposed on a learnt dictionary $\boldsymbol{D}$. The motion field is obtained through the minimization of the resulting cost function

$$\min_{\boldsymbol{\alpha},\boldsymbol{U}} \{E_{\mathrm{Q}}(\boldsymbol{U},\boldsymbol{I}) + \lambda_{\mathrm{W}} E_{\mathrm{W}}(\boldsymbol{U},\boldsymbol{\alpha}) + \lambda_{\mathrm{S}} E_{\mathrm{S}}(\boldsymbol{U})\} \tag{4.5}$$

where $\boldsymbol{\alpha}$ is the sparse coefficient vector, λ_{W} and λ_{S} are two regularization parameters that control the influence of the two regularizations. Prior to the motion estimation, the motion dictionaries are learnt offline from a set of training cardiac motion fields as in Chapter 3. In a second step, the motion of each pair of test images is estimated using the minimization problem (4.5). Further details about the way the data fidelity term and regularizations of (4.5) are defined are provided in the following sections.

4.3.2 Robust data fidelity term

This chapter considers an OF data fidelity term for motion estimation. OF estimation methods have shown their efficiency for cardiac motion estimation [Alessandrini 2013d, **?**]. In differential OF methods, the motion is usually estimated by linking the spatial and temporal image intensity variations within a least squares estimation (see Section 2.2.1.2). The main drawback of this quadratic formulation is the lack of robustness to outliers. For example, attenuated or noisy image pixels result in large residuals leading to important biases in the estimated motions. In order to address this issue, we propose to penalize less strictly the violations of the brightness constancy assumption using a weight matrix $\boldsymbol{Q} = \mathrm{diag}[\boldsymbol{q}(1), ..., \boldsymbol{q}(N)] \in \mathbb{R}^{N\times N}$. More specifically, data outliers are assigned low weights (*i.e.,* $\boldsymbol{q}(i)$ close to 0) while inlying estimates are not affected by the weighting process (*i.e.*, $\boldsymbol{q}(i)$ close to 1), where i is the pixel index.

Denoting as $\partial_t \boldsymbol{I}$ the temporal derivative of $\boldsymbol{I}$ at time t and as $\nabla \boldsymbol{I}^T \in \mathbb{R}^{N\times 2N}$ the spatial intensity gradients such that

$$\nabla \boldsymbol{I}^T = [\mathrm{diag}(\boldsymbol{\nabla} I_x), \mathrm{diag}(\boldsymbol{\nabla} I_y)] \tag{4.6}$$

where $\boldsymbol{\nabla} I_x$ and $\boldsymbol{\nabla} I_y$ are the spatial intensity gradients in both directions and ∇ indicates the gradient operator, the proposed robust data fidelity term is defined as follows

$$E_{\mathrm{Q}}(\boldsymbol{U}, \boldsymbol{I}) = \|\boldsymbol{Q}^{1/2}(\partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T \boldsymbol{U})\|_2^2. \tag{4.7}$$

The spatially-variant data weights in $\boldsymbol{Q}$, which allows the outlying estimates to be down-weighted, are obtained using a robust weight function (*i.e.*, w_L or w_B) introduced in Section 4.2 and the corresponding residual error $\boldsymbol{e}_d$ such that,

$$\boldsymbol{q}(i) = w_{L,B}[\boldsymbol{e}_d(i)], \forall i \quad \text{and} \quad \boldsymbol{e}_d = \partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T \boldsymbol{U} \tag{4.8}$$

where $w_{L,B}$ means that the weighting function can be w_L defined in (4.2) or w_B defined in (4.3). As explained in Section 4.2, the scale parameter σ_d is iteratively and jointly estimated with the motion estimates using (4.4). Note that it would be more difficult to robustify the data fidelity term E_{CD_2} employed in Chapter 3. The choice of the OF-based data fidelity term (4.7) in the present chapter was motivated by its simplicity.

4.3.3 Robust spatial regularization

The spatial regularization term ensures the smoothness of the motion estimates. A classical choice is the term (3.13) employed in Chapter 3, *i.e.*, $E_{\mathrm{TV}}(\boldsymbol{U}) = \|\nabla \boldsymbol{U}\|_2^2$ [Horn 1981]. This spatial regularization enforces weak spatial gradients on the two motion components. However, due to the use of the ℓ_2-norm, motion discontinuities are also penalized leading to over-smoothing and estimation errors around motion boundaries. In cardiac motion estimation, the regions associated with over-smoothing typically correspond to the inner and outer contours of the myocardium. In this work, we propose a weighted spatial regularization preserving motion discontinuities by assigning them lower weights ($\boldsymbol{s}(i)$ close to 0), while still imposing smoothness in homogeneous regions (where $\boldsymbol{s}(i)$ is close to 1). The associated robust spatial regularization term is formulated as follows

$$E_{\mathrm{S}}(\boldsymbol{U}) = \|\boldsymbol{S}^{1/2} \nabla \boldsymbol{U}\|_2^2 \tag{4.9}$$

where $\boldsymbol{S} = \mathrm{diag}[\boldsymbol{s}(1), ..., \boldsymbol{s}(2N)] \in \mathbb{R}^{2N \times 2N}$. The spatial weighting matrix $\boldsymbol{S}$ is computed using the error $\boldsymbol{e}_s$ associated with the magnitude of the motion field gradient such that,

$$\boldsymbol{s}(i) = w_{L,B}[\boldsymbol{e}_s(i)] \quad \text{and} \quad \boldsymbol{e}_s = \nabla \boldsymbol{U} \tag{4.10}$$

with $i = 1, ..., 2N$. Higher values of $\boldsymbol{e}_s$ result from rapidly varying motions, *e.g.*, at motion discontinuities. Because these discontinuities are of the same nature in both directions, the horizontal and vertical weights share the same spatial scale σ_s computed using (4.4). Note that the horizontal and vertical motion fields $\boldsymbol{u}$ and $\boldsymbol{v}$ are weighted separately. Note also that distinct weights could have been assigned to the horizontal and vertical gradient components, resulting in a set of four weights for each pixel. However,

in the case of cardiac motion estimation, the motion discontinuities usually occur in anatomical boundaries between the myocardium and the background, characterized by a similar discontinuous motion in both horizontal and vertical directions.

4.3.4 Robust Sparse Regularization

The proposed sparse regularization consists in finding the motion field $\boldsymbol{U}$ that is best described by a few atoms of a dictionary containing typical patterns of cardiac motion. This strategy has been introduced in Chapter 3, where the sparse regularization was performed patch-wise, so that each pair of motion patches $\boldsymbol{P}_p\boldsymbol{U}$ is constrained to have a sparse representation with respect to the motion dictionary $\boldsymbol{D}$, *i.e*,

$$E_{\mathrm{P}}(\boldsymbol{U}, \boldsymbol{\alpha}) = \sum\nolimits_p \|\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p\|_2^2 \tag{4.11}$$

where $\boldsymbol{P}_p \in \mathbb{R}^{2n\times 2N}$ is an operator that extracts the pth pair of patches in the horizontal and vertical directions from $\boldsymbol{U}$[1], $\boldsymbol{\alpha} \in \mathbb{R}^{2q\times N_p}$ is a sparse coding matrix whose columns are $\boldsymbol{\alpha}_p = (\boldsymbol{\alpha}_{u,p}^T, \boldsymbol{\alpha}_{v,p}^T)^T$, $\boldsymbol{D} \in \mathbb{R}^{2n\times 2q}$ is a block diagonal matrix whose blocks are $\boldsymbol{D}_u \in \mathbb{R}^{n\times q}$ and $\boldsymbol{D}_v \in \mathbb{R}^{n\times q}$, n denotes the patch size, q the number of atoms in each dictionary and N_p the number of patches. In order to ensure a sparse decomposition of the patches in the dictionary, the vectors $\boldsymbol{\alpha}_p$ in (4.11) are constrained to be sparse, *i.e.*, $\|\boldsymbol{\alpha}_{u,p}\|_0 \leq K$ and $\|\boldsymbol{\alpha}_{v,p}\|_0 \leq K$, with K a fixed maximum number of non zero coefficients.

The sparse prior in (4.11) is based on the assumption of a Gaussian error (expressed by the ℓ_2-norm) between the motion patches and their sparse representation in the learnt dictionaries. This assumption can be violated for outliers, *i.e,* patches containing displacements far from the patterns contained in the dictionary. This is the case, for example, when there are in-patch multiple motions or when the training data is not sufficiently rich. Since the dictionary contains only trained motions of the myocardium, typical outliers in cardiac motion estimation are the background motions and the patches located on the contours of the myocardium. In order to ensure robustness to outliers, we propose a weighting approach similar to the one adopted for (4.7) and (4.9). The influence of each patch is controlled by varying weights, *i.e.,* the pixels in the patches corresponding to outlying motions are assigned lower weights ($\boldsymbol{w}_p(i)$ close to 0), whereas the patch elements that are sufficiently close to the dictionary have higher weights ($\boldsymbol{w}_p(i)$ close to 1). The robust sparse regularization term is formulated as follows

$$E_{\mathrm{W}}(\boldsymbol{U}, \boldsymbol{\alpha}) = \sum\nolimits_p \|\boldsymbol{W}_p^{1/2}(\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p)\|_2^2 \tag{4.12}$$

where $\boldsymbol{W}_p = \mathrm{diag}[\boldsymbol{w}_{u,p}(1), ..., \boldsymbol{w}_{u,p}(n), \boldsymbol{w}_{v,p}(1), ..., \boldsymbol{w}_{v,p}(n)] \in \mathbb{R}^{2n\times 2n}$ is the weight matrix associated with the pth patch (n is the patch size). More specifically, the sparse coding weights are computed separately for the horizontal and vertical motion components

[1] $\boldsymbol{P}_p$ is a block diagonal matrix whose blocks are $\boldsymbol{P}'_p$, which extracts the pth patch in the horizontal or vertical direction.

according to

$$\boldsymbol{w}_{u,p}(i) = w_{L,B}[\boldsymbol{e}_{u,p}(i)] \tag{4.13}$$

and

$$\boldsymbol{w}_{v,p}(i) = w_{L,B}[\boldsymbol{e}_{v,p}(i)] \tag{4.14}$$

where $\boldsymbol{e}_{u,p}$ and $\boldsymbol{e}_{v,p}$ are the residual sparse coding errors of the pth patch for the horizontal and vertical motion vectors such that,

$$\boldsymbol{e}_{u,p} = \boldsymbol{P}'_p\boldsymbol{u} - \boldsymbol{D}_u\boldsymbol{\alpha}_{u,p} \tag{4.15}$$

and

$$\boldsymbol{e}_{v,p} = \boldsymbol{P}'_p\boldsymbol{v} - \boldsymbol{D}_v\boldsymbol{\alpha}_{v,p}. \tag{4.16}$$

Finally, the combination of (4.7), (4.9) and (4.12) results in a fully robust cardiac OF estimation, *i.e.,* with robust data fidelity, spatial smoothness and sparsity constraints, which allows the outliers to be mitigated for a better motion estimation. Section 4.4 studies the optimization algorithm that will be used to solve (4.5). The choice of the different weighting functions will be discussed in the section devoted to experimental results.

4.4 Implementation

In a first step, the motion dictionaries $\boldsymbol{D}$ are learnt offline as in Chapter 3 (see Section 3.4.1 for more details). After fixing the dictionaries, the proposed cardiac motion estimation can be formulated as the following minimization problem

$$\min_{\boldsymbol{U},\boldsymbol{\alpha}} \left\{ \|\boldsymbol{Q}^{1/2}(\partial_t\boldsymbol{I} + \nabla\boldsymbol{I}^T\boldsymbol{U})\|_2^2 + \lambda_{\mathrm{S}}\|\boldsymbol{S}^{1/2}\nabla\boldsymbol{U}\|_2^2 + \lambda_d \sum\nolimits_p \|\boldsymbol{W}_p^{1/2}(\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p)\|_2^2 \right\} \tag{4.17}$$

where the weights $\boldsymbol{Q}$, $\boldsymbol{S}$ and $\boldsymbol{W}_p$ have been defined in Section 4.3. The problem (4.17) can be solved using an alternate minimization strategy as in Chapter 3. It is based on an iterative approach, where the optimization alternates with respect to the motion $\boldsymbol{U}$ and the sparse codes $\boldsymbol{\alpha}$ for fixed regularization parameters λ_{S} and λ_{W}, before increasing the sparsity parameter λ_{W} and repeating the process. In this chapter, this strategy allows us to incorporate an iterative re-weighted minimization of (4.5), where the weights are determined in closed form and jointly with the motion estimates and the corresponding sparse coefficients at each iteration. More specifically, all the weights are initialized to 1 (no weighting). The residuals of the energy terms are then used to update the weights at each iteration according to the considered weight function (see Section 4.2). This approach allows outliers (*i.e.,* estimates with high residuals) to be removed from the estimation by gradually assigning them lower weights. Further details about these two steps are provided in the following.

Iteratively re-weighted minimization

1. *Sparse coding and weight estimation*
   The motion vectors in $\boldsymbol{U}$ are fixed and the optimization is performed with respect to $\boldsymbol{\alpha}$. The horizontal sparse vectors at the current iteration are determined by solving
   $$\min_{\boldsymbol{\alpha_u}} \sum\nolimits_p \|\boldsymbol{P'}_p \boldsymbol{u} - \boldsymbol{D_u}\boldsymbol{\alpha}_{u,p}\|_2^2 \text{ subject to } \|\boldsymbol{\alpha}_{u,p}\|_0 \leq K, \forall p.$$
   This sparse coding problem is NP-hard and is solved using the orthogonal matching pursuit (OMP) algorithm [Pati 1993]. A similar problem is solved to find the vertical sparse codes $\boldsymbol{\alpha}_v$. The sparse regularization weights $\boldsymbol{W}_p$ are then updated for each patch using (4.13) and (4.13) as explained in Section 4.3.4 .

2. *Motion field, data weight and spatial weight estimation*
   Once the sparse codes $\boldsymbol{\alpha}$ have been determined, the motion field $\boldsymbol{U}$ is updated by solving the minimization problem (4.17) with respect to $\boldsymbol{U}$ using the matrices $\boldsymbol{Q}$ and $\boldsymbol{S}$ that have been determined at the previous iteration. The minimization problem (4.17) is solved using the scaled conjugate gradient algorithm (SCG) [Moller 1993]. The data and spatial weights $\boldsymbol{Q}$ and $\boldsymbol{S}$ are then computed for the next iteration using respectively (4.8) and (4.10) (see Sections 4.3.2 and 4.3.3).

A full description of the sparse coding and motion estimation steps with an iterative weighting is provided in Algorithm 1. Note that the algorithm is stopped before the maximum number of steps *Outersteps* and *Innersteps* if the difference between two successive values of the total cost function is smaller than a selected threshold.

Scale parameter estimation

As explained in Section 4.2.2, the scale parameters σ_d, σ_s, σ_u and σ_v are computed jointly with the motions using (4.4). In this work, the scale parameters σ_u and σ_v corresponding to the weighting matrices $\boldsymbol{W}_p$ are computed using the global reconstructed errors $\boldsymbol{e}_{u,g} = \sum_p \boldsymbol{P}_p'^T \boldsymbol{e}_{u,p}$ and $\boldsymbol{e}_{v,g} = \sum_p \boldsymbol{P}_p'^T \boldsymbol{e}_{v,p}$ instead of the patch-wise ones (*i.e.*, resulting in a different outlier threshold for each patch). While the latter approach works well for motion boundaries, where a few pixels have large errors with respect to the rest of the patch, it does not guarantee the rejection of an entire outlying patch. The global errors $\boldsymbol{e}_{u,g}$ and $\boldsymbol{e}_{v,g}$ allow a common threshold to be computed for all patches. The patches with high but homogeneous errors are thus discarded.

Algorithm 2: Robust motion field estimation.

Input : $\boldsymbol{I}$,$\boldsymbol{D}$,K,λ_{S},λ_{W}, OuterSteps,InnerSteps

Initialization: $\boldsymbol{U}=\boldsymbol{0}$, $\boldsymbol{Q}=\boldsymbol{1}$,$\boldsymbol{W_p}=\boldsymbol{1}$, $\boldsymbol{S}=\boldsymbol{1}$

for $k=1$ **to** *OuterSteps* **do**

  for $j=1$ **to** *InnerSteps* **do**

    `%Sparse coding`

    $\boldsymbol{\alpha} \leftarrow \mathrm{OMP}(\boldsymbol{U},\boldsymbol{D},K)$;

    `%Sparse weight update`

    - Compute the residuals $\boldsymbol{e}_{u,p} \leftarrow \boldsymbol{P}'_p\boldsymbol{u} - \boldsymbol{D}_u\boldsymbol{\alpha}_{u,p}$, $\boldsymbol{e}_{v,p} \leftarrow \boldsymbol{P}'_p\boldsymbol{v} - \boldsymbol{D}_v\boldsymbol{\alpha}_{v,p}$
    - Reconstruct the global errors $\boldsymbol{e}_{u,g} \leftarrow \sum_p \boldsymbol{P}'^T_p \boldsymbol{e}_{u,p}$, $\boldsymbol{e}_{v,g} \leftarrow \sum_p \boldsymbol{P}'^T_p \boldsymbol{e}_{v,p}$
    - Compute the scales $\sigma_u \leftarrow \sigma_0\mathrm{MAD}(\boldsymbol{e}_{u,g})$, $\sigma_v \leftarrow \sigma_0\mathrm{MAD}(\boldsymbol{e}_{v,g})$
    - Update the weights $\boldsymbol{W}_p$ for $p=1,...,N_p$

    `%Motion estimation`

    $$\boldsymbol{U} \leftarrow \min_{\boldsymbol{U}} \|\boldsymbol{Q}^{1/2}(\partial_t\boldsymbol{I} + \nabla\boldsymbol{I}^T\boldsymbol{U})\|_2^2 + \lambda_{\mathrm{S}}\|\boldsymbol{S}^{1/2}\nabla\boldsymbol{U}\|_2^2 + \lambda_d \sum_p \|\boldsymbol{W}_p^{1/2}(\boldsymbol{P}_p\boldsymbol{U} - \boldsymbol{D}\boldsymbol{\alpha}_p)\|_2^2;$$

    `%Data and spatial weight update`

    - Compute the residuals $\boldsymbol{e}_d \leftarrow (\partial_t\boldsymbol{I} + \nabla\boldsymbol{I}^T\boldsymbol{U})$ and $\boldsymbol{e_s} \leftarrow \nabla\boldsymbol{U}$
    - Compute the scales $\sigma_d \leftarrow \sigma_0\mathrm{MAD}(\boldsymbol{e}_d)$ and $\sigma_s \leftarrow \sigma_0\mathrm{MAD}(\boldsymbol{e}_s)$
    - Update the weights $\boldsymbol{Q}$ and $\boldsymbol{S}$

  end

  - Adjust λ_{W}

end

Output : Motion $\boldsymbol{U}$, weights $\boldsymbol{Q}$, $\boldsymbol{W}_p$ and $\boldsymbol{S}$ and sparse codes $\boldsymbol{\alpha}$.

4.5 Experimental results

This section evaluates the proposed robust method using images with synthetic motions, realistic US simulations (with a controlled ground-truth) and real data. For the data with available ground-truth, the performance was evaluated using the endpoint error described in Section 3.5.2. As in Chapter 3 the dictionary was learnt using the simulation sequence LADdist, which contains realistic motion fields generated according to [Alessandrini 2016b]. The parameters used for learning this dictionary are provided in Section 3.5.3.

4.5.1 Synthetic data

The synthetic data consists of pairs of images that were used to evaluate the robust data fidelity term and the spatial and sparse regularizations used separately and jointly. In particular, the interest of the data fidelity and spatial regularization terms was highlighted using simple motions defined by translations. Pairs of synthetic cardiac images were then used to investigate the influence of the robust sparse regularization. After generating the different motions, the images were corrupted using a multiplicative Rayleigh noise, which is widely accepted in UI [Goodman 2007]. In this section, we use the Tukey Biweight function (4.3) for all robust terms. This choice was motivated by the fact that the synthetic images present abrupt transitions with clear boundaries (see Section 4.2). Cross-validation was used to determine suitable parameters for all the synthetic experiments, leading to $c_p = c_d = c_s = 7.4$ for all energy terms and $\lambda_{\text{S}} = 0.1$ for the spatial regularization parameter.

4.5.1.1 Robust data fidelity and spatial regularization

This section first considers the example of a pair of images with a simple translation as in [Black 1996]. The images correspond to a simple motion boundary with two parts with different intensities: the first one translates 1 pixel to the right in the horizontal direction, while the second one remains static. A vertical boundary separates the two regions, which are both contaminated by a multiplicative Rayleigh noise. In order to investigate the influence of the robust data fidelity and spatial terms, the sparse regularization was first removed by setting $\lambda_{\text{W}} = 0$. The motion field was then estimated for four different cases as follows.

1. First, we consider a non-robust formulation of (4.5) for which the weights are not updated, *i.e.*, $\boldsymbol{q}(i) = \boldsymbol{s}(i) = 1, \forall i$.

2. The weights are then updated for the spatial regularization term alone, *i.e.*, $\boldsymbol{q}(i) = 1, \forall i$ and $\boldsymbol{s}$ is computed as in (4.10).

3. The weights are updated for the data fidelity term alone, *i.e.*, $\boldsymbol{s}(i) = 1, \forall i$ and $\boldsymbol{q}$ is computed as in (4.8).

4. Finally the motion vectors are estimated using the fully robust formulation (4.5), with $\lambda_W = 0$.

Fig. 4.2 shows the resulting horizontal (top row) and vertical (bottom row) motion fields. The non-robust formulation asscociated with Fig. 4.2 (a,b) shows errors resulting from the data fidelity term and the over-smoothing at the boundary. By introducing a robust spatial regularization, discontinuities are allowed and the resulting motion in Fig. 4.2(c,d) is noisy. The use of a robust data fidelity term allows this noise to be mitigated by relaxing the brightness constancy assumption. The resulting motion displayed in Fig. 4.2 (e,f) is smooth with discontinuities that are not allowed at the boundary. Finally, the fully robust formulation results in a smooth motion field shown in Fig. 4.2 (g,h) with less data errors and a clear discontinuity in the horizontal flow between the two parts of the images. Quantitative results associated with the different formulations are reported in Table 4.1. The fully robust formulation provides competitive results in terms of mean and std of the endpoint errors.

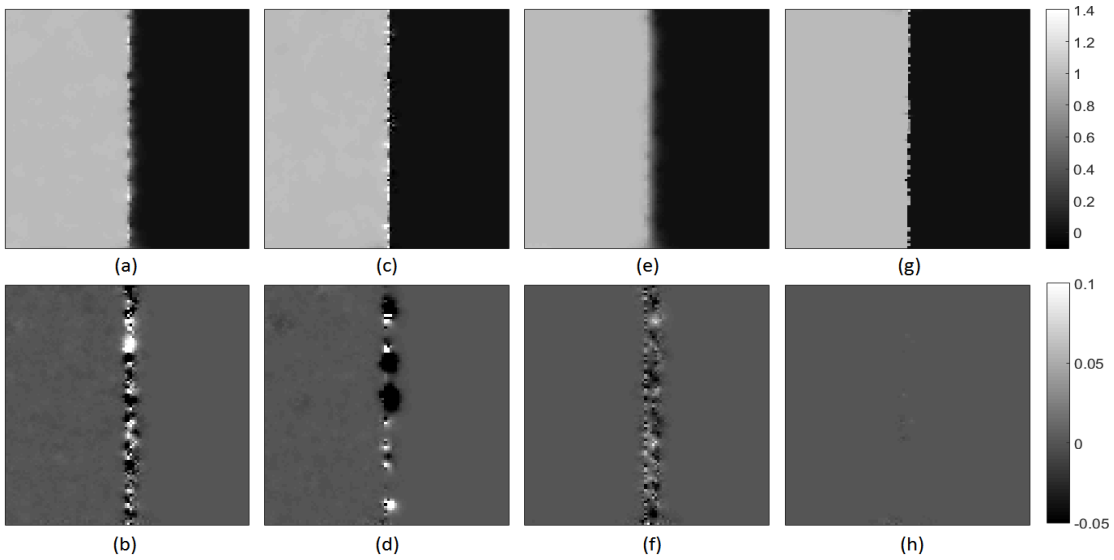


Figure 4.2: Estimated horizontal (top) and vertical (bottom) motion fields (in pixels) using the non-robust (a,b), robust spatial regularization only (c,d), robust data fidelity only (e,f) and fully robust (g,h) formulations.

Method	Non-robust	Robust spatial	Robust data	Fully robust
Error	0.020±0.10	0.018±0.14	0.023±0.10	**0.004±0.05**

Table 4.1: Means and stds of the errors for the translation images.

The final weights $\boldsymbol{Q}$ and $\boldsymbol{S}$ associated with the fully robust formulation are shown in Fig. 4.3. The lowest weight values correspond to the remaining outliers and are located

near the motion boundary. Note that due to the absence of vertical motion in both parts of the images, the associated weights $\boldsymbol{S}_v$ do not outline any spatial outliers in Fig. 4.3 (middle).

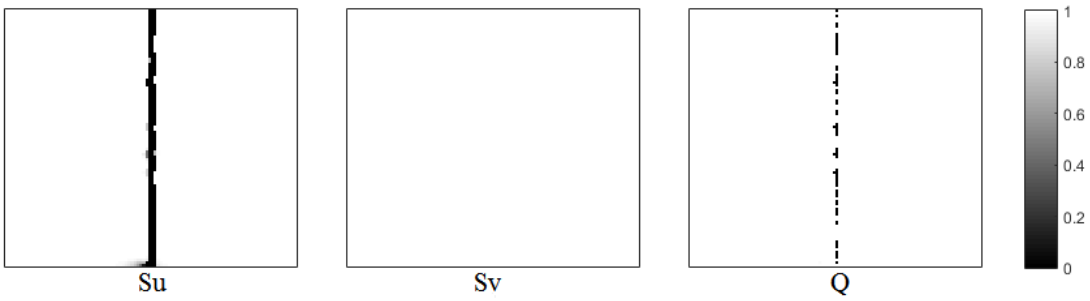


Figure 4.3: Estimated weights $\boldsymbol{S}$ and $\boldsymbol{Q}$ for the images subjected to translations.

4.5.1.2 Robust sparse regularization

This section investigates the interest of using a robust sparse regularization term. A pair of synthetic cardiac images was considered to compute the motion accuracy and analyze the resulting robust weights. The images were generated by corrupting an initial myocardium mask with a multiplicative Rayleigh noise and moving it according to peak-systole ground-truth displacements (*i.e.*, the largest displacements in the sequence). The background motion was generated using a mixture of two Gaussian distributions with variances equal to 5 and 25. The masks and the ground-truth motions were taken from the LADprox sequence of realistic simulations in [Alessandrini 2016b]. In order to simulate a UI artefact, a region of size 15×15 pixels was translated from 1 pixel in both horizontal and vertical directions. The amplitude of the artefact was set to 10 dB above the images (see Fig. 4.4).

Fig. 4.4 shows the estimated weights for the two images. The horizontal and vertical sparse weights were merged such that $\boldsymbol{W} = \text{diag}^{-1}(\boldsymbol{W}_u \boldsymbol{W}_v)$. The spatial weights $\boldsymbol{S}_u$ and $\boldsymbol{S}_v$ were combined similarly. The motion boundaries between the myocardium and the background produced the lowest spatial and sparse weights. As seen in Section 4.3, these motion discontinuities result in high gradient values and sparse coding residuals due to the absence of boundary elements among the training atoms. The low weights in these regions prevent over-smoothing, but also allow the sparsity constraint to be relaxed. Table 4.2 compares the estimation accuracy obtained with robust and non-robust sparse regularizations. These results are in favour of the fully robust formulation for the considered pair of images.

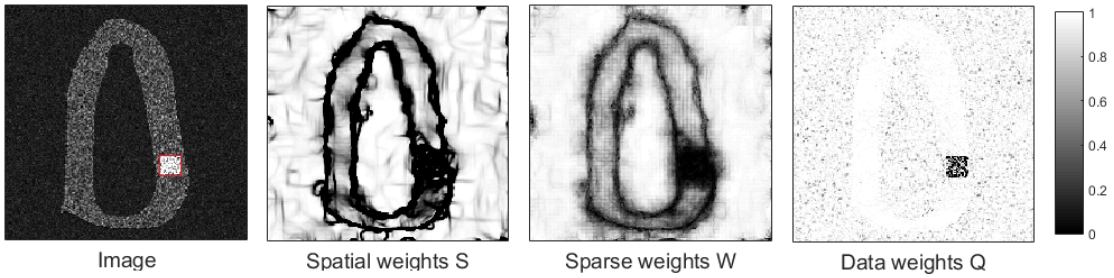


Figure 4.4: Synthetic cardiac image and its sparse weights $\boldsymbol{W}$, spatial weights $\boldsymbol{S}$ and data weights $\boldsymbol{Q}$. The colorbar indicates the weight values.

Method	Non-robust	Robust data and spatial only	Fully robust
Error	0.262±0.177	0.226±0.154	**0.194±0.153**

Table 4.2: Means and stds of the errors for a pair of synthetic cardiac images .

4.5.2 Realistic simulations

This section evaluates the proposed robust method using the realistic simulations dataset introduced in Section 3.5.5 [Alessandrini 2016b]. The sequences considered in this chapter include the healthy sequence (*i.e.,* Normal), the two ischemic cases with occlusions of the proximal and distal parts of the left anterior descending coronary artery (*i.e.*, LADprox and LADdist) and the sequence with occlusion of the left circumflex coronary artery (*i.e.*, LCX). The true displacements of the ischemic sequence LADdist were used to learn the dictionaries. As in the previous chapter, this choice allowed us to evaluate the method for different scenarios, *i.e.,* when the dictionary contains patterns of similar or different pathologies when compared to the test sequence. More specifically, the motion estimation accuracy was evaluated using the healthy sequence (*i.e.*, Normal), a sequence with a pathology similar to the training sequence (*i.e.*, LADprox) and a sequence with a distinct pathology (*i.e.*, LCX). Finally, a different imaging plane (*i.e.*, the short-axis view (SAX)) was considered for the LADprox sequence.

Tests were first conducted for the original sequences, containing only native outliers, *e.g.*, motion boundaries (see Section 4.5.2.2). In a second step, synthetic artefacts were introduced in order to corrupt the LADprox sequence (see Section 4.5.2.3). The performance of the proposed method was compared with two different methods. The first one is a robust motion estimation algorithm referred to as BA (for Black and Anandan) [Black 1996]. The BA method uses a robust OF-based data fidelity term with a robust smoothness constraint based on the gradient of the motion field. The second non-robust method (referred to as NR for non-robust) is the method studied in Chapter 3 and is based on an energy minimization framework. Its energy is defined using the data fidelity term (3.9) based on the assumption of multiplicative Rayleigh noise, the quadratic spatial smoothness term (3.13) and the sparse prior based on dictionary learning (3.12).

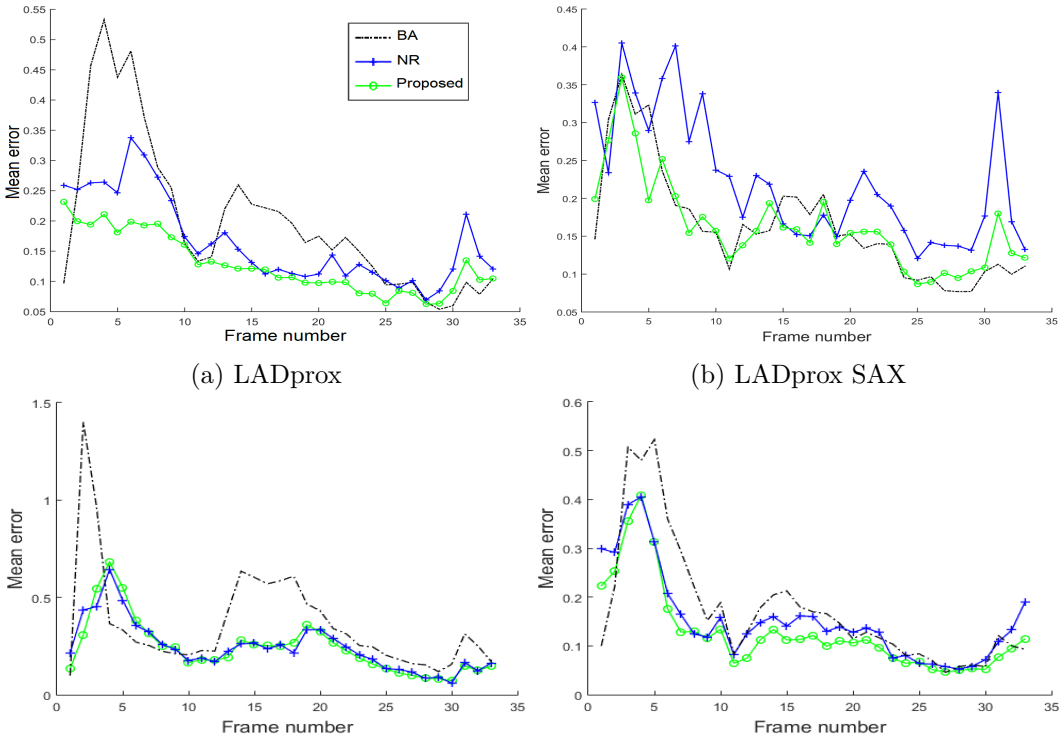
4.5.2.1 Robust estimation and regularization parameters

Using realistic simulation sequences with available ground-truth allowed the selection of motion estimation parameters providing the smallest average error (using cross-validation). For the proposed method, the optimal spatial parameter was $\lambda_S = 0.05$ for the LADprox SAX (short-axis view). For all other sequences, we chose to determine a suitable common spatial regularization parameter, *i.e.*, $\lambda_S = 0.2$. For each outer iteration, the sparse regularization parameter λ_W was logarithmically increased from 10^{-4} to 10 (see Section 4.4) in 4 iterations [Sulam 2015]. The Lorentzian weight function was used for all the considered sequences. This choice was motivated by the fact that the images used in this section have realistic motion boundaries with more gradual transitions, contrary to the experiments considered in Section 4.5.1. The parameters used for the computation of the robust weights were fixed to $c_d = 1$ for the data fidelity term, and to $c_p = c_s = 2.38$ for the sparse and spatial regularization weights, whereas the corresponding scale parameters were computed as explained in Section 4.3.

The parameters providing the best performance were also selected for the methods considered for comparison. For the BA method, the Lorentzian robust norm was used for the data and spatial terms, with a control parameter $\sigma_{BA1} = 0.01$ for the data fidelity term and $\sigma_{BA2} = 0.1$ for the spatial regularization. The smoothness parameter was fixed to $\lambda_{BA} = 1$ for the LADprox SAX sequence and $\lambda_{BA} = 5$ for all other sequences. The value of the spatial parameter for the NR method was adjusted by cross-validation, leading to $\lambda_{NR} = 0.75$ for the LADprox sequence, $\lambda_{NR} = 0.25$ for the LADprox SAX sequence and $\lambda_{NR} = 0.5$ for the data with artefacts. The dictionary learning and sparse regularization parameters of the NR method were adjusted as in Section 3.5.3.

4.5.2.2 Data without artefacts

The proposed robust method was first tested for the sequences without artefacts. Note that native outliers such as motion boundaries or motions from outside the myocardium can be present in the images. A comparison between the different methods in terms of the mean and stds of the motion estimation errors is provided in Table 4.3. Fig. 4.5 shows the mean errors for each frame of the cardiac cycle for the considered sequences. An improved performance can be observed, specifically for the large displacements in the beginning of the sequences (systole) for the LADprox sequence. These results show that robustness can also be beneficial for motion estimation accuracy in the absence of UI artefacts. For the Normal and LCX sequences without artefacts, a slightly better performance can be observed for the robust method when compared with the NR algorithm. This lower gain in performance can be explained by the fact that the regularization parameters in this chapter have not been tuned individually for each sequence contrary to the NR method (as explained in Section 4.5.2.1). Overall, the proposed robust approach consistently provides a higher or similar accuracy in terms of the mean endpoint error.



(a) LADprox (b) LADprox SAX

(c) Normal (d) LCX

Figure 4.5: Mean endpoint error for the uncorrupted LADprox, Normal and LCX sequences.

Sequence	Method		
	NR	BA	Proposed
LADprox	0.1410 ± 0.112	0.175 ± 0.128	**0.101** ± **0.089**
LADprox SAX	0.222 ± 0.147	0.164 ± **0.120**	**0.163** ± 0.124
Normal	0.240 ± **0.161**	0.363 ± 0.269	**0.237** ± 0.167
LCX	0.150 ± 0.151	0.169 ± 0.142	**0.125** ± **0.128**

Table 4.3: Error means ± stds for the sequences without artefacts.

In order to show the impact of the proposed robust approach on motion estimates, Fig. 4.6 shows the motion maps obtained for the 4th frame of the LADprox sequence compared to the NR method. It is interesting to outline that the smoothing effects close to the myocardium boundaries are significantly reduced for the robust approach. The large motions of the valves have also less impact on the regions near the base of the myocardium (see Fig. 4.9 for the corresponding error maps).

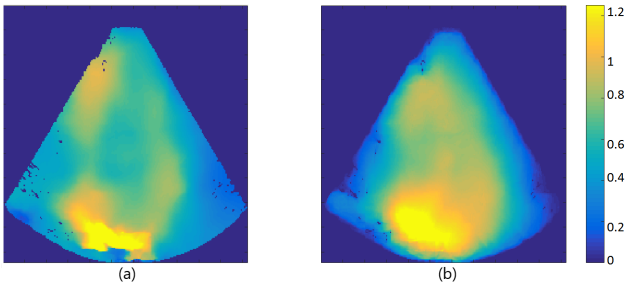


Figure 4.6: Estimated motions (in pixels) for the 4th frame of the LADprox sequence for (a) the proposed robust method and (b) the NR method of Chapter 3.

4.5.2.3 Data with artefacts

In order to further investigate the proposed robust approach, the LADprox sequence was corrupted using two different types of artefacts. The attenuation of an image region was used to simulate shadowing or loss of signal, while the increase in amplitude of a part of the image was used to simulate moving reverberations, reflections or brighter speckles. Two sequences *Corrupted1* and *Corrupted2* were created with different attenuation and reflection magnitudes. For each pair of consecutive images, only one frame was attenuated in a region of size 15×15 pixels, with attenuations of 10 dB and 15 dB. For all the images, a region of the same size was corrupted using reflection amplitudes of 5 dB and 10 dB. Between each pair of consecutive frames, the artefact moved 1 pixel in the horizontal and vertical directions. Fig. 4.7 shows the first two frames of the *Corrupted1* sequence.

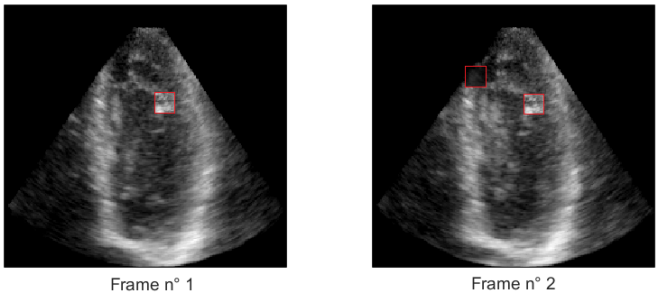
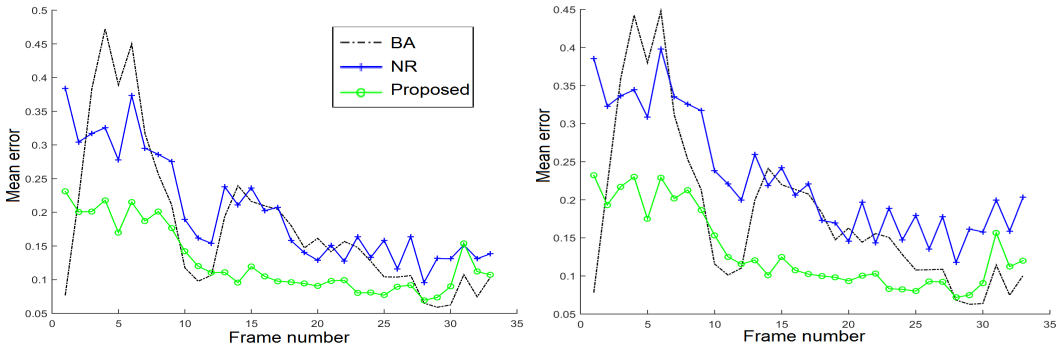


Figure 4.7: First and second frames of the *Corrupted1* simulation sequence. The red boxes indicate the regions with attenuation and clutter artefacts.

Endpoint errors

Table 4.4 summarizes the results obtained for the two corrupted sequences in terms of the global endpoint error. Note that the errors were computed for the uncorrupted regions only. The proposed robust and BA methods may result in different outliers. However, only the known outliers, *i.e,* added attenuation and clutter artefacts, are discarded for the error computation in order to ensure a comprehensible and efficient evaluation. The results show that the proposed method provides a competitive performance in terms of the error means and stds. In particular, the proposed method clearly outperforms the non-robust NR algorithm. Finally, the performance of the BA method was similar for all the sequences, with larger mean endpoint errors with respect to the proposed method.

Fig. 4.8 shows the time evolution of the mean endpoint errors for the entire cardiac cycle. Large differences with respect to the BA method can be observed at the beginning of the sequence, *i.e.,* the beginning of the cardiac cycle, where the displacements are large. These differences are less pronounced at the end of the sequence and around the 12th frame where the displacements are relatively small. Note that unlike the proposed method and NR, the BA algorithm uses a coarse-to-fine estimation scheme to cope with large motions. Moreover, the performance gap with respect to the NR method is consistent over the entire cardiac cycle.



(a) Corrupted1 (b) Corrupted2

Figure 4.8: Mean endpoint error for the two corrupted sequences.

Sequence	Method		
	NR	BA	Proposed
Corrupted1	0.196 ± 0.147	0.176 ± 0.124	**0.117** ± **0.116**
Corrupted2	0.222 ± 0.190	0.176 ± **0.121**	**0.122** ± 0.132

Table 4.4: Error means ± stds for the sequences with UI artefacts.

Spatial analysis of the errors

In order to understand the impact of robustness, Fig. 4.9 shows the error maps for the 4th (peak-systole) and the 12th (end systole) frames for the two corrupted sequences as well as the original LADprox sequence. For both corrupted sequences the NR method provided large errors around the artefacts, due to over-smoothing for the clutter artefact and data errors for the attenuation. Also, large errors can be observed in some uncorrupted portions of the frame, particularly for the 4th frame. This is due to the fact that in the non-robust approach (NR) the solution was highly impacted by outliers. Note that the errors for the corresponding regions in the uncorrupted LADprox sequence were much smaller for the NR method. The proposed robust as well as the BA methods did not suffer from this kind of errors since they allowed the impact of outliers to be mitigated. Globally, the proposed method provided more accurate estimates in comparison with the BA algorithm for the 4th frame and a similar performance can be observed for the 12th frame.

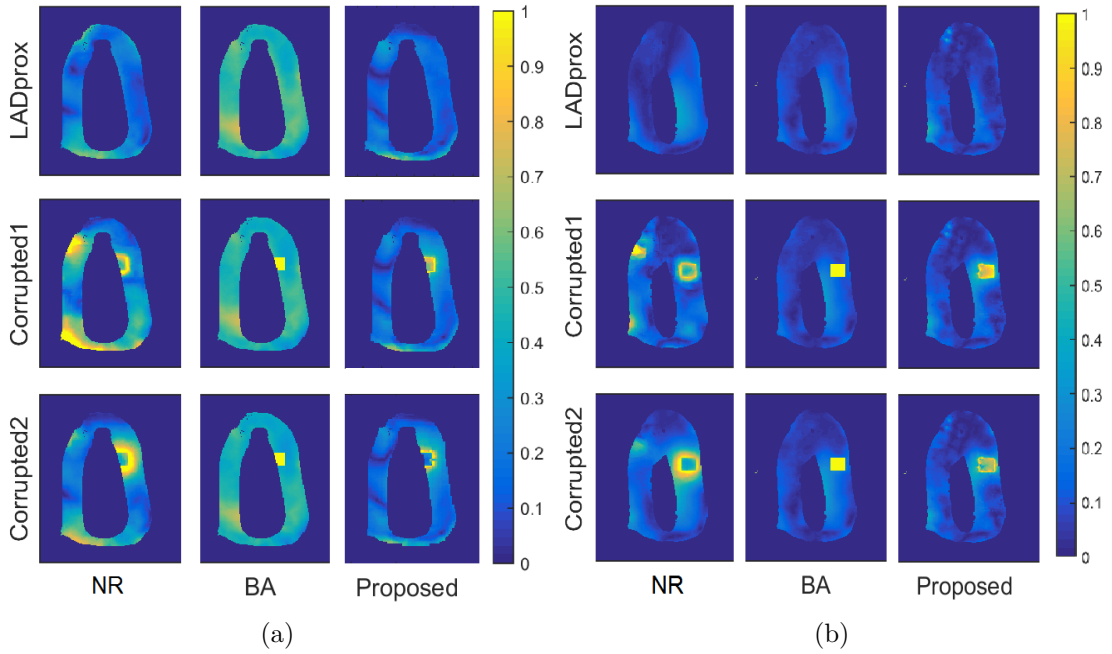


Figure 4.9: Error maps (in pixels) of (a) the 4th (maximum displacement) and (b) 12th (end systole) frames for the LADprox, *Corrupted1* and *Corrupted2* simulated sequences.

Finally, Fig. 4.10 shows the resulting data, spatial and sparse weights for the 4th (top row) and 12th (bottom) frames of the *Corrupted1* sequence. The horizontal and vertical spatial and sparse weights were merged as in Section 4.5.1. This figure shows that outliers are assigned lower weights compared to those of the inlying estimates whose values are close to 1. More specifically, the lowest data weights can be observed for the attenuation region and the outside of the myocardium. Attenuated regions are

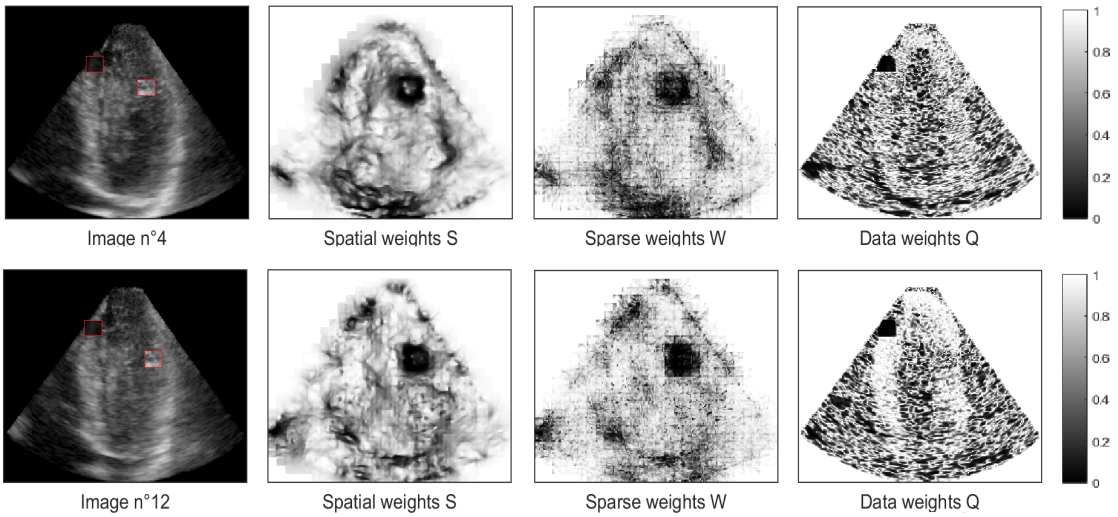


Figure 4.10: The 4th and 12th images of the *Corrupted1* sequence (the red boxes indicate the added attenuation and clutter artefacts) and the resulting robust weights.

typical data outliers, since the signal is attenuated or missing in one or both images. Overall, the data weights have low values since the $4th$ frame corresponds to the largest displacements of the sequence. For the spatial term, the lowest weights are assigned to motion discontinuities (as expected), *i.e.,* the contours of the added corruptions and the myocardium. The sparse weights are also low in the locations corresponding to motion boundaries. This can be explained by the fact that the training atoms did not contain motions lying between the inside and the outside of the myocardium.

4.5.3 In vivo

This section evaluates the proposed robust motion estimation method using two sequences of real US cardiac images. The sequences *in vivo 1* (image size 445×399) and *in vivo 2* (image size 510×372) were acquired at the Toulouse university hospital (CHU Rangueil, Cardiology service) using a GE Vingmed Ultrasound Vivid E9 machine equipped with an XDclear active matrix single crystal phased array transducer working at 1.5–4.6 MHz. The acquired sequences span a cardiac cycle of 50 frames. The first patient (*i.e., in vivo 1*) is a 60-year-old man referred for primary systemic (AL) with cardiac amyloidosis and congestive heart failure. The second sequence (*i.e., in vivo 2*) was acquired from a 18-year-old female referred for exercise dyspnea, with a transthoracic echocardiography showing normal left ventricular systolic function. The patch-size was set to 20×20 with motion dictionaries of size 400×600. The spatial regularization parameters were tuned to give the best visual tracking results (see Fig. 4.13), leading to $\lambda_{\mathrm{S}} = \lambda_{\mathrm{NR}} = 0.2$ for the proposed approach and the NR method and $\lambda_{\mathrm{BA}} = 1$, while the other parameters were the same as in Section 4.5.2.

Visual analysis

The interest of the robust approach is first illustrated by analysing the weights and motions obtained for a pair of images in each sequence. Fig. 4.11 shows two images and the corresponding data, spatial and sparse weights using the Lorentzian function. As in Section 4.5.2, this choice was motivated by the fact that *in vivo* images contain realistic motion boundaries with gradual transitions. In both cases, the lowest weights were assigned to the contours of the myocardium, the valves and some regions of the background. For the *in vivo 2* sequence, the reflection artefact (top right of the image Fig. 4.11 (e)) was also assigned lower spatial and sparse weights. Note that, as in the previous sections, the horizontal and vertical spatial and sparse weights were merged. A visual analysis of the motion fields obtained for the images in Fig. 4.11 is provided in Fig. 4.12. For the *in vivo 1* sequence, the regularization across the borders of the myocardium is reduced for the proposed method because of the low weights assigned to these motion boundaries. Also, the motions resulting from isolated brighter speckles are less spread to the neighboring mid and apical regions in comparison with the NR method. The same behaviour can be seen for the *in vivo 2* sequence for the proposed and BA methods in the regions near the valves (*i.e.*, the basal segments of the myocardium) characterized by clear motion discontinuities.

Comparison with a manual tracking

The manual tracking of 6 landmarks in 10 consecutive frames was used to compute motion estimation errors for the proposed method and the NR and BA algorithms. The landmarks were located on the endorcardium for both sequences, in the diastole phase for the *in vivo 1* sequence (frames 26 to 36) and during systole for the *in vivo 2* sequence (frames 16 to 26). The mean errors for the 10 frames are provided in Table 4.5. This table shows that the smallest error means and stds were obtained for both sequences using the proposed method. Fig. 4.13 provides the errors corresponding to three examples of images in each sequence. This figure also shows the displaced positions obtained for the 6 landmarks in comparison with the manual tracking (red circles). For the non-robust NR method, the displacements of the landmarks do not follow the true motion of the endocardium and result in the largest errors for the considered landmarks. In contrast, the motions obtained using the proposed and BA methods were closer to the manual tracking, with smaller errors for the proposed method.

Sequence	Method		
	NR	BA	Proposed
in vivo 1	5.239 ± 3.423	2.599 ± 2.157	**1.375 ± 0.59**
in vivo 2	8.304 ± 5.378	10.027 ± 8.161	**3.995 ± 2.518**

Table 4.5: Error means ± stds for 10 frames of the *in vivo* sequences.

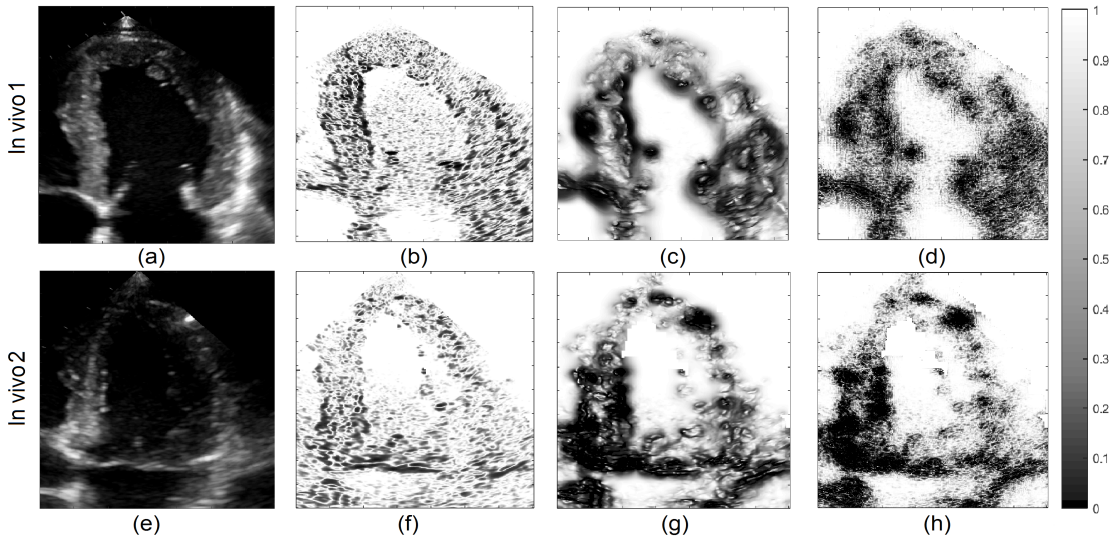


Figure 4.11: Weights for two examples of *in vivo* images obtained for two different patients. Images are displayed in (a) and (e) whereas the data, spatial and sparse weights are shown in (b)(c)(d) and (f)(g)(h).

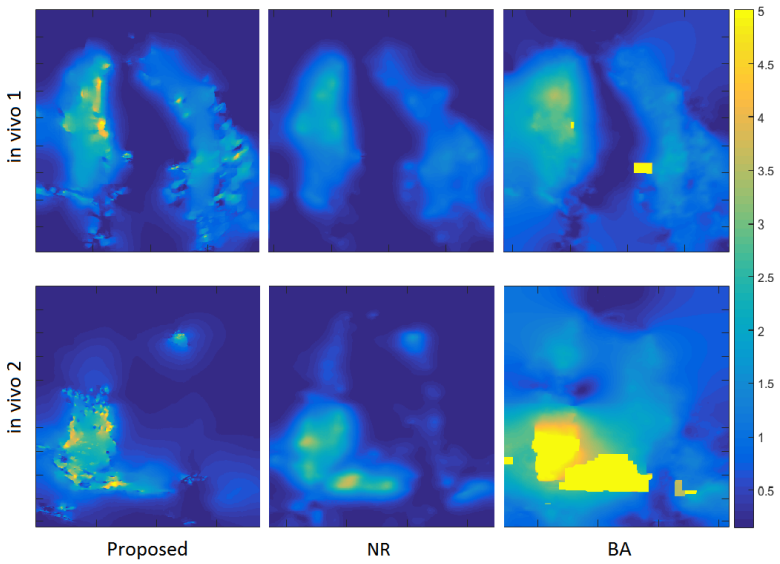
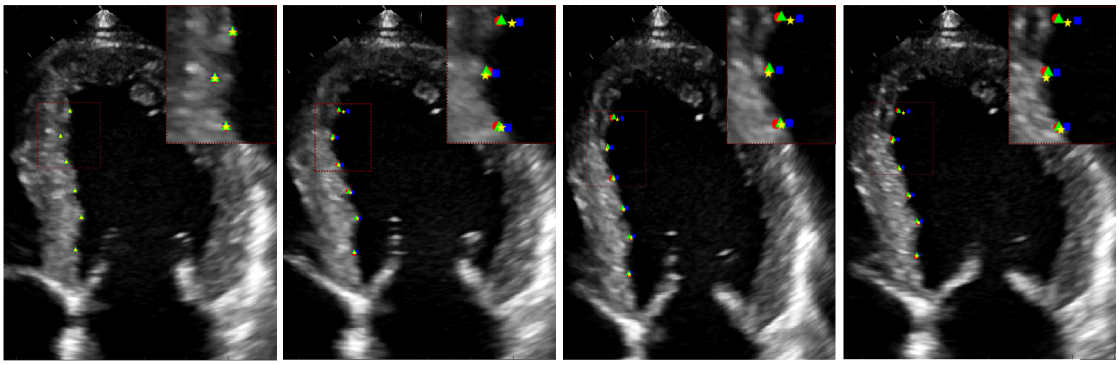
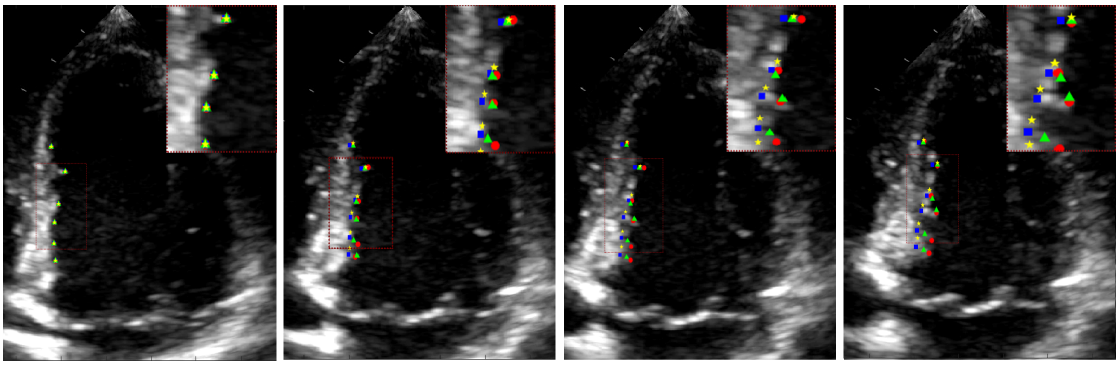


Figure 4.12: Displacement maps (in pixels) obtained for the proposed, NR and BA methods for the *in vivo* images in Fig. 4.11.



(a) Frame 26 (*in vivo 1*) Frame 30: $\epsilon = 1.38$, $\epsilon_{NR} = 5.51$, $\epsilon_{BA} = 2.9$ Frame 32: $\epsilon = 1.52$, $\epsilon_{NR} = 6.8$, $\epsilon_{BA} = 2.94$ Frame 35: $\epsilon = 1.79$, $\epsilon_{NR} = 7.13$, $\epsilon_{BA} = 3.50$



(b) Frame 16 (*in vivo* 2) Frame 20: $\epsilon = 3.41$, $\epsilon_{NR} = 7.77$, $\epsilon_{BA} = 9.38$ Frame 22: $\epsilon = 4.95$, $\epsilon_{NR} = 10.48$, $\epsilon_{BA} = 12.53$ Frame 25: $\epsilon = 5.17$, $\epsilon_{NR} = 12.28$, $\epsilon_{BA} = 14.43$

Figure 4.13: Tracking results for 6 landmarks on the endocardial wall (magnified in the top-right part of the images) for (a) the *in vivo 1* and (b) *in vivo 2* sequences. The average errors for each frame in comparison with the manual tracking (red circles) are provided for the proposed (green triangles), NR (blue squares) and BA (yellow stars) methods.

4.6 Conclusions and discussion

This chapter introduced a new robust motion estimation method for 2D cardiac US images. The main objective of this method was to robustify the cardiac motion estimation algorithm of Chapter 3 (based on spatial and sparse regularizations) in order to mitigate the effects of outliers. The motion estimation problem was formulated as a weighted energy minimization in an OF framework with combined spatial and sparse regularizations. Robustness was introduced using weight functions derived from M-estimators. In order to ensure a fully robust estimation, the weighting was applied jointly to the data fidelity term and the spatial and sparse regularizations. The obtained fully robust approach allowed us to deal with the problem of native outliers, *e.g.*, motion boundaries or background motions, as well as UI artefacts and image noise. The effectiveness of this fully robust formulation was demonstrated using synthetic realistic simulation sequences. Finally, *in vivo* images were used to show the interest of the method for real data contaminated by artefacts.

It is worth mentioning at this point that other strategies have been proposed in the literature to address the problem of cardiac motion estimation outliers (see Section 4.1). For example, in [McLeod 2015] the myocardium was segmented prior to the motion estimation, allowing to down-weight the displacements located at the epicardial borders, and thus, to prevent over-smoothing in this area. In contrast with the method studied in [McLeod 2015], the proposed method addressed the problem of spatial outliers for the entire motion field (*i.e.*, using pixel-wise weights). It allowed us to deal not only with discontinuities at the contours, but also with outliers located inside the myocardium. In addition, the proposed strategy did not require a beforehand segmentation (which may be difficult to obtain in some practical applications), allowing spatial discontinuities to be directly compensated from the estimated motions. More generally, the proposed approach showed the interest of jointly robustifying the data fidelity and regularization terms in a variational approach.

For future work, it would be interesting to consider an extension to 3D US images. Other prospects include the robustification of the dictionary learning step. A robust learning of the cardiac motion dictionary can be especially useful when using corrupted learning data. Furthermore, a joint motion estimation and segmentation would be possible using the resulting weight matrices.

Chapter 5

Time Consistent Optical Flow for Cardiac Ultrasound

Contents

5.1 Introduction

The previous chapters introduced two cardiac motion estimation methods that use only pairs of consecutive frames. One drawback of these pairwise approaches is that they do not exploit the temporal information embedded in the whole image sequence. One way of taking advantage of this information is to estimate the motions for all the data simultaneously, *i.e.*, using a groupewise approach. More specifically, a global estimation using all consecutive US frames can enforce temporal consistency of the resulting motion fields, *i.e.*, exploits the fact that the motions at a specific time instant are affected by the motions at other close time instants. Temporal consistency is an important property, especially in the case of cyclic displacement patterns such as cardiac motion. Another way of exploiting the temporal information available in image sequences consists in incorporating *a priori* knowledge about the temporal evolution of motions, typically assuming smooth trajectories. In the case of 2D echocardiography, the temporal smoothness assumption is useful because of the presence of speckle decorrelations, particularly due to out-of-plane motions or attenuations caused, for example, by shadowing.

The temporal aspects of motion have been studied in the context of OF-based motion estimation, *e.g.*, in [Volz 2011] where temporal smoothness has been incorporated in a multi-frame OF estimation approach. In medical imaging, temporal continuity has been exploited for the respiratory motions of lungs using 4D CT images [Jef 2010]. Furthermore, temporal consistency and smoothness have been investigated for cardiac motion estimation in MRI [Morais 2013, Shi 2013] as well as in the context of UI [Ledesma-Carbayo 2005, De Craene 2012, Shi 2013]. These methods typically extend the spatial FFD motion model into the time domain using B-splines. The motion fields of all the frames in the sequence are then estimated with respect to one reference frame [Shi 2013, Ledesma-Carbayo 2005, De Craene 2012]. The main drawback of this approach is that the estimation accuracy decreases for frames located far from the reference one. Temporal properties, such as smoothness, have also been investigated using sequential motion estimation methods where the motions of each frame depend on the previous estimates [Morais 2013]. However, the sequential methods do not allow temporal consistency to be incorporated.

The main objective of this chapter is to investigate the benefits of temporal consistency and smoothness for cardiac motion estimation in comparison with the methods of Chapters 3 and 4. As opposed to the pairwise approaches introduced in the previous chapters, the proposed groupewise implementation allows both temporal consistency and regularization to be incorporated. Moreover, the proposed implementation method allows us to estimate the motions for all the frames in the sequence without relying on a single reference frame. More specifically, the motion estimation problem is formulated within a general OF framework using the spatial and sparse regularizations, whose interest has been outlined in the previous chapters of this PhD thesis. In addition, a temporal regularization term imposing piecewise temporally smooth motions is introduced. In order to account for temporal discontinuities in the case of abrupt transitions, such as between the systole and diastole phases, a robust weighting method

is investigated. More precisely, an efficient optimization strategy based on an alternating direction method of multipliers (ADMM) using the constrained split augmented Lagrangian shrinkage algorithm (C-SALSA) [Afonso 2011] is proposed.

This chapter is organized as follows. Section 5.2.2 formulates the groupewise motion estimation problem and introduces the proposed temporal regularization term. Section 5.3 describes the proposed C-SALSA-based optimization strategy and provides implementation details. A preliminary experimental study is presented in Section 5.4. More precisely, the performance of the proposed groupewise method is compared to the methods presented in Chapters 3 and 4 using one sequence of realistic simulated 2D US images [Alessandrini 2016b]. Finally, perspectives and concluding remarks are reported in Section 5.5.

5.2 OF with spatial, sparse and temporal regularizations

5.2.1 Problem formulation

We consider the problem of 2D cardiac motion estimation for a sequence of images $\boldsymbol{I} = [\boldsymbol{I}_1^T, ..., \boldsymbol{I}_M^T] \in \mathbb{R}^{m_n}$, where $m_n = NM$ with N the number of image pixels and M the total number of frames such that $\boldsymbol{I}_t$ is the tth vectorized US image and $t = 1, ..., M$. The 2D motion fields to be estimated are denoted as $\boldsymbol{U}_t \in \mathbb{R}^{2N}$ and are concatenated in a vector $\boldsymbol{X} = [\boldsymbol{U}_1^T, ..., \boldsymbol{U}_M^T] \in \mathbb{R}^{2m_n}$ where $\boldsymbol{U}_t$ contains the motions between $\boldsymbol{I}_t$ and $\boldsymbol{I}_{t+1}$ for $t = 1, ..., M-1$ and $\boldsymbol{U}_M$ contains the motions between $\boldsymbol{I}_M$ and $\boldsymbol{I}_1$. The proposed motion estimation method is formulated as an energy minimization with an OF-based data fidelity term denoted as E_{OF}. In this chapter, we exploit the temporal smoothness of the motions by introducing a temporal regularization term denoted as E_{T}. As in the previous Chapters 3 and 4, we also make use of a spatial regularization term E_{TV} that ensures a smooth spatial variation of the motion fields, as well as a sparse regularization term E_{P} that exploits the patch-wise sparse properties of the motions in $\boldsymbol{X}$ when decomposed on a learnt dictionary of cardiac motions denoted as $\boldsymbol{D} \in \mathbb{R}^{2n \times 2q}$. In order to ensure a sparse decomposition of the motion patches in $\boldsymbol{D}$, the sparse coding vectors in $\boldsymbol{\Omega}$ are constrained to be sparse by introducing the term E_1, which enforces sparsity using the ℓ_1-norm (see Section 2.3.1). The choice of this approximation will be discussed in Section 5.2.2. The motion fields are finally obtained through the minimization of an appropriate energy function as follows

$$\min_{\boldsymbol{\Omega}, \boldsymbol{X}} \{E_{\text{OF}}(\boldsymbol{I}, \boldsymbol{X}) + \lambda_{\text{TV}} E_{\text{TV}}(\boldsymbol{X}) + \lambda_{\text{P}} E_{\text{P}}(\boldsymbol{X}, \boldsymbol{\Omega}) + \lambda_{\text{T}} E_{\text{T}}(\boldsymbol{X}) + \lambda E_1(\boldsymbol{\Omega})\} \tag{5.1}$$

where $\boldsymbol{\Omega} = [\boldsymbol{\alpha}_1^T, ..., \boldsymbol{\alpha}_{m_p}^T] \in \mathbb{R}^{2q \times m_p}$ contains the sparse codes associated with all the patches, with m_p the total number of patches in the sequence and $\boldsymbol{\alpha}_p = (\boldsymbol{\alpha_{u,p}}^T, \boldsymbol{\alpha_{v,p}}^T)^T \in \mathbb{R}^{2q}$ contains the horizontal and vertical sparse codes associated with the pth patch. The regularization parameters allowing the influence of each energy term to be controlled are denoted as λ_{TV}, λ_{P}, λ_{T} and λ.

As in the previous chapters, the dictionaries $\boldsymbol{D}$ are learnt offline from a set of training cardiac motion fields. Once the dictionaries have been determined, the motion fields of the considered sequence are estimated through the minimization of (5.1), where the data fidelity term and regularizations are detailed in the following section.

5.2.2 OF with spatial and sparse regularizations

In this chapter, we formulate the motion estimation problem using a groupewise approach allowing us to process the image sequence $\boldsymbol{I}$ as a whole. This groupewise motion estimation strategy leads to temporally consistent motions in the sense that the displacements in each frame depend on the entire sequence. The interest of this property will be investigated in Section 5.4 devoted to experimental results. The proposed strategy does not need to define a reference frame, which may be a problem in some applications because it would require the estimation of large displacements between distant frames. In addition, the proposed groupewise approach further exploits the temporal characteristics of the image sequence by incorporating a temporal smoothness term as shown in Section 5.3.2.

OF-based data fidelity term

The data fidelity term is based on the classical OF formulation (2.4) introduced in Chapter 2. In this chapter we propose to reformulate the data fidelity term using all the frames of the image sequence simultaneously, *i.e.*,

$$E_{\text{OF}}(\boldsymbol{I}, \boldsymbol{X}) = \frac{1}{2}\|\boldsymbol{Y} + \boldsymbol{A}\boldsymbol{X}\|_F^2 \tag{5.2}$$

where $\|.\|_F$ denotes the Frobenius norm and

- $\boldsymbol{Y} = [\partial_t \boldsymbol{I}_1^T, ..., \partial_t \boldsymbol{I}_M^T]^T \in \mathbb{R}^{m_n}$ is a vector containing the temporal derivatives of the intensities $\partial_t \boldsymbol{I}_t \in \mathbb{R}^N$ at each time instant $t = 1, ..., M$,
- $\boldsymbol{A} \in \mathbb{R}^{mn \times 2mn}$ is a block diagonal matrix whose blocks are $\boldsymbol{L}\boldsymbol{I}_t{}^T \in \mathbb{R}^{N \times 2N}$ such that

$$\boldsymbol{L}\boldsymbol{I}_t{}^T = [\text{diag}(\boldsymbol{L}_h \boldsymbol{I}_t), \text{diag}(\boldsymbol{L}_v \boldsymbol{I}_t)] \tag{5.3}$$

  where $\boldsymbol{L}_h \boldsymbol{I}_t$ and $\boldsymbol{L}_v \boldsymbol{I}_t$ are the spatial intensity gradients in both directions and $\boldsymbol{L}$ indicates the gradient operator, leading to

$$\boldsymbol{A}\boldsymbol{X} = [(\boldsymbol{L}\boldsymbol{I}_1{}^T \boldsymbol{U}_1)^T, ..., (\boldsymbol{L}\boldsymbol{I}_M{}^T \boldsymbol{U}_M)^T]^T. \tag{5.4}$$

Note that the term (5.2) incorporates the OF data fidelity terms of all consecutive frames and thus does not use a single reference frame.

Sparse and spatial regularizations

The proposed OF estimation method exploits the combination of sparse and spatial regularizations promoted in the previous chapters. We formulate the regularization terms following the groupewise strategy, *i.e.*, enforcing spatial smoothness and sparsity for all the motions in $\boldsymbol{X}$ simultaneously. More specifically, the pairwise spatial regularization (3.13) introduced in Chapter 3 is reformulated as follows

$$E_{\mathrm{TV}}(\boldsymbol{X}) = \|L_h\boldsymbol{X}\|_F^2 + \|L_v\boldsymbol{X}\|_F^2 \tag{5.5}$$

where $\boldsymbol{L}_h$ and $\boldsymbol{L}_h$ stand for the horizontal and vertical gradient operators respectively. In a similar way, the pairewise sparse regularization (3.12) can be reformulated for the entire sequence, *i.e.*,

$$E_{\mathrm{P}}(\boldsymbol{X}, \boldsymbol{\Omega}) = \|\mathcal{P}(\boldsymbol{X}) - \boldsymbol{D}\boldsymbol{\Omega}\|_F^2 \tag{5.6}$$

where $\mathcal{P}(.) : \mathbb{R}^{m_n} \mapsto \mathbb{R}^{2n \times m_p}$ is an operator that extracts m_p overlapping patches of size $2n$ from $\boldsymbol{X}$. In order to enforce sparsity, the sparse codes in $\boldsymbol{\Omega}$ are constrained to have a sparse decomposition in the dictionary $\boldsymbol{D}$. In this chapter, the sparsity inducing ℓ_0-pseudonorm is approximated by the convex ℓ_1-norm. This choice allows us to ensure the convexity of the cost function in (5.1), which will be a useful property for the proposed optimization method (see Section 5.3). The sparsity inducing energy term is finally defined as follows

$$E_1(\boldsymbol{\Omega}) = \|\boldsymbol{\Omega}\|_1. \tag{5.7}$$

5.2.3 Temporal regularization

In addition to the temporal consistency, which is made possible by the groupewise estimation strategy, the proposed temporal regularization term E_{T} enforces piecewise smooth trajectories. A straightforward way of imposing temporal smoothness is to penalize the differences between the displacements of consecutive frames. However, using only one frame (*i.e.*, the previous or next one) can lead to biased estimates that propagate towards the end or beginning of the sequence. In particular, when the motion is not cyclic, one cannot penalize the differences between the last and first frames of the sequence. Therefore, we propose to employ backward and forward temporal constraints simultaneously. More specifically, the motion of each pixel at a time instant t is constrained to be close to the motion of the corresponding pixel in the previous and next frames $t-1$ and $t+1$.

Let $\boldsymbol{X}_b = [\boldsymbol{U}_M^T, \boldsymbol{U}_1^T, ..., \boldsymbol{U}_{M-1}^T]$ and $\boldsymbol{X}_f = [\boldsymbol{U}_2^T, ..., \boldsymbol{U}_M, \boldsymbol{U}_1^T]$ be the time shifted versions of $\boldsymbol{X}$ containing the backward and forward motions with respect to each frame of the sequence. Note that the pixel coordinates $i + \boldsymbol{X}_b(i)$ and $i + \boldsymbol{X}(i)$ correspond to the new positions of the ith pixel in the frames t and $t+1$. In order to enforce temporal smoothness, one can penalize the differences $\boldsymbol{X}(i + \boldsymbol{X}_b(i)) - \boldsymbol{X}_b(i)$ and $\boldsymbol{X}(i) - \boldsymbol{X}_f(i +$

$\boldsymbol{X}(i)$) separately by minimizing the following energy term

$$E_{\mathrm{T}}(\boldsymbol{X}) = E_{\mathrm{T},b}(\boldsymbol{X}) + E_{\mathrm{T},f}(\boldsymbol{X}) \tag{5.8}$$

such that

$$\begin{array}{ll} E_{\mathrm{T},b}(\boldsymbol{X}) & = \sum_i [\ \boldsymbol{X}(i + \boldsymbol{X}_b(i)) \ - \ \boldsymbol{X}_b(i)\ ]^2 \\ E_{\mathrm{T},f}(\boldsymbol{X}) & = \sum_i [\ \boldsymbol{X}(i) \ - \ \boldsymbol{X}_f(i + \boldsymbol{X}(i))\ ]^2 \end{array} \tag{5.9}$$

where the differences in (5.9) represent the finite difference approximations of the first derivatives of the motions in the trajectorial direction. In this way, the motions in $\boldsymbol{X}$ are constrained to be close to the adjacent frames at each instant t. Another way of imposing smooth trajectories is by promoting a constant velocity. More precisely, one can use a finite difference approximation of the second derivatives of motions [Volz 2011], *i.e.*,

$$E_{\mathrm{T}}(\boldsymbol{X}) = \sum\nolimits_i [\ 2\boldsymbol{X}(i') \ - \ \boldsymbol{X}_b(i) \ - \ \boldsymbol{X}_f(i'')\ ]^2 \tag{5.10}$$

where the coordinates i' and i'' correspond to the new positions of the pixel i in the frames t and $t+1$, *i.e.*, $i' = i + \boldsymbol{X}_b(i)$ and $i'' = i' + \boldsymbol{X}(i')$. The difficulty that arises at this point is the non linearity of (5.10), which is due to the change in the pixel positions. In this work, we propose to bypass this difficulty by taking advantage of the iterative implementation strategy presented in Section 5.3. More specifically, the positions i, i' and i'' as well as the backward and forward frames are fixed from the previous iteration of the algorithm. In this way, the term $\boldsymbol{X}_f(i'')$ can be computed by a simple interpolation using the motions and positions of the previous iteration. For the term $\boldsymbol{X}(i')$, one could also employ an interpolation operator. However, this term contains the motions to be estimated (*i.e.,* not fixed). Therefore, the estimation of the target motions in $\boldsymbol{X}$ would require the computation of the adjoint interpolation operation. As an alternative, an approximation based on the 1st order Taylor expansion of the corresponding term is introduced as follows

$$\begin{aligned} \boldsymbol{X}[t, i + \boldsymbol{X}(t-1,i)] &\approx \boldsymbol{X}(t-1,i) + \partial_t \boldsymbol{X}(t-1,i) + \boldsymbol{L}\boldsymbol{X}(t-1,i)^T \boldsymbol{X}(t-1,i) \\ &\approx \boldsymbol{X}_b + \partial_t \boldsymbol{X}_b + \boldsymbol{L}\boldsymbol{X}_b^T \boldsymbol{X}_b \end{aligned} \tag{5.11}$$

where $\partial_t \boldsymbol{X}_b$ denotes the temporal gradients and $\boldsymbol{L}\boldsymbol{X}_b^T$ the spatial gradients of the motion fields. The effectiveness of this approximation will be shown experimentally in Section 5.4. Using the approximation (5.11), the temporal regularization term (5.10) can be reformulated as follows

$$E_{\mathrm{T}}(\boldsymbol{X}) = \|2\boldsymbol{X} - (\boldsymbol{X}_b - 2\boldsymbol{L}\boldsymbol{X}_b^T \boldsymbol{X}_b + \boldsymbol{X}''_f)\|_F^2 \tag{5.12}$$

where $\boldsymbol{X}''$ stands for the motions $\boldsymbol{X}_f$ at the positions i''. As mentioned above, the

proposed iterative implementation strategy allows the term $\boldsymbol{X}_b - 2\boldsymbol{L}\boldsymbol{X}_b^T\boldsymbol{X}_b + \boldsymbol{X}''_f$ to be computed from the previous iteration (see Section 5.3 for more details). In the rest of this chapter, we will use the term (5.12) since it has provided better experimental results for the considered data when compared to the temporal regularization term (5.9), as shown in Section 5.4.

Weighted temporal regularization

The temporal smoothness assumption does not always hold in echocardiography. More specifically, frame rate limitations can lead to a loss of temporal continuity for the rapid motions of the heart. Typically, the transitions between the systole and diastole frames are characterized by a discontinuous motion in time. Moreover, other discontinuities can occur in the same phase of the cardiac cycle, for example, at motion boundaries. In order to cope with temporal discontinuities, we use a robust weighting approach allowing us to reduce the effect of the temporal regularization for the pixels corresponding to large temporal discontinuities. More precisely, the proposed weighted regularization term is formulated as follows

$$E_{\mathrm{T}}(\boldsymbol{X}) = \|\boldsymbol{W}_{\mathrm{L}}^{1/2}[2\boldsymbol{X} - (\boldsymbol{X}_b - 2\boldsymbol{L}\boldsymbol{X}_b^T\boldsymbol{X}_b + \boldsymbol{X}''_f)]\|_F^2 \tag{5.13}$$

where $\boldsymbol{W}_{\mathrm{L}} = \mathrm{diag}([w_{\mathrm{L}}(\boldsymbol{e}_1), ..., w_{\mathrm{L}}(\boldsymbol{e}_{m_n})] \in \mathbb{R}^{m_n \times m_n}$, w_{L} is the Lorentzian weight function introduced in Section 4.2 and $\boldsymbol{e}_i$ is the residual of the ith pixel defined as

$$\boldsymbol{e} = 2\boldsymbol{X} - (\boldsymbol{X}_b - 2\boldsymbol{L}\boldsymbol{X}_b^T\boldsymbol{X}_b + \boldsymbol{X}''_f). \tag{5.14}$$

Note that the Lorentzian robust function is a redescending M-estimator that allows the influence of the discontinuities to be decreased to zero (i.e., $w_{\mathrm{L}}(i)$ close to 0). This property allows us to account for large shifts in the displacements, *e.g.*, at the end of systole.

5.3 Groupewise motion estimation

The problem (5.1) and the associated energy terms lead to the following optimization problem

$$\begin{aligned}\min_{\boldsymbol{X},\boldsymbol{\Omega}} \quad & \tfrac{1}{2}\|\boldsymbol{Y} + \boldsymbol{A}\boldsymbol{X}\|_F^2 + \lambda_{\mathrm{TV}}(\|\boldsymbol{L}_h\boldsymbol{X}\|_F^2 + \|\boldsymbol{L}_v\boldsymbol{X}\|_F^2) \\ & + \lambda_{\mathrm{P}}\|\mathcal{P}(\boldsymbol{X}) - \boldsymbol{D}\boldsymbol{\Omega}\|_F^2 + \lambda\|\boldsymbol{\Omega}\|_1 + \lambda_{\mathrm{T}}E_{\mathrm{T}}(\boldsymbol{X}).\end{aligned} \tag{5.15}$$

The problem (5.15) is jointly convex with respect to $(\boldsymbol{X}, \boldsymbol{\Omega})$ but hard to solve because of the high dimensionality of the unknown matrices $\boldsymbol{X}$ and $\boldsymbol{\Omega}$ as well as the presence of the non-quadratic sparsity enforcing regularization term $\|\boldsymbol{\Omega}\|_1$. In this chapter, this problem is addressed using the C-SALSA algorithm [Afonso 2011]. This choice allows us

to cope with the above-mentioned issues in a fast and efficient manner. More specifically, C-SALSA is based on the ADMM approach, which separates the minimization problem into simpler sub-problems, most of them having closed form solutions as explained in Section 5.3.2. Further details about the implementation of the C-SALSA method are provided in the following subsections.

5.3.1 Optimization method based on C-SALSA

The ADMM approach allows the problem (5.15) to be solved by iterating between simple optimization sub-problems. These problems are formulated by introducing auxiliary variables (*i.e.*, the so-called splittings) and handled in an approximate way by solving the associated augmented Lagrangian. In this work, we follow the C-SALSA splitting technique, leading to 7 auxiliary variables that stand for the motion fields in $\boldsymbol{X}$ and the sparse codes $\boldsymbol{\Omega}$. The optimization problem (5.15) is thus reformulated as follows

$$\begin{aligned}\min_{\boldsymbol{V}_1-\boldsymbol{V}_5} \quad & \tfrac{1}{2}\|\boldsymbol{Y}+\boldsymbol{A}\boldsymbol{V}_1\|_F^2+\lambda_{\mathrm{TV}}(\|\boldsymbol{L}_h\boldsymbol{V}_{21}\|_F^2+\|\boldsymbol{L}_v\boldsymbol{V}_{22}\|_F^2) \\ & +\lambda_{\mathrm{P}}\|\mathcal{P}(\boldsymbol{V}_{31})-\boldsymbol{D}\boldsymbol{V}_{32}\|_F^2+\lambda\|\boldsymbol{V}_4\|_1+\lambda_{\mathrm{T}}E_{\mathrm{T}}(\boldsymbol{V}_5)\end{aligned} \tag{5.16}$$

$$\text{subject to} \quad \begin{aligned}\boldsymbol{V}_1 &= \boldsymbol{X} \\ \boldsymbol{V}_{21} &= \boldsymbol{L}_h\boldsymbol{X} \\ \boldsymbol{V}_{22} &= \boldsymbol{L}_v\boldsymbol{X} \\ \boldsymbol{V}_{31} &= \boldsymbol{X} \\ \boldsymbol{V}_{32} &= \boldsymbol{\Omega} \\ \boldsymbol{V}_4 &= \boldsymbol{\Omega} \\ \boldsymbol{V}_5 &= \boldsymbol{X}.\end{aligned} \tag{5.17}$$

Note that a different splitting strategy could have been considered. However, the choice of the splittings (5.17) was motivated by the fact that it leads to simpler optimization sub-problems (see Section 5.3.2). For conciseness, we introduce the matrices $\boldsymbol{V}$, $\boldsymbol{H}$ and $\tilde{\boldsymbol{X}}$ such that

$$\boldsymbol{V}=\begin{bmatrix}\boldsymbol{V}_1^T\\ \boldsymbol{V}_{21}^T\\ \boldsymbol{V}_{22}^T\\ \boldsymbol{V}_{31}^T\\ \boldsymbol{V}_{32}^T\\ \boldsymbol{V}_4^T\\ \boldsymbol{V}_{5^T}\end{bmatrix},\ \boldsymbol{H}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{L}_h^T & \boldsymbol{0}\\ \boldsymbol{L}_v^T & \boldsymbol{0}\\ \boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\\ \boldsymbol{0} & \boldsymbol{I}\\ \boldsymbol{I} & \boldsymbol{0}\end{bmatrix} \text{ and } \tilde{\boldsymbol{X}}=\begin{bmatrix}\boldsymbol{X}^T\\ \boldsymbol{\Omega}^T\end{bmatrix}.$$

The augmented Lagrangian associated with the problem (5.16) is defined as follows

$$\mathcal{L}(\tilde{\boldsymbol{X}},\boldsymbol{V},\boldsymbol{G})=E(\boldsymbol{V})+\frac{\mu}{2}\|\boldsymbol{H}\tilde{\boldsymbol{X}}-\boldsymbol{V}-\boldsymbol{G}\|_F^2 \tag{5.18}$$

where $\boldsymbol{G}$ contains the Lagrange multipliers corresponding to $\boldsymbol{V}$, $\mu > 0$ is the penalty parameter and $E(\boldsymbol{V})$ is the cost function associated with the problem (5.16), *i.e.*,

$$\begin{aligned} E(\boldsymbol{V}) = \; & \tfrac{1}{2}\|\boldsymbol{Y} + \boldsymbol{A}\boldsymbol{V}_1\|_F^2 + \lambda_{\text{TV}}(\|\boldsymbol{L}_h\boldsymbol{V}_{21}\|_F^2 + \|\boldsymbol{L}_v\boldsymbol{V}_{22}\|_F^2) \\ & + \lambda_{\text{P}}\|\mathcal{P}(\boldsymbol{V}_{31}) - \boldsymbol{D}\boldsymbol{V}_{32}\|_F^2 + \lambda\|\boldsymbol{V}_4\|_1 + \lambda_{\text{T}}E_{\text{T}}(\boldsymbol{V}_5). \end{aligned} \tag{5.19}$$

Solving the problem (5.16) consists in alternating minimizations with respect to the motions $\boldsymbol{X}$, the sparse codes $\boldsymbol{\Omega}$ and the auxiliary variables $\boldsymbol{V}$. After these minimizations, the Lagrange multipliers $\boldsymbol{G}$ are updated. The different steps required to solve the problem (5.16) are summarized in Algorithm 5. Note that Algorithm 5 satisfies the conditions for the convergence of C-SALSA provided in [Afonso 2011]. More precisely, the matrix $\boldsymbol{H}$ is full column rank due to the presence of the identity matrices and the function E in (5.19) is closed, proper and convex. Further details about the solutions with respect to $\boldsymbol{X}$, $\boldsymbol{\Omega}$ and $\boldsymbol{V}$ are provided in the following section.

Algorithm 3: Motion estimation using C-SALSA

Input : Images $\boldsymbol{I}$, regularization parameters λ_{TV}, λ_{P}, λ, λ_T, penalty μ and initializations $\boldsymbol{V}^0$, $\boldsymbol{G}^0$, $\tilde{\boldsymbol{X}}^0$.

1 **for** $k = 1, ..., k_{\max}$ **do**

`%Motions and sparse codes`

2 $\tilde{\boldsymbol{X}}^k \in \underset{\tilde{\boldsymbol{X}}}{\text{argmin}}\ \mathcal{L}(\tilde{\boldsymbol{X}}, \boldsymbol{V}^{k-1}, \boldsymbol{G}^{k-1})$;

`%Proximal computations`

3 $\boldsymbol{V}^k \in \underset{\boldsymbol{V}}{\text{argmin}}\ \mathcal{L}(\tilde{\boldsymbol{X}}^k, \boldsymbol{V}, \boldsymbol{G}^{k-1})$;

`%Lagrange multipliers`

4 $\boldsymbol{G}^k = \boldsymbol{G}^{k-1} - (\boldsymbol{H}\tilde{\boldsymbol{X}}^k - \boldsymbol{V}^k)$;

5 **end**

Output: The motions $\boldsymbol{X}$ and the associated sparse codes $\boldsymbol{\Omega}$.

5.3.2 Details about the different minimizations

The minimizations with respect to $\boldsymbol{X}$ and $\boldsymbol{\Omega}$ are simple quadratic problems, whereas the solutions for the problems involving the auxiliary variables are provided by the corresponding Moreau proximity operators [Combettes 2005]. Note that the computations are straightforward for the variables $\boldsymbol{V}_{21}$, $\boldsymbol{V}_{22}$, $\boldsymbol{V}_4$ and $\boldsymbol{V}_5$. More specifically, the minimization with respect to $\boldsymbol{V}_4$ corresponds to a soft-thresholding operation and the other optimizations are simple problems, leading to closed-form expressions. However, the problems involving the variables $\boldsymbol{V}_1$, $\boldsymbol{V}_{31}$ and $\boldsymbol{V}_{32}$ are non trivial. More specifically, the minimization with respect to $\boldsymbol{V}_1$ results in a sparse linear system of equations that can be solved efficiently using numerical optimization methods, *e.g.*, the preconditioned

conjugate gradient (PCG) method [Kaasschieter 1988]. As shown in Section 5.4, running only a few iterations of the PCG algorithm allows an efficient optimization to be achieved. At this point it is interesting to mention that a linearized-ADMM [Lin 2011] approach could be considered for this minimization problem. More specifically, the latter strategy would allow the inversion of the operator $\boldsymbol{A}$ to be avoided by linearizing the corresponding penalty term and introducing an additional proximal term.

The minimization problem involving the term $\lambda_{\mathrm{P}}\|\mathcal{P}(\boldsymbol{V}_{31}) - \boldsymbol{D}\boldsymbol{V}_{32}\|_F^2$ requires the simultaneous minimization with respect to the non-separable auxiliary variables $\boldsymbol{V}_{31}$ and $\boldsymbol{V}_{32}$. One way of tackling this problem is to employ an approximate alternate minimization approach. The experimental results presented in Section 5.4 will show that only a few iterations are required to achieve an efficient optimization. Note that this choice was motivated by its simplicity. However, it is worth mentioning that it would be possible to consider a coupled formulation of the problem (5.15) using $\tilde{\boldsymbol{X}}$. The latter approach would allow a simultaneous minimization with respect to $\boldsymbol{X}$ and $\boldsymbol{\Omega}$ and the resulting auxiliary variables to be conducted. Further details about the beforementioned minimization problems and the corresponding solutions are provided in the rest of this section.

Minimizations with respect to $\boldsymbol{X}$ and $\boldsymbol{\Omega}$

The minimizations with respect to $\boldsymbol{X}$ and $\boldsymbol{\Omega}$ are quadratic problems, *i.e.,*

$$\begin{cases} \boldsymbol{X}^k \in \underset{\boldsymbol{X}}{\operatorname{argmin}} & \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{V}_1^{k-1} - \boldsymbol{G}_1^{k-1}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{L}_h\boldsymbol{X} - \boldsymbol{V}_{21}^{k-1} - \boldsymbol{G}_{21}^{k-1}\|_F^2 \\ & + \frac{\mu}{2}\|\boldsymbol{L}_v\boldsymbol{X} - \boldsymbol{V}_{22}^{k-1} - \boldsymbol{G}_{22}^{k-1}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{V}_{31}^{k-1} - \boldsymbol{G}_{31}^{k-1}\|_F^2 \\ & + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{V}_5^{k-1} - \boldsymbol{G}_5^{k-1}\|_F^2 \\ \boldsymbol{\Omega}^k \in \underset{\boldsymbol{\Omega}}{\operatorname{argmin}} & \frac{\mu}{2}\|\boldsymbol{\Omega} - \boldsymbol{V}_{32}^{k-1} - \boldsymbol{G}_{32}^{k-1}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{\Omega} - \boldsymbol{V}_4^{k-1} - \boldsymbol{G}_4^{k-1}\|_F^2 \end{cases} \tag{5.20}$$

where the superscript k denotes the current iteration. The solutions of (5.20) are as follows

$$\begin{cases} \boldsymbol{X}^k = & (3\boldsymbol{I} + \boldsymbol{L}_h^T\boldsymbol{L}_h + \boldsymbol{L}_v^T\boldsymbol{L}_v)^{-1}[(\boldsymbol{V}_1^{k-1} + \boldsymbol{G}_1^{k-1}) + (\boldsymbol{V}_{31}^{k-1} + \boldsymbol{G}_{31}^{k-1}) + (\boldsymbol{V}_5^{k-1} + \boldsymbol{G}_5^{k-1}) \\ & + \boldsymbol{L}_h^T(\boldsymbol{V}_{21}^{k-1} + \boldsymbol{G}_{21}^{k-1}) + \boldsymbol{L}_v^T(\boldsymbol{V}_{22}^{k-1} + \boldsymbol{G}_{22}^{k-1})] \\ \boldsymbol{\Omega}^k = & \frac{1}{2}[(\boldsymbol{V}_{32}^{k-1} + \boldsymbol{G}_{32}^{k-1}) + (\boldsymbol{V}_4^{k-1} + \boldsymbol{G}_4^{k-1})]. \end{cases} \tag{5.21}$$

Since the gradient operators $\boldsymbol{L}_h$ and $\boldsymbol{L}_v$ represent 2D cyclic convolutions leading to block cyclic matrices, the minimization with respect to $\boldsymbol{X}$ can be solved efficiently using the Fast Fourier Transform (FFT) (see [Afonso 2011]. Note that the inverse $(3\boldsymbol{I} + \boldsymbol{L}_h^T\boldsymbol{L}_h + \boldsymbol{L}_v^T\boldsymbol{L}_v)^{-1}$ can be computed beforehand.

Minimization with respect to $\boldsymbol{V}_1$

The minimization with respect to $\boldsymbol{V}_1$ can be written as follows

$$\boldsymbol{V}_1{}^k \in \underset{\boldsymbol{V}_1}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{Y} + \boldsymbol{A}\boldsymbol{V}_1\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X}^k - \boldsymbol{V}_1 - \boldsymbol{G}_1{}^{k-1}\|_F^2. \tag{5.22}$$

The problem (5.22) results in the following sparse linear system of equations

$$(\boldsymbol{A}^T\boldsymbol{A} + \mu\boldsymbol{I})\boldsymbol{V}_1^k = \mu(\boldsymbol{X}^k - \boldsymbol{G}_1^{k-1}) - \boldsymbol{A}^T\boldsymbol{Y} \tag{5.23}$$

where $\boldsymbol{A}^T\boldsymbol{A}$ and can be computed only once and the matrix $\boldsymbol{A}^T\boldsymbol{A}+\mu\boldsymbol{I}$ is full rank due to the presence of the identity matrix $\boldsymbol{I}$. The system (5.23) can be solved using a gradient descent method. In this work, we perform a few steps of the preconditioned conjugate gradient (PCG) algorithm.

Minimization with respect to $\boldsymbol{V}_{21}$ and $\boldsymbol{V}_{22}$

The minimization problem involving the auxiliary variables $\boldsymbol{V}_{21}$ and $\boldsymbol{V}_{22}$ is formulated as follows

$$\begin{aligned}\boldsymbol{V}_{21}^k, \boldsymbol{V}_{22}^k \in \underset{\boldsymbol{V}_{21},\boldsymbol{V}_{22}}{\operatorname{argmin}} \lambda_{\mathrm{TV}}(\|\boldsymbol{V}_{21}\|_F^2 + \|\boldsymbol{V}_{22}\|_F^2) \quad &+\tfrac{\mu}{2}\|\boldsymbol{L}_h\boldsymbol{X}^k - \boldsymbol{V}_{21} - \boldsymbol{G}_{21}^{k-1}\|_F^2 \\ &+\tfrac{\mu}{2}\|\boldsymbol{L}_v\boldsymbol{X}^k - \boldsymbol{V}_{22} - \boldsymbol{G}_{22}^{k-1}\|_F^2.\end{aligned} \tag{5.24}$$

The above problem has the solutions

$$\begin{cases} \boldsymbol{V}_{21}^k = \frac{\mu}{2\lambda_{\mathrm{TV}}+\mu}(\boldsymbol{L}_h\boldsymbol{X}^k - \boldsymbol{G}_{21}^{k-1}) \\ \boldsymbol{V}_{22}^k = \frac{\mu}{2\lambda_{\mathrm{TV}}+\mu}(\boldsymbol{L}_v\boldsymbol{X}^k - \boldsymbol{G}_{22}^{k-1}) \end{cases} \tag{5.25}$$

where the terms $\boldsymbol{L}_h\boldsymbol{X}^k$ and $\boldsymbol{L}_v\boldsymbol{X}^k$ are computed using the FFT.

Minimization with respect to $\boldsymbol{V}_{31}$ and $\boldsymbol{V}_{32}$

The minimization problem with respect to $\boldsymbol{V}_{31}$ and $\boldsymbol{V}_{32}$ is formulated as follows

$$\begin{aligned}\boldsymbol{V}_{31}^k, \boldsymbol{V}_{32}^k \in \underset{\boldsymbol{V}_{31},\boldsymbol{V}_{32}}{\operatorname{argmin}} \lambda_{\mathrm{P2}}\|\mathcal{P}(\boldsymbol{V}_{31}) - \boldsymbol{D}\boldsymbol{V}_{32}\|_F^2 \quad &+\tfrac{\mu}{2}\|\boldsymbol{V}_{31} - (\boldsymbol{X}^k - \boldsymbol{G}_{31}^{k-1})\|_F^2 \\ &+\tfrac{\mu}{2}\|\boldsymbol{V}_{32} - (\boldsymbol{\Omega}^k - \boldsymbol{G}_{32}^{k-1})\|_F^2.\end{aligned} \tag{5.26}$$

Since the problem (5.26) involves the non-separable variables $\boldsymbol{V}_{31}$ and $\boldsymbol{V}_{32}$ we alternate minimizations with respect to each variable. More specifically, the solutions for each

step j are obtained as follows

$$\begin{cases} \boldsymbol{V}_{31}^{k,j} = & (2\lambda_{\mathrm{P}}\beta + \mu\boldsymbol{I})^{-1}[2\lambda_{\mathrm{P}}\mathcal{P}^{*}(\boldsymbol{D}\boldsymbol{V}_{32}^{k,j-1}) + \mu(\boldsymbol{X}^{k} - \boldsymbol{G}_{31}^{k-1})] \\ \boldsymbol{V}_{32}^{k,j} = & (2\lambda_{\mathrm{P}}\boldsymbol{D}^{T}\boldsymbol{D} + \mu\boldsymbol{I})^{-1}[2\lambda_{\mathrm{P}}\boldsymbol{D}^{T}\mathcal{P}(\boldsymbol{V}_{31}^{k,j}) + \mu(\boldsymbol{\Omega}^{k} - \boldsymbol{G}_{32}^{k-1})] \end{cases} \tag{5.27}$$

where j indicates the current step, $\boldsymbol{V}_{32}^{k,j-1}$ is computed at the previous step and the term β is a constant weighting factor depending on the number of times each pixel is considered. The term $\mathcal{P}^{*}(\boldsymbol{D}\boldsymbol{V}_{32}^{k,j-1})$ contains the reconstructions of the motion fields $\boldsymbol{X}$ from the sparse representations obtained at step $j-1$. The matrix $(2\lambda_{\mathrm{P}}\boldsymbol{D}^{T}\boldsymbol{D} + \mu\boldsymbol{I})$ is invertible (due to the presence of the identity matrix $\boldsymbol{I}$) and its inverse can be computed in advance.

Minimization with respect to $\boldsymbol{V}_4$

The solution with respect to $\boldsymbol{V}_4$ corresponds to the following minimization problem

$$\boldsymbol{V}_{4}^{k} \in \underset{\boldsymbol{V}_4}{\operatorname{argmin}}\ \lambda\|\boldsymbol{V}_4\|_1 + \frac{\mu}{2}\|\boldsymbol{V}_4 - (\boldsymbol{\Omega}_k - \boldsymbol{G}_4^{k-1})\|_F^2. \tag{5.28}$$

The solution of (5.28) is obtained using the soft-threshold operator, *i.e.,*

$$\boldsymbol{V}_{4}^{k} = \operatorname{sign}(\boldsymbol{\Omega}_k - \boldsymbol{G}_4^{k-1})\max\left\{0, \|\boldsymbol{\Omega}_k - \boldsymbol{G}_4^{k-1}\|_F - \frac{\lambda}{\mu}\right\}. \tag{5.29}$$

Minimization with respect to $\boldsymbol{V}_5$

In order to compute the auxiliary variable corresponding to the temporal regularization term (5.13), we use the motions resulting from the previous iteration $k-1$. Since the previous, current and next frames are fixed from the previous iteration, the minimization problems with respect to $\boldsymbol{V}_5$ can be formulated as follows

$$\boldsymbol{V}_{5}^{k} \in \underset{\boldsymbol{V}_5}{\operatorname{argmin}}\ \lambda_{\mathrm{T}}\|\boldsymbol{W}_{\mathrm{L}}^{1/2}(2\boldsymbol{V}_5 - \boldsymbol{B})\|_F^2 + + \frac{\mu}{2}\|\boldsymbol{V}_5 - (\boldsymbol{X}^{k} - \boldsymbol{G}_5{}^{k-1})\|_F^2 \tag{5.30}$$

where $B = \boldsymbol{X}_b - 2\boldsymbol{L}\boldsymbol{X}_b^{T}\boldsymbol{X}_b + \boldsymbol{X}_f''$ results from the Taylor expansion introduced in (5.11) and is computed using the motions of the previous iteration $k-1$. The corresponding solution is as follows

$$\boldsymbol{V}_{5}^{k} = (8\lambda_{\mathrm{T}}\boldsymbol{W}_{\mathrm{L}} + \mu\boldsymbol{I})^{-1}[4\lambda_{\mathrm{T}}\boldsymbol{W}_{\mathrm{L}}\boldsymbol{B} + \mu(\boldsymbol{X}^{k} - \boldsymbol{G}_5^{k-1})] \tag{5.31}$$

where the weights $\boldsymbol{W}_{\mathrm{L}}$ are computed using the residuals of the previous iteration as explained in Section . Note that, in our experiments, the temporal regularization is introduced after a few iterations. This strategy provides a more accurate initialization

for the matrix $\boldsymbol{B}$. However, note also that it would be possible to perform the temporal regularization starting from the first iteration of the algorithm using, *e.g.*, a varying value for the penalty parameter μ.

5.4 Experimental results

This section evaluates the proposed groupewise motion estimation method using the LADprox sequence already considered in Section 3.5.5 [Alessandrini 2016b]. For the purpose of this preliminary study, the proposed method is compared to the pairwise motion estimation method based on a sparse regularization (denoted as NR for non-robust) introduced in Chapter 3 and the robust approach of Chapter 4 (referred to as R for robust). In order to evaluate the interest of the proposed temporal regularization, a comparison with the groupewise OF method using only spatial and sparse regularizations (*i.e.*, with $\lambda_{\mathrm{T}} = 0$) is also included. The motion estimation accuracy is evaluated in terms of the endpoint error (see Section 3.5.2). In addition, we compute the endpoint error between the estimated and groundtruth trajectories, *i.e.*,

$$\epsilon_t(i,t) = \sqrt{[\boldsymbol{x}(i,t) - \hat{\boldsymbol{x}}(i,t)]^2 + [\boldsymbol{y}(i,t) - \hat{\boldsymbol{y}}(i,t)]^2} \tag{5.32}$$

where $\boldsymbol{x}(i,t), \boldsymbol{y}(i,t)$ and $\hat{\boldsymbol{x}}(i,t), \hat{\boldsymbol{y}}(i,t)$ are the true and estimated horizontal and vertical coordinates of pixel i in the tth frame.

5.4.1 Regularization and optimization parameters

The regularization parameters providing the smallest endpoint errors for the motions of the considered sequence were selected using cross-validation. The final parameters for the proposed method are provided in Table 5.1 where c denotes the threshold parameter for the Lorentzian weight function. Note that since the displacements of the first and last frames of the LADprox sequence do not match, the temporal regularization parameter was set to $\lambda_{\mathrm{T}} = 0$ for these two frames (*i.e.*, no cyclic motions). Note that the dictionary learning parameters and regularization parameters used for the methods NR and R have been provided in Sections 3.5.3 and 4.5.2.1.

Parameter	λ_{TV}	λ_{P}	λ	λ_{T}	c
Value	2.10^{-1}	10^{-4}	10^{-4}	10^{-2}	2.10^{-2}

Table 5.1: Regularization parameters for the proposed motion estimation method.

The proposed optimization strategy based on the C-SALSA method was run using 100 iterations, with a penalty parameter $\mu = 10^{-2}$. The algorithm was stopped after

the convergence of all auxiliary variables [Simões 2015], *i.e.*,

$$\begin{cases} \|\boldsymbol{\Omega} - \boldsymbol{V}_j\|_F < \tau, & \text{for } j \in \{4, 32\} \\ \|\boldsymbol{X} - \boldsymbol{V}_j\|_F < \tau, & \text{otherwise} \end{cases} \tag{5.33}$$

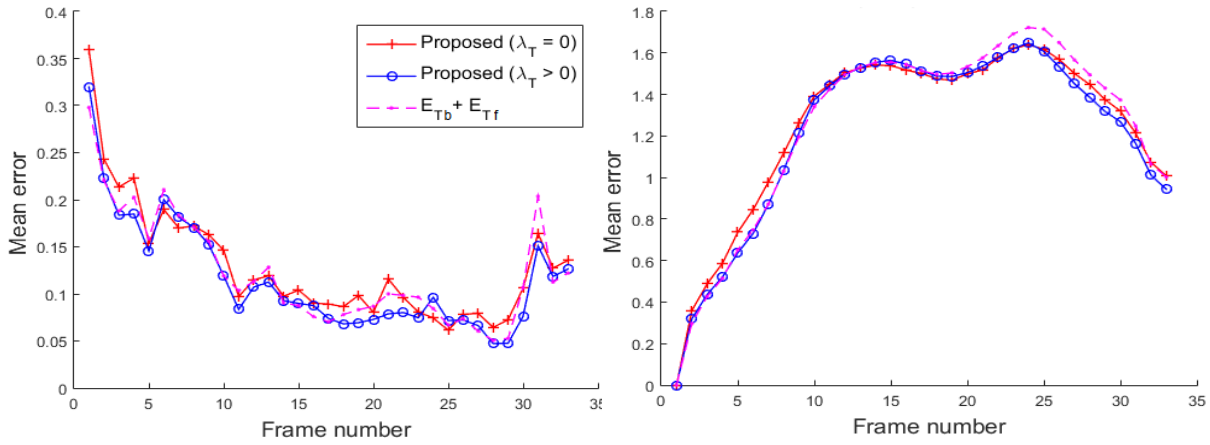
where τ is a predefined threshold set to $\tau = 10^{-3}$ in our experiments. The minimization with respect to the auxiliary variable $\boldsymbol{V}_1$ was conducted by running 20 iterations of the PCG algorithm. Finally, the alternate minimization with respect to $\boldsymbol{V}_{31}$ and $\boldsymbol{V}_{32}$ was conducted using 2 iterations.

5.4.2 Temporal regularization

The global endpoint errors for the LADprox sequence are reported in Table 5.2. The evolution of the mean displacement and trajectory errors for all the frames of the considered sequence are also shown in Fig. 5.1. The errors provided by the temporal regularization term (5.8) are shown for comparison (referred to as $E_{T,b} + E_{T,f}$). These errors show that the proposed regularization (5.13) provides smaller global endpoint errors. The time evolution of the errors shows that the temporal regularization using two separate terms as in (5.8) is outperformed by the proposed regularization in terms of displacement and trajectory errors, particularly, for the frames corresponding to the diastole phase of the cardiac cycle. In order to gain more insight into the effects of these two regularizations, Fig. 5.2 shows the obtained displacements for a point in the mid segment of the myocardium. This figure shows that the regularization with two separate terms (*i.e.*, $E_{T,b} + E_{T,f}$) leads to piecewise constant motions with more discontinuities than the proposed temporal regularization. Note that the two regularization approaches were weighted as explained in Section 5.3.2. These findings support the fact that the proposed regularization is more adapted to the motions in the considered data.

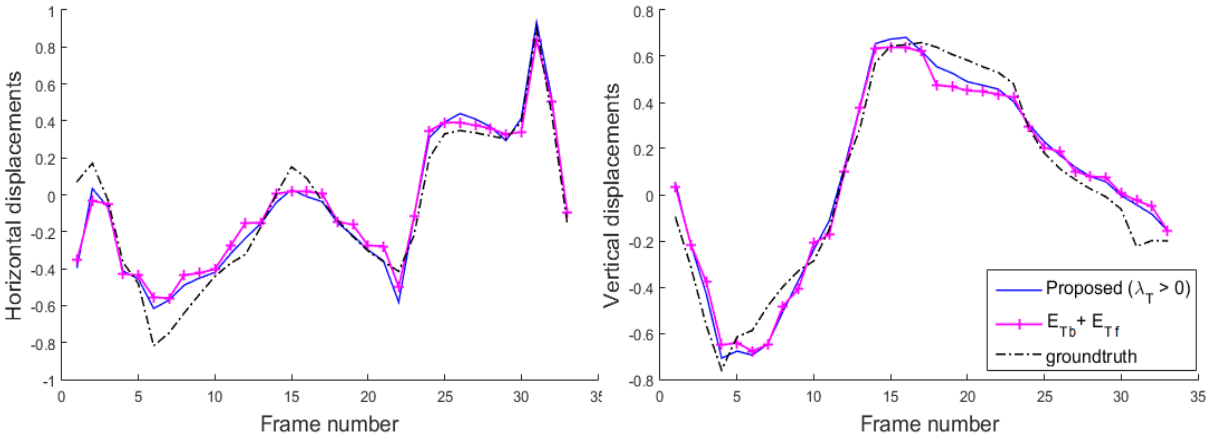
Method	NR	R	Proposed ($\lambda_{\mathrm{T}} = 0$)	Proposed ($\lambda_{\mathrm{T}} \neq 0$)	$E_{T,b} + E_{T,f}$
Displacements (ϵ)	0.14±0.11	**0.10±0.089**	0.12±0.11	0.11±0.10	0.12±0.10
Trajectories (ϵ_t)	1.74±1.36	**1.26**±1.09	1.28.±1.04	**1.26**±1.04	1.29±**1.03**

Table 5.2: Means ± stds of the global endpoint errors for the LADprox sequence.



(a) Displacement errors (b) Trajectory errors

Figure 5.1: Mean endpoint errors of the proposed method with and without temporal regularization for the displacements and trajectories of the LADprox sequence.



(a) (b)

Figure 5.2: Comparison between the displacements of a point in the mid segment of the myocardium for the proposed temporal regularization and the separated temporal regularization term (5.8).

In order to analyze the effect of the proposed temporal regularization we show the obtained displacements for two points on the myocardial wall when compared to the proposed groupewise approach (i) without temporal regularization and (ii) with the proposed temporal regularization without weighting (*i.e.*, $\boldsymbol{W}_{\mathrm{L}} = \boldsymbol{I}$). In particular, Fig 5.3 shows the displacements for two points in the mid and basal segments of the myocardium. This figure shows that the absence of temporal regularization leads to noisy displacements for the selected points. It is also clear from this figure that the temporal regularization without weighting yields smooth motions but does not capture

the discontinuities present in the groundtruth. Thus, the temporal regularization results in over-smoothing for some frames of the sequence. On the other hand, the proposed weighting allows us to relax the temporal smoothness for the frames corresponding to these discontinuities, resulting in piecewise smooth displacements.

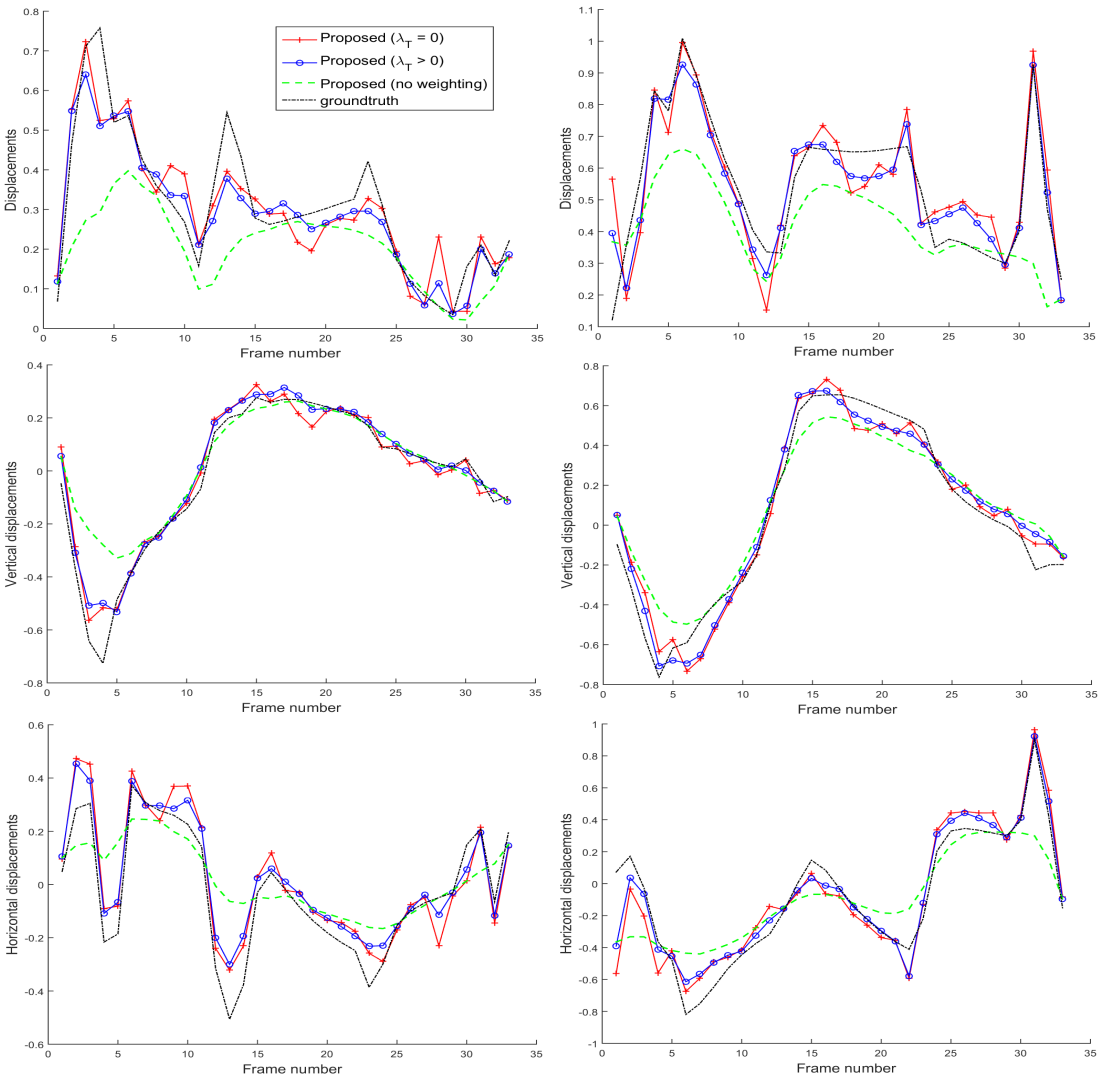
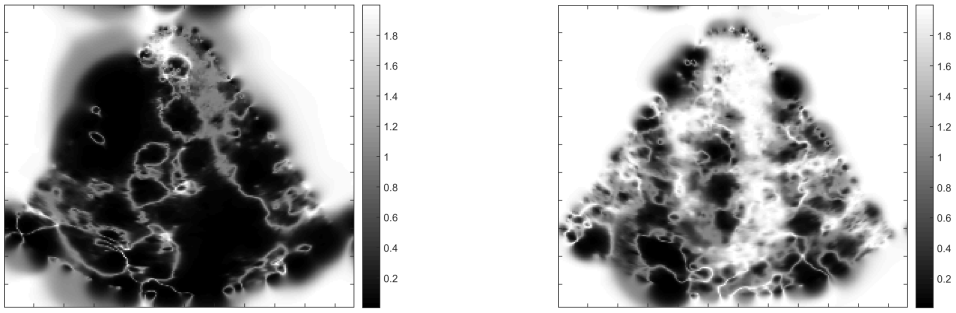


Figure 5.3: Displacements for two points in (a) the mid and (b) basal segments of the heart.

Fig. 5.4 shows an example of the final weights for a frame with a temporal discontinuity (*i.e.*, frame 4) and a frame with temporally smooth displacements (*i.e.*, frame 19). Note that the displayed weights are the sums of the horizontal and vertical weights. This figure shows that low weights are assigned to the pixels in the 4th frame allowing

the temporal regularization to be mitigated. In contrast, the pixels corresponding to the myocardium region in the 19th frame do not produce low weights (*i.e.*, the weights are close to 1), leading to a temporal regularization for these pixels. As shown in Figs. 5.3 and 5.4, the proposed weighting strategy clearly allowed temporal discontinuities to be incorporated. However, it is worth mentioning at this point that it would be necessary to further investigate the effects of a pixel-wise weighting approach for 2D US, since it could enforce discontinuities due to speckle decorrelations. A different way of weighting the temporal regularization term would consist in assigning the same weight to all the pixels of one frame using, for example, the pixel-wise weights themselves or *a priori* information about the time instants corresponding to the temporal discontinuities (*e.g.*, the end systole frame). We think that this idea deserves to be investigated in future work.



(a) (b)

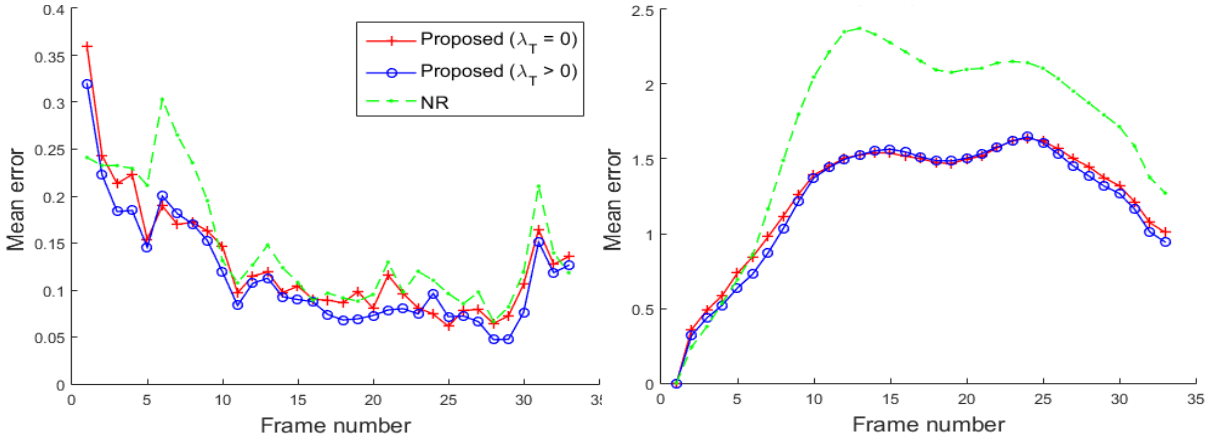
Figure 5.4: Resulting weights for (a) the 4th and (b) 19th frames of the LADprox sequence.

Finally, Table 5.2 shows that the proposed temporal regularization does not provide a significant gain in terms of displacement accuracies when compared to the proposed groupewise approach without a temporal regularization (*i.e.*, $\lambda_{\mathrm{T}} = 0$). The observed gain in accuracy for the resulting trajectories is also small when enforcing temporal piecewise smoothness. However, we think that further investigations using, for example, *in vivo* data would be necessary to evaluate the proposed temporal regularization. In fact, real data containing more speckle decorrelations and out-of-plane motions could further benefit from the temporal smoothness assumption as mentioned in [Ledesma-Carbayo 2005]. It would also be necessary to compare the performance of the proposed temporal regularization with other state-of-the-art spatio-temporal motion estimation methods and conduct tests using other sequences of the realistic simulations dataset [Alessandrini 2016b].

5.4.3 Comparison with the NR method

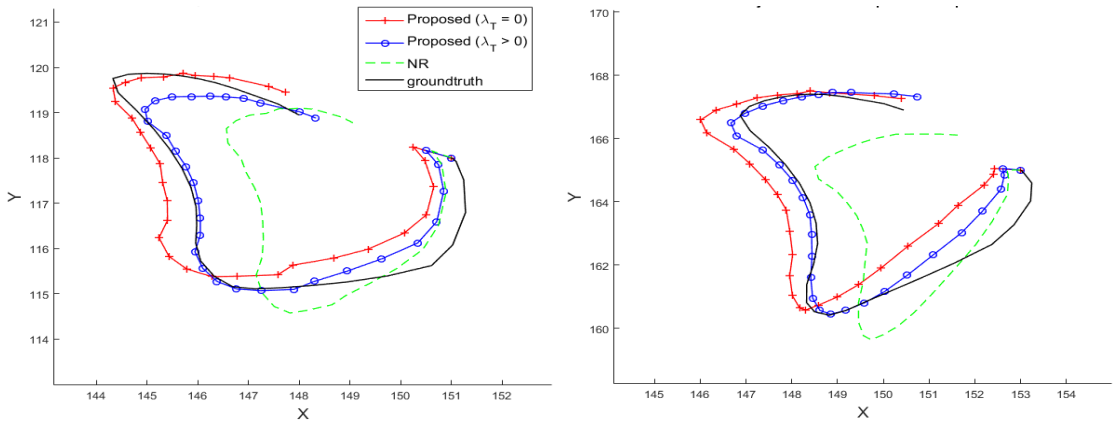
This subsection investigates the benefits of the proposed motion estimation method with temporal regularization when compared to the NR method presented in Chapter 3. The global endpoint errors for the proposed and NR methods are provided in Table 5.2. We

can see that the proposed groupewise approach provides smaller displacement and trajectory errors when compared to the NR method. The time evolution of the endpoint errors during the cardiac cycle are shown in Fig. 5.5. This figure shows that the proposed groupewise method results in smaller mean errors for almost all the frames of the sequence when compared to the NR method. In particular, a significant gain in trajectory accuracy can be observed starting from the 5th frame of the sequence. Fig. 5.6 shows an example of the obtained trajectories for two points in the mid and basal segments of the myocardium. Note that the trajectories are smooth for all the methods due to the interpolation used to compute these trajectories. These results clearly show that the proposed temporal consistent approach can be beneficial in terms of motion estimation accuracy when compared to the nonrobust pairwise method NR.



(a) Displacement errors (b) Trajectory errors

Figure 5.5: Mean endpoint errors of the proposed and NR methods.

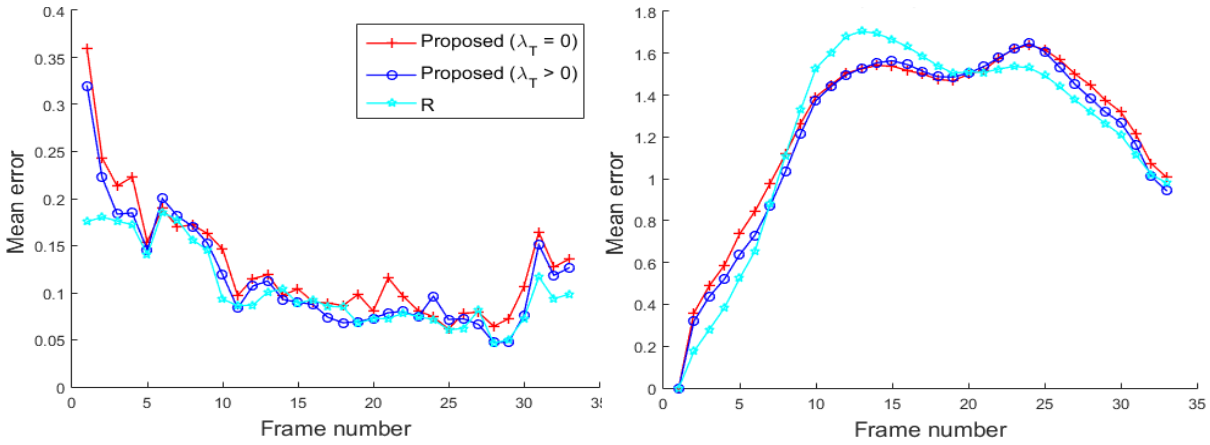


(a) Mid (b) Base

Figure 5.6: Examples of trajectories obtained for two points in the myocardium.

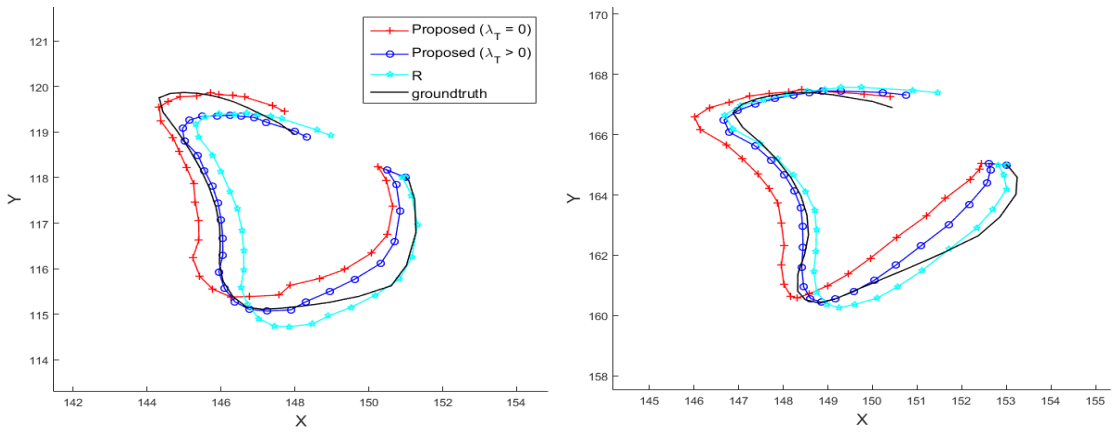
5.4.4 Comparison with the robust R method

In order to further investigate the interest of the proposed method, the motion estimation accuracy is compared to the robust method of Chapter 4. The global endpoint errors of the R method are reported in Table 5.2. This table shows that the best performance in terms of displacement accuracy is provided by the robust method. In terms of trajectory accuracy, comparable errors were obtained for both approaches. The evolution of these errors versus time is shown in Fig. 5.7. We can see that the robust approach provides the best mean displacement errors, in particular, at the beginning of the sequence. Furthermore, the R method outperformed the proposed method in terms of trajectory accuracy at the beginning and end of the sequence, while being less accurate during end systole and beginning of diastole phases (*i.e.,* frames 7-19). Two examples of the obtained trajectories are shown in Fig. 5.8 for two points in the mid and basal segments of the myocardium. This figure confirms that the proposed temporal regularization results in more accurate trajectories in the mid-part of the sequence. Again, the trajectories are smooth for all the methods due to the interpolation used to compute these trajectories.



(a) Displacement errors (b) Trajectory errors

Figure 5.7: Mean endpoint errors of the proposed and R methods for the displacements and trajectories of the LADprox sequence.



(a) Mid (b) Base

Figure 5.8: Examples of trajectories obtained for two points in the myocardium using the proposed and R methods.

Thus, the results obtained with the proposed method did not show a significant gain in performance when compared to the robust approach of Chapter 4. More precisely, the results showed that both methods can have a similar impact on the estimated motions. This is possibly due to the fact that the temporal regularization reduces the sensitivity of motion estimation to noise and discrepancies in the data, leading to a robust estimation. Moreover, the groupewise estimation approach can also result in more stable estimates when compared to the pairwise approach (as shown in Section 5.4.3). At this point, it is worth pointing out that the R method relies strongly on the tuning of several parameters, in particular, the threshold parameters for the three weight functions. In contrast, the adjustment of the parameters was simpler for the proposed temporally consistent approach approach, which is clearly an interesting property.. It is also interesting to note that the robust method is more computationally intensive than the proposed groupewise estimation method. Moreover, the proposed global implementation strategy also provided faster execution times when compared to the NR method of Chapters 3 as shown in Table 5.3. Finally, it is worth mentioning that the proposed framework would allow us to introduce robustness using, *e.g.*, the weighted approach presented in Chapter 4 or robust norms (with their corresponding proximal operators) for all the proposed energy terms.

Method	NR	R	Proposed ($\lambda_T = 0$)	Proposed ($\lambda_T \neq 0$)
Time (sec)	52.33	87.32	10.27	12.54

Table 5.3: Mean execution times (in seconds) for one frame of the LADprox sequence.

5.5 Conclusions and perspectives

This chapter presented a groupewise motion estimation method for 2D cardiac US sequences. The proposed framework allowed the motions of an entire sequence of ultrasound images to be estimated simultaneously, while incorporating the sparse regularization term promoted in Chapter 3 as well as an additional regularization term enforcing temporal smoothness. The aim of this chapter was to analyze the interest of a temporally consistent motion estimation approach when compared to the methods introduced in the previous Chapters 3 and 4. More precisely, the benefits of a temporal regularization of the cardiac motion estimation problem was investigated. A preliminary study allowed us to appreciate some benefits of the proposed method, in particular, an improvement of the motion estimation accuracy when compared to the non-robust pairewise method of Chapter 3. The proposed C-SALSA optimization strategy also led to faster execution times when compared to the two methods considered for comparison. However, the benefits of the temporal consistent approach were unclear with respect to the robust method of Chapter 4. Thus, it would be necessary to conduct further experimentations using, for example, real 2D US data. More specifically, 2D *in vivo* images contain more speckle decorrelations and out-of-plane motions, allowing us to further investigate the benefits of the proposed temporal regularization term. Testing the proposed method using a dataset with cyclic motions would also be more appropriate for analysing the effects of temporal consistency for cardiac motion patterns. Finally, it would be interesting to compare the proposed approach to other state-of-the-art groupewise and temporal smoothness-based methods.

For future work, it would be possible to incorporate a cardiac motion specific temporal regularization in the proposed C-SALSA framework. More precisely, it would be possible to learn a temporal or spatio-temporal dictionary, *i.e.*, encoding usual cardiac trajectory patterns, including discontinuities. Another interesting prospect would be to robustify the proposed method by replacing the ℓ_2-norm with robust alternatives. Finally, some issues related to the proposed implementation strategy would deserve to be considered. More specifically, the alternate minimization between the motions and the sparse codes as well as the solving of the linear system of equations resulting from the OF-based data fidelity term could be handled differently. These challenges could be addressed using, for example, a coupled variable approach and linearized-ADMM method [Lin 2011].

Chapter 6

Conclusions and Perspectives

Cardiac motion estimation is of critical importance for the diagnosis of cardiovascular diseases. However, achieving a reliable cardiac function assessment is a difficult task, making the development of efficient motion estimation methods still an open problem. In this context, the objective of this PhD thesis was to propose novel approaches for motion estimation in 2D echocardiography. The guiding principle of this work was the use of dictionary learning for motion estimation. Each of the presented contributions approached this problem from a different perspective. More specifically, our first work developed a novel regularization exploiting the fact that most motions of interest can be decomposed in an appropriate dictionary of typical displacements. A robustification of the first algorithm was then proposed in order to mitigate the impact of abnormal motions. Finally, the temporal aspect of ultrasound image sequences was considered. The interest of the different contributions was supported by experiments using simulated images and *in vivo* sequences of healthy and diseased patients.

The first work focused on developing a regularization based on a sparse representation in a dictionary of typical motions. The classical spatial constraints can lead to over-smoothing and loss of the local structures of motions, which are crucial for the analysis of cardiac function. As an alternative, a dictionary learning-based regularization was proposed. More specifically, the introduced regularization strategy promoted the sparsity of the displacements when decomposed on a learned dictionary of cardiac motions. The dictionary was learned using small patches of realistic simulated cardiac motions, allowing local structures to be captured. The experimental results showed that by incorporating learned patterns of motions into the regularization strategy, one could prevent over-smoothing and recover more structures locally. These findings were supported by competitive results in terms of motion estimation accuracy when compared to state-of-the-art methods. Moreover, a reduced sensitivity to large displacements was observed, which was an interesting property of the proposed method. Overall, the obtained results demonstrated that a learning-based regularization is an interesting alternative to the classical spatial constraints. However, some questions remain regarding how the motion dictionaries should be learned. The present work showed that learning the dictionary using one pathological sequence led to a gain in performance. Nonetheless, further investigations are still necessary to show how different pathologies are impacted by the choice of the learning dataset.

The second contribution of this thesis is related to the problem of outliers in 2D ultrasound imaging. From a general point of view, motion estimation is a difficult task because it is inherently affected by outliers, such as motion boundaries. These native outliers are compounded by several limitations proper to 2D echocardiography, such as acquisition artefacts, image noise and random motions in the blood. Taking into consideration these drawbacks of ultrasound imaging, the second work aimed at robustifying the cardiac motion estimation algorithm introduced in the first chapter. In order to mitigate the impact of outliers, a robust iteratively reweighted method was introduced. More precisely, the weights were computed directly from the estimated motions. Thus, the proposed method did not require a preliminary intervention, such as a segmentation of the myocardium, to handle outliers. The experimental results clearly

showed that introducing robustness led to a gain in the estimation accuracy. Moreover, the presented tests highlighted the interest of jointly robustifying the data fidelity and regularization terms. However, the proposed method relied on the tuning of several parameters and the execution times were higher when compared to the other algorithms presented in this thesis. Therefore, more thought should be given to the interactions between the different terms when adjusting these parameters as well as to the efficiency of the optimization strategy. One possible way of alleviating the computational cost and the difficulties related to the parameter setting, would be to set fixed thresholds for the weight computations. More specifically, one could use *a priori* knowledge about the nature of outliers, *e.g.*, related to the motion variability associated with anatomical boundaries.

The objective of the final contribution was to analyze the interest of a temporal regularization. Since cardiac motion estimation aims at quantifying the deformation of the tissues in both space and time, it seemed appropriate to exploit the temporal information embedded in ultrasound image sequences. More precisely, a comparison with the previously introduced methods was presented in a preliminary experimental study. The novelties of this approach were twofold. First, a groupewise estimation framework allowed all the motions of an image sequence to be estimated simultaneously by minimization of an appropriate global cost function. A C-SALSA-based algorithm was proposed to solve this optimization problem. Secondly, this global estimation strategy allowed us to incorporate a weighted temporal regularization term. The experimental results showed some benefits of introducing a temporal regularization, particularly, with respect to the first non-robust method presented in this thesis. The comparison with the robust approach promoted in the second contribution showed that the proposed method can have a similar impact on the estimation, which is a property that would deserve further investigations. It is worth noting that the proposed method presented the advantages of having a relatively simple parameter tuning with faster execution times. However, the benefits in terms of accuracy were not clear and would deserve further investigations in future experiments using, for example, real and cyclic cardiac sequences . Moreover, the introduced weighting approach would deserve further investigations. The proposed framework being flexible, alternative temporal regularizations would be possible to consider in future investigations. Possible ways of improvement will be further discussed in the next part of this chapter.

The next part of this chapter presents some issues and perspectives that would deserve to be investigated in future work.

Emerging ultrasound imaging techniques

The frame rate is of key importance when quantifying the motion of rapidly moving organs. This is particularly true for the heart, whose contraction occurs in a relatively short period of time. In standard echocardiography, these large systolic motions are less accurately evaluated by state-of-the-art motion estimation methods. Thus, it would be interesting to extend the presented contributions to new ultrasound imaging techniques that allow these fast motions to be captured more efficiently. More specifically, recent advances have made it possible to reach ultrafast frame rates (> 1000 frames per second) by means of the so-called plane wave or diverging wave imaging techniques [Tanter 2014]. The feasibility of cardiac motion estimation from this type of high-frame-rate images has been shown in [Joos 2018]. In this work, accurate motion estimates are obtained with standard block-matching methods using diverging wave-based images. However, it is worth noting that new challenges would arise when using ultrafast ultrasound data. For example, it would be necessary to deal with the limitations associated with the decrease in the spatial resolution of the images.

Taking into account the increased use of 3D ultrasound imaging, it is also worth considering an extension of the proposed methods to 3D. This thesis has focused on 2D echocardiography, which despite the advantages of its relatively high temporal resolution and small data size, can present some shortcomings. In particular, the problems of out-of-plane motions and incomplete geometrical information could be bypassed with 3D ultrasound imaging. At this point, it is worth mentioning that the data fidelity and regularization terms employed in the proposed methods are not inherently limited to 2D. Nevertheless, it should be pointed out that motion estimation from 3D ultrasound images would be affected by new limitations, such as lower frame rate and spatial resolution in the azimuthal direction.

Dictionary learning strategies

In this work, we have mainly addressed the problem of offline dictionary learning. However, a framework including a dictionary update step using the current estimation was proposed. Preliminary experimental results have not brought conclusive evidence for the interest of this online learning approach. However, we think that the latter strategy could be a relevant choice for cardiac motion estimation, especially when lacking appropriate or sufficient training data. Moreover, the choice of a learning dataset may not be straightforward, for example, when dealing with distinct pathologies. Online learning would allow us to bypass this difficulty since the dictionary would be adapted to the estimated motions themselves. Thus, it seems necessary to further investigate the online strategy, especially using *in vivo* data, where the motions are potentially far from the patterns of a predefined simulation dataset. At this point, it is worth mentioning that an

online learning strategy can be more affected by erroneous motion estimates. Therefore, a robust learning strategy would be worth considering [Lu 2013]. More specifically, when learning a dictionary using estimated cardiac motions fields, it would be necessary to mitigate the impact of random motions in the blood. Finally, note that improving the computational cost of the online learning strategy would also deserve to be considered in future work.

In this thesis, the dictionaries were learned separately for the horizontal and vertical motion fields. However, other ways of learning the dictionary could have been possible. For example, a dictionary could be learned jointly for the two directions (or three directions in the 3D case). The differences between the latter approach and the proposed dictionary learning method have not been investigated in this thesis, but would also deserve consideration in future work. Nonetheless, note that separating the horizontal and vertical components would allow a faster parallel computing strategy to be implemented. Furthermore, the proposed dictionary has not exploited the temporal properties of cardiac motion. Another prospect would consist in integrating this aspect by learning motion dictionaries that take into account the sparsity of the motion versus time. The latter point is developed in more detail in the following subsection.

Temporal regularization

The last contribution of this thesis introduced a temporal regularization strategy for cardiac motion estimation. The major limitation of smoothness-based temporal regularizations is that they are very prone to over-smoothing. As a consequence, it is crucial to take into account temporal discontinuities. In this thesis, an iterative reweighting of the temporal regularization term was introduced, allowing us to reduce the smoothing at temporal discontinuities. However, it is worth mentioning that a pixel-wise weighting is affected by discrepancies in the image intensities. Therefore, it would be necessary to combine such a weighted approach with a robust data fidelity term that would mitigate the effect of data outliers. However, the benefits of the temporal regularization in this scenario remain unclear. An ideal solution would discard discrepancies in the intensities and still allow for temporal discontinuities that originate from the motions themselves. A possible way of achieving this compromise is by introducing *a priori* knowledge about the time instants corresponding to the temporal discontinuities (*e.g.*, the end systole frame). However, such information may not be available. Thus, an alternative approach would be to use an adaptive temporal regularization. More specifically, one could learn a temporal or spatio-temporal dictionary that encodes typical cardiac trajectories, including discontinuities. Considering the cyclic nature of cardiac motion, exploiting the sparsity of the motion in time is clearly an interesting prospect. Note that sparsity in the temporal domain has been investigated for cardiac motion in [Shi 2013]. In this work, sparsity is imposed in the parametric space of a temporal free form deformation model. However, it should be pointed out that this method is not based on a learning approach as suggested herein. Finally, it is interesting to note that the proposed C-SALSA-based framework would allow such a cardiac motion specific temporal regularization to be

incorporated.

Exploiting the sparse codes, dictionaries and robust weights

The primary goal of cardiac motion estimation is the quantification of the myocardial displacements. In the context of medical imaging, some works have investigated methods that perform motion estimation jointly with other problems, for example, segmentation [Onofrey 2015, Lu 2012, Vemuri 2003]. The motion estimates obtained using the methods proposed in this thesis could be exploited similarly. In addition, the proposed approaches have employed a sparsity based-regularization using a learned dictionary of typical cardiac motions. Therefore, exploiting the knowledge that can be gained from this regularization strategy would be an interesting prospect. The analysis of the sparse codes could, for instance, be exploited for segmentation by using separate motion dictionaries for distinct anatomical regions. Other examples include the detection of abnormal motions or specific cardiac malfunctions encoded in associated motion dictionaries. Furthermore, the robust method of the second chapter results in three types of weights, *i.e.*, associated with the intensities, motion boundaries and sparse coding outliers. Thus, it would be interesting to take advantage of the information provided by these weights, for example, to perform a joint cardiac motion estimation and segmentation. The weights associated with the temporal regularization introduced in the last contribution could be exploited similarly.

Appendix

Appendix A

Likelihood term

In this appendix we detail the calculations leading to the final likelihood term (3.8) (see [Cohen 2002b] for similar derivations).

Starting from the observation model (3.5) we can write

$$\begin{aligned} \boldsymbol{I}_{k+1}(n+\boldsymbol{u}_k(n)) &= \boldsymbol{I}_k(n) + b[\boldsymbol{a}_{b,k+1}(n+\boldsymbol{u}_k(n)) - \boldsymbol{a}_{b,k}(n)] \\ &= \boldsymbol{I}_k(n) + b\log(\eta_k) \end{aligned} \tag{A.1}$$

where $\eta_k = \boldsymbol{a}_{k+1}(n+\boldsymbol{u}_k(n))/\boldsymbol{a}_k(n)$ is the ratio between the multiplicative Rayleigh noises.

Assuming that $\boldsymbol{a}_{k+1}[n+\boldsymbol{u}_k(n)]$ and $\boldsymbol{a}_k(n)$ are two independent Rayleigh noises with equal variances σ^2, the joint probability density function of $\boldsymbol{a}_{k+1}(n+\boldsymbol{u}_k(n))$ and $\boldsymbol{a}_k(n)$ is as follows

$$\begin{aligned} p(\boldsymbol{a}_{k+1}, \boldsymbol{a}_k) &= p(\boldsymbol{a}_{k+1})p(\boldsymbol{a}_k) \\ &= \frac{\boldsymbol{a}_{k+1}}{\sigma^2}\exp\left(\frac{-\boldsymbol{a}_{k+1}}{2\sigma^2}\right)\frac{\boldsymbol{a}_k}{\sigma^2}\exp\left(\frac{-\boldsymbol{a}_k}{2\sigma^2}\right)\mathbb{1}_{\mathbb{R}^+}(\boldsymbol{a}_k)\mathbb{1}_{\mathbb{R}^+}(\boldsymbol{a}_{k+1}) \end{aligned}$$

where $\sigma \in \mathbb{R}^+$ is a scale parameter, $\mathbb{1}_{\mathbb{R}^+}(.)$ denotes the indicator function on $\mathbb{R}^+$ and $p(a)$ is the probability density function of a Rayleigh distributed noise a and is defined as

$$p(a) = \frac{a}{\sigma^2}\exp\left(\frac{-a^2}{2\sigma^2}\right)\mathbb{1}_{\mathbb{R}^+}(a) \tag{A.2}$$

After applying the change of variable $\boldsymbol{a}_{k+1} = \eta_k\boldsymbol{a}_k$, the joint probability density function of η_k and $\boldsymbol{a}_k$ becomes

$$\pi(\eta_k, \boldsymbol{a}_k) = p(\eta_k\boldsymbol{a}_k, \boldsymbol{a}_k)\boldsymbol{a}_k$$

and

$$\begin{aligned}
\pi(\eta_k) &= \int p(\eta_k \boldsymbol{a}_k, \boldsymbol{a}_k) \boldsymbol{a}_k d\boldsymbol{a}_k \\
&= \int \boldsymbol{a}_k \frac{\eta_k \boldsymbol{a}_k}{\sigma^2} \exp\left(\frac{-\boldsymbol{a}_k^2 \eta_k^2}{2\sigma^2}\right) \frac{\boldsymbol{a}_k}{\sigma^2} \exp\left(\frac{-\boldsymbol{a}_k^2}{2\sigma^2}\right) d\boldsymbol{a}_k \\
&= \int \frac{\boldsymbol{a}_k^3}{\sigma^4} \eta_k \exp\left(\frac{-\boldsymbol{a}_k^2}{2\sigma^2}(\eta_k^2 + 1)\right) d\boldsymbol{a}_k \\
&= \frac{2\sigma^4 \eta_k}{(\eta_k^2 + 1)^2} \mathbb{1}_{\mathbb{R}^+}(\eta_k) \qquad \text{(A.3)}
\end{aligned}$$

Finally, using (A.1) and (A.3), the conditional likelihood of $\boldsymbol{I}_{k+1}(n + \boldsymbol{u}(n))$ given $\boldsymbol{I}_k(n)$ and $\boldsymbol{u}(n)$ can be formulated as follows

$$\begin{aligned}
p[\boldsymbol{I}_{k+1}(n + \boldsymbol{u}_k(n)) | \boldsymbol{I}_k(n), \boldsymbol{u}_k(n)] &= \pi(\eta_k) \frac{d\eta_k}{d\boldsymbol{I}_{k+1}} \\
&= \frac{1}{b} \pi(\eta_k) \exp\left[\frac{1}{b}(\boldsymbol{I}_{k+1}(n + \boldsymbol{u}_k(n)) - \boldsymbol{I}_k(n))\right] \\
&= \frac{2\sigma^4 \eta_k}{b(\eta_k^2 + 1)^2} \exp\left[\frac{1}{b}(\boldsymbol{I}_{k+1}(n + \boldsymbol{u}_k(n)) - \boldsymbol{I}_k(n))\right] \\
&= \frac{2\sigma^4 \eta_k^2}{b(\eta_k^2 + 1)^2} \mathbb{1}_{\mathbb{R}^+}[\boldsymbol{I}_{k+1}(n + \boldsymbol{u}_k(n))]
\end{aligned}$$

where $\eta_k = \exp[\frac{1}{b}(\boldsymbol{I}_{k+1}(n + \boldsymbol{u}_k(n)) - \boldsymbol{I}_k(n))]$. After applying the negative log-transform as in (3.2), the data fidelity term can be obtained for the displacement vector $\boldsymbol{u}_k$ as follows

$$E_{\mathrm{CD}_2}(\boldsymbol{u}) = -2d_k(\boldsymbol{u}) + 2\log[e^{2d_k(\boldsymbol{u})} + 1] + \mathrm{cst} \qquad \text{(A.4)}$$

where $d_k(\boldsymbol{u}) = \frac{1}{b} \sum_{n=1}^{N} [\boldsymbol{I}_{k+1}(n + \boldsymbol{u}_k(n)) - \boldsymbol{I}_k(n)]$ and $\mathrm{cst} = -\log(2\sigma^4/b)$ is a constant.

Bibliography

[Abraham 2007] Theodore P. Abraham, Veronica L. Dimaano and Hsin-Yueh Liang. *Role of Tissue Doppler and Strain Echocardiography in Current Clinical Practice.* Circulation, vol. 116, no. 22, pages 2597–2609, 2007.

[Achim 2001] A. Achim, A. Bezerianos and P. Tsakalides. *Novel Bayesian multiscale method for speckle removal in medical ultrasound images.* IEEE Trans. Med. Imag., vol. 20, no. 8, pages 772–783, Aug 2001.

[Afonso 2011] M. V. Afonso, J. M. Bioucas-Dias and M. A. T. Figueiredo. *An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems.* IEEE Trans. Image Process., vol. 20, no. 3, pages 681–695, March 2011.

[Aghajani 2016] Khadijeh Aghajani, Mohammad T. Manzuri and Rohollah Yousefpour. *A robust image registration method based on total variation regularization under complex illumination changes.* Computer Methods and Programs in Biomedicine, vol. 134, no. Supplement C, pages 89–107, 2016.

[Alessandrini 2013a] M Alessandrini, A Basarab, H Liebgott and O. Bernard. *Myocardial Motion Estimation from Medical Images Using the Monogenic Signal.* IEEE Trans. Image Process., vol. 22, no. 3, pages 1084–1095, 2013.

[Alessandrini 2013b] M Alessandrini, A Basarab, H Liebgott and O. Bernard. *Myocardial Motion Estimation from Medical Images Using the Monogenic Signal.* IEEE Trans. Image Process., vol. 22, no. 3, pages 1084–1095, 2013.

[Alessandrini 2013c] Martino Alessandrini, Hervé Liebgott, Daniel Barbosa and Olivier Bernard. Monogenic phase based optical flow computation for myocardial motion analysis in 3D echocardiography, volume 7746 of *Lecture Notes in Computer Science*, chapter 19, pages 159–168. 2013.

[Alessandrini 2013d] Martino Alessandrini, Hervé Liebgott, Daniel Barbosa and Olivier Bernard. *Monogenic Phase Based Optical Flow Computation for Myocardial Motion Analysis in 3D Echocardiography.* In Proc STACOM workshop of the 15th Int. Conf. Med. Image Computing and Comput. Assisted Intervention (MICCAI'2012), volume 7746, pages 159–168, Nice, France, Oct. 2013.

[Alessandrini 2014] M. Alessandrini, A. Basarab, L. Boussel, X. Guo, A. Serusclat, D. Friboulet, D. Kouamé, O. Bernard and H. Liebgott. *A New Technique for the Estimation of Cardiac Motion in Echocardiography Based on Transverse Oscillations: A Preliminary Evaluation In Silico and a Feasibility Demonstration In Vivo*. IEEE Trans. Med. Imag., vol. 33, no. 5, pages 1148–1162, May 2014.

[Alessandrini 2015] Martino Alessandrini, Mathieu De Craene, Olivier Bernard, Sophie Giffard-Roisin, Pascal Allain, Juergen Weese, Eric Saloux, Hervé Delingette, Maxime Sermesant and Jan D'Hooge. *A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database*. IEEE Trans. Med. Imag., vol. 34, no. 8, pages 1436–1451, Jul. 2015.

[Alessandrini 2016a] M Alessandrini, B Heyde, S Queiros, S Cygan, M Zontak, O Somphone, O Bernard, M De Craene, M O'Donnell and J. D'hooge. *Detailed Evaluation of Five 3D Speckle Tracking Algorithms using Synthetic Echocardiographic Recordings*. IEEE Trans. Med. Imag., vol. 35, no. 8, pages 1915–1926, Aug. 2016.

[Alessandrini 2016b] M Alessandrini, B Heyde, S Queiros, S Cygan, M Zontak, O Somphone, O Bernard, M De Craene, M O'Donnell and J. D'hooge. *Detailed Evaluation of Five 3D Speckle Tracking Algorithms using Synthetic Echocardiographic Recordings*. IEEE Trans. Med. Imag., vol. 35, no. 8, pages 1915–1926, Aug. 2016.

[Angelsen 2000] B. Angelsen. Ultrasound Imaging: Waves, Signals, and Signal Processing. Ultrasound Imaging: Waves, Signals, and Signal Processing. Emantec, 2000.

[Aviles 2017] Angelica I Aviles, Thomas Widlak, Alicia Casals, Maartje M Nillesen and Habib Ammari. *Robust cardiac motion estimation using ultrafast ultrasound data: a low-rank topology-preserving approach*. Physics in Medicine & Biology, vol. 62, no. 12, pages 4831–4851, 2017.

[Baraldi 1996] Patrizia Baraldi, A. Sarti, C. Lamberti, Alessandro Prandini and Fiorella Sgallari. *Evaluation of differential optical flow techniques on synthesized echo images*. IEEE Trans. Biomed. Eng., vol. 43, no. 3, pages 259–272, 1996.

[Behar 2004a] V. Behar, D. Adam, P. Lysyansky and Z. Friedman. *The combined effect of nonlinear filtration and window size on the accuracy of tissue displacement estimation using detected echo signals*. Ultrasonics, vol. 41, no. 9, pages 743–753, 2004.

[Behar 2004b] V. Behar, D. Adam, P. Lysyansky and Z. Friedman. *Improving motion estimation by accounting for local image distortion*. Ultrasonics, vol. 43, no. 1, pages 57–65, 2004.

[Bentoutou 2005] Y. Bentoutou, N. Taleb, K. Kpalma and J. Ronsin. *An Automatic Image Registration for Applications in Remote Sensing*. IEEE Trans. Geosci. and Remote Sens., vol. 43, no. 9, pages 2127–2137, Sept 2005.

[Bjaerum 2002] S. Bjaerum, H. Torp and K. Kristoffersen. *Clutter filter design for ultrasound color flow imaging.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 49, no. 2, pages 204–216, Feb. 2002.

[Black 1996] M. J. Black and P. Anandan. *The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields.* Computer Vision and Image Understanding, vol. 63, no. 1, pages 75–104, Jan. 1996.

[Boukerroui 2003] Djamal Boukerroui, J. Alison Noble and Michael Brady. Velocity estimation in ultrasound images: A block matching approach, pages 586–598. Ambleside, UK, July 2003.

[Brox 2004] T. Brox, A. Bruhn, N. Papenberg and J. Weickert. *High accuracy optical flow estimation based on a theory for warping.* In Proc. 8th Eur. Conf. Comput. Vision (ECCV '04), pages 25–36, Prague, Czech Republic, May 2004.

[Byram 2013] B.C. Byram and G.E Trahey. *Bayesian Speckle Tracking. Part I: An Implementable Perturbation to the Likelihood Function for Ultrasound Displacement Estimation.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 60, no. 1, pages 132—-143, 2013.

[Capon 1969] J. Capon. *High-resolution frequency-wavenumber spectrum analysis.* Proceedings of the IEEE, vol. 57, no. 8, pages 1408–1418, Aug 1969.

[Cohen 2002a] B. Cohen and I. Dinstein. *New maximum likelihood motion estimation schemes for noisy ultrasound images.* Pattern Recognition, vol. 35, no. 2, pages 455–463, 2002.

[Cohen 2002b] B. Cohen and I. Dinstein. *New maximum likelihood motion estimation schemes for noisy ultrasound images.* Pattern Recognition, vol. 35, no. 2, pages 455–463, 2002.

[Combettes 2005] P. Combettes and V. Wajs. *Signal Recovery by Proximal Forward-Backward Splitting.* Multiscale Modeling & Simulation, vol. 4, no. 4, pages 1168–1200, 2005.

[Compas 2012] C. B. Compas, E. Y. Wong, X. Huang, S. Sampath, B. A. Lin, X. Papademetris, K. Thiele, D. P. Dione, A. J. Sinusas, M. O'Donnell and J. S. Duncan. *A combined shape tracking and speckle tracking approach for 4D deformation analysis in echocardiography.* In Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI'12), pages 458–461, Barcelona, Spain, 2012.

[Compas 2014] C. B. Compas, E. Y. Wong, X. Huang, S. Sampath, B. A. Lin, P. Pal, X. Papademetris, K. Thiele, D. P. Dione, M. Stacy, L. H. Staib, A. J. Sinusas, M. O'Donnell and J. S. Duncan. *Radial Basis Functions for Combining Shape and Speckle Tracking in 4D Echocardiography.* IEEE Trans. Med. Imag., vol. 33, no. 6, pages 1275–1289, June 2014.

[Cottrell 2010] Caroline Cottrell and James N. Kirkpatrick. *Echocardiographic strain imaging and its use in the clinical setting.* Expert Review of Cardiovascular Therapy, vol. 8, no. 1, pages 93–102, 2010.

[Dawood 2006] M. Dawood, N. Lang, Xiaoyi Jiang and K. P. Schafers. *Lung motion correction on respiratory gated 3-D PET/CT images.* IEEE Trans. Med. Imag., vol. 25, no. 4, pages 476–485, April 2006.

[De Craene 2012] Mathieu De Craene, Gemma Piella, Oscar Camara, Nicolas Duchateau, Etelvino Silva, Adelina Doltra, Jan D'hooge, Josep Brugada, Marta Sitges and Alejandro F. Frangi. *Temporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography.* Medical Image Analysis, vol. 16, no. 2, pages 427–450, 2012.

[D'hooge 2002] J. D'hooge, E. Konofagou, F. Jamal, A. Heimdal, L. Barrios, B. Bijnens, J. Thoen, F. Van de Werf, G. Sutherland and P. Suetens. *Two-dimensional ultrasonic strain rate measurement of the human heart in vivo.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 49, no. 2, pages 281–286, 2002.

[D'hooge 2015] Jan D'hooge, Daniel Barbosa, Hang Gao, Piet Claus, David Prater, Jamie Hamilton, Peter Lysyansky, Yasuhiko Abe, Yasuhiro Ito, Helene Houle *et al. Two-dimensional speckle tracking echocardiography: standardization efforts based on synthetic ultrasound data.* Eur. Heart J.-Cardiovascular Imag., vol. 17, no. 6, pages 693–701, Aug. 2015.

[Dong 2011] W. Dong, X. Li, L. Zhang and G. Shi. *Sparsity-based image denoising via dictionary learning and structural clustering.* In Proc. IEEE Conf. Comput. Vision and Pattern Recognition, pages 457–464, Providence, RI, June 2011.

[Duan 2007] Q. Duan, E. Angelini, S. Homma and A. Laine. *Validation of Optical-Flow for Quantification of Myocardial Deformations on Simulated RT3D Ultrasound.* In Proc. IEEE 4th Int. Symp. Biomed. Imag. (ISBI), pages 944–947, Los Alamitos, CA, 2007.

[Elad 2006] M. Elad and M. Aharon. *Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries.* IEEE Trans. Image Process., vol. 15, no. 12, pages 3736–3745, Dec 2006.

[Elen 2008] A. Elen, Choi Hon Fai, D. Loeckx, Gao Hang, P. Claus, P. Suetens, F. Maes and J. D'Hooge. *Three-Dimensional Cardiac Strain Estimation Using Spatio-Temporal Elastic Registration of Ultrasound Images: A Feasibility Study.* IEEE Trans. Med. Imag., vol. 27, no. 11, pages 1580–1591, 2008.

[Engan 1999] K. Engan, S. O. Aase and J. Hakon Husoy. *Method of optimal directions for frame design.* In Proc. IEEE ICASSP-99, volume 5, pages 2443–2446, 1999.

[Furht 2012] B. Furht, J. Greenberg and R. Westwater. Motion Estimation Algorithms for Video Compression. The Springer Int. Series Eng. and Computer Science. Springer US, 2012.

[Goodman 2007] J.W. Goodman. Speckle phenomena in optics: Theory and applications. Roberts & Company, 2007.

[Héas 2008] P. Héas and E. Mémin. *Optical-flow for 3D atmospheric motion estimation.* In Proc. Int. Conf. Comput. Vision Theory and Appl. (VISAPP'08), volume 2, pages 399–406, Funchal, Portugal, Jan. 2008.

[Héas 2012] Patrick Héas, Cedric Herzet and Étienne Mémin. *Bayesian Inference of Models and Hyperparameters for Robust Optical-Flow Estimation.* IEEE Trans. Image Process., vol. 21, pages 1437–1451, April 2012.

[Heyde 2013] Brecht Heyde, Stefaan Bouchez, Sabine Thieren, Michael Vandenheuvel, Ruta Jasaityte, Daniel Barbosa, Piet Claus, Frederik Maes, Patrick Wouters and Jan D'hooge. *Elastic Image Registration to Quantify 3-D Regional Myocardial Deformation from Volumetric Ultrasound: Experimental Validation in an Animal Model.* Ultrasound in Medicine & Biology, vol. 39, no. 9, pages 1688–1697, 2013.

[Horn 1981] Berthold K.P. Horn and Brian G. Schunck. *Determining optical flow.* Artificial Intelligence, vol. 17, no. 1, pages 185–203, 1981.

[Huang 2014] Xiaojie Huang, Donald P. Dione, Colin B. Compas, Xenophon Papademetris, Ben A. Lin, Alda Bregasi, Albert J. Sinusas, Lawrence H. Staib and James S. Duncan. *Contour tracking in echocardiographic sequences via sparse representation and dictionary learning.* Medical Image Analysis, vol. 18, no. 2, pages 253–271, 2014.

[Huber 1981] P.J. Huber, J. Wiley and W. InterScience. Robust statistics. Wiley New York, NY, 1981.

[Ingels 1975] N B Ingels, G T Daughters, E B Stinson and E L Alderman. *Measurement of midwall myocardial dynamics in intact man by radiography of surgically implanted markers.* Circulation, vol. 52, no. 5, pages 859–867, 1975.

[Jef 2010] Vandemeulebroucke Jef, Rit Simon, Kybic Jan, Clarysse Patrick and Sarrut David. *Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs.* Medical Physics, vol. 38, no. 1, pages 166–178, 2010.

[Jia 2011a] Wang X. Jia K. *Optical flow estimation using learned sparse model.* In Proc. IEEE Int. Conf. Comput. Vision, pages 2391—-2398, Barcelona, Spain, 2011.

[Jia 2011b] Wang X. Jia K. *Optical flow estimation using learned sparse model.* In Proc. IEEE Int. Conf. Comput. Vision, pages 2391—-2398, Barcelona, Spain, 2011.

[Joos 2018] P. Joos, J. Porée, H. Liebgott, D. Vray, M. Baudet, J. Faurie, F. Tournoux, G. Cloutier, B. Nicolas and D. Garcia. *High-Frame-Rate Speckle-Tracking Echocardiography.* IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, vol. 65, no. 5, pages 720–728, May 2018.

[Kaasschieter 1988] E.F. Kaasschieter. *Preconditioned conjugate gradients for solving singular systems.* Journal of Computational and Applied Mathematics, vol. 24, no. 1, pages 265–275, 1988.

[Kaluzynski 2001] K. Kaluzynski, Xunchang Chen, S. Y. Emelianov, A. R. Skovoroda and M. O'Donnell. *Strain rate imaging using two-dimensional speckle tracking.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 48, no. 4, pages 1111–1123, July 2001.

[Khamis 2016] H. Khamis, S. Shimoni, A. Hagendorff, N. Smirin, Z. Friedman and D. Adam. *Optimization-Based Speckle Tracking Algorithm for Left Ventricle Strain Estimation: A Feasibility Study.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 63, no. 8, pages 1093–1106, Aug 2016.

[Kim 2006] Yeon-Ho Kim and A. C. Kak. *Error analysis of robust optical flow estimation by least median of squares methods for the varying illumination model.* IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pages 1418–1435, Sept 2006.

[Kindberg 2010] Katarina Kindberg. *Invasive and Non-Invasive Quantification of Cardiac Kinematics.* PhD thesis, Linköping University, 2010.

[Kontogeorgakis 1994] C. Kontogeorgakis, M.G. Strintzis, N. Maglaveras and I. Kokkinidis. *Tumor detection in ultrasound B-mode images through motion estimation using a texture detection algorithm.* In Computers in Cardiology, pages 117–120. IEEE Comput. Soc. Press, 1994.

[Kotropoulos 1994] C. Kotropoulos, X. Magnisalis, I. Pitas and M. G. Strintzis. *Nonlinear ultrasonic image processing based on signal-adaptive filters and self-organizing neural networks.* IEEE Trans. Image Process., vol. 3, no. 1, pages 65–77, Jan 1994.

[Krim 1996] H. Krim and M. Viberg. *Two decades of array signal processing research: the parametric approach.* IEEE Signal Process. Mag., vol. 13, no. 4, pages 67–94, Jul 1996.

[Ledesma-Carbayo 2005] M.J. Ledesma-Carbayo, J. Kybic, M. Desco, A. Santos, M. Suhling, P. Hunziker and M. Unser. *Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation.* IEEE Trans. Med. Imag., vol. 24, no. 9, pages 1113–1126, 2005.

[Ledesma-Carbayo 2008] Maria J. Ledesma-Carbayo, J. Andrew Derbyshire, Smita Sampath, Andrés Santos, Manuel Desco and Elliot R. McVeigh. *Unsupervised estimation of myocardial displacement from tagged MR sequences using nonrigid*

registration. Magnetic Resonance in Medicine, vol. 59, no. 1, pages 181–189, 2008.

[Lin 2011] Zhouchen Lin, Risheng Liu and Zhixun Su. *Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation.* In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 24, pages 612–620. Curran Associates, Inc., 2011.

[Liu 2009] Xiaofeng Liu, Khaled Z. Abd-Elmoniem and Jerry L. Prince. *Incompressible Cardiac Motion Estimation of the Left Ventricle Using Tagged MR Images.* In Proc. 12th Int. Conf. Med. Image Computing and Comput. Assisted Intervention: Part II (MICCAI'09), pages 331–338, Berlin, Heidelberg, 2009. Springer-Verlag.

[Loizou 2005] C. P. Loizou, C. S. Pattichis, C. I. Christodoulou, R. S. H. Istepanian, M. Pantziaris and A. Nicolaides. *Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery.* vol. 52, no. 10, pages 1653–1669, Oct 2005.

[Lopata 2009] Richard G. P. Lopata, Maartje M. Nillesen, Hendrik H. G. Hansen, Inge H. Gerrits, Johan M. Thijssen and Chris L. de Korte. *Performance Evaluation of Methods for Two-Dimensional Displacement and Strain Estimation Using Ultrasound Radio Frequency Data.* Ultrasound in Medicine & Biology, vol. 35, no. 5, pages 796–812, 2009.

[Lopata 2011] Richard G.P. Lopata, Maartje M. Nillesen, Johan M. Thijssen, L. Kapusta and Chris L. de Korte. *Three-Dimensional Cardiac Strain Imaging in Healthy Children Using RF-Data.* Ultrasound in Medicine & Biology, vol. 37, no. 9, pages 1399 – 1408, 2011.

[Lorintiu 2015] O. Lorintiu. *Compressed sensing reconstruction for 3D and Doppler medical ultrasound.* PhD thesis, INSA-Lyon, 2015.

[Lu 2012] C. Lu, S. Chelikani, D. A. Jaffray, M. F. Milosevic, L. H. Staib and J. S. Duncan. *Simultaneous Nonrigid Registration, Segmentation, and Tumor Detection in MRI Guided Cervical Cancer Radiation Therapy.* IEEE Trans. Med. Imag., vol. 31, no. 6, pages 1213–1227, June 2012.

[Lu 2013] C. Lu, J. Shi and J. Jia. *Online Robust Dictionary Learning.* In IEEE Conference on Computer Vision and Pattern Recognition, pages 415–422, June 2013.

[Luo 2008] J. Luo and E.E Konofagou. *High-frame rate, full-view myocardial elastography with automated contour tracking in murine left ventricles in vivo.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 55, no. 1, pages 240–248, 2008.

[Mailloux 1989] Guy E. Mailloux, F. Langlois, P.Y. Simard and M. Bertrand. *Restoration of the velocity field of the heart from two-dimensional echocardiograms.* IEEE Trans. Med. Imag., vol. 8, no. 2, pages 143–153, 1989.

[Mairal 2008a] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman. *Discriminative learned dictionaries for local image analysis.* In Proc. IEEE Conf. Comp. Vision and Pattern Recognition, pages 1–8, Anchorage, AK, June 2008.

[Mairal 2008b] Julien Mairal, Guillermo Sapiro and Michael Elad. *Learning Multiscale Sparse Representations for Image and Video Restoration.* Multiscale Modeling & Simulation, vol. 7, no. 1, pages 214–241, 2008.

[Mairal 2009] Julien Mairal, Francis Bach, Jean Ponce and Guillermo Sapiro. *Online Dictionary Learning for Sparse Coding.* In Proc. 26th Annu. Int. Conf. Mach. Learning (ICML '09), pages 689–696, Montreal, Quebec, Canada, 2009.

[Mallat 1993] S. G. Mallat and Z. Zhang. *Matching Pursuits with Time-Frequency Dictionaries.* IEEE Trans. Signal Process., pages 3397—-3415, 1993.

[McLeod 2012] Kristin McLeod, Adityo Prakosa, Tommaso Mansi, Maxime Sermesant and Xavier Pennec. *An Incompressible Log-Domain Demons Algorithm for Tracking Heart Tissue.* In Oscar Camara, Ender Konukoglu, Mihaela Pop, Kawal Rhode, Maxime Sermesant and Alistair Young, editors, Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges, pages 55–67, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[McLeod 2015] K. McLeod, M. Sermesant, P. Beerbaum and X. Pennec. *Spatio-Temporal Tensor Decomposition of a Polyaffine Motion Model for a Better Analysis of Pathological Left Ventricular Dynamics.* IEEE Trans. Med. Imag., vol. 34, no. 7, pages 1562–1575, July 2015.

[Meer 1991] Peter Meer, Doron Mintz, Azriel Rosenfeld and Dong Yoon Kim. *Robust regression methods for computer vision: A review.* Int. Journal Computer Vision, vol. 6, pages 59–70, 1991.

[Michailovich 2006] A. Michailovich O. V.and Tannenbaum. *Despeckling of medical ultrasound images.* vol. 53, no. 1, pages 64–78, Jan 2006.

[Michal Aharon 2006] Michael Elad Michal Aharon and Alfred Bruckstein. *K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation.* IEEE Trans. Signal Process., vol. 54, no. 11, pages 4311–4322, 2006.

[Mignotte 2001] M. Mignotte and J. Meunier. *A multiscale optimization approach for the dynamic contour-based boundary detection issue.* Computerized Medical Imaging and Graphics, vol. 25, no. 3, pages 265–275, 2001.

[Mirea 2016] O Mirea, J Duchenne and JU Voigt. *Recent advances in echocardiography: strain and strain rate imaging [version 1; referees: 3 approved].* F1000Research, vol. 5, no. 787, 2016.

[Moller 1993] Martin Fodslette Moller. *A scaled conjugate gradient algorithm for fast supervised learning.* Neural Networks, vol. 6, no. 4, pages 525–533, 1993.

[Morais 2013] Pedro Morais, Brecht Heyde, Daniel Barbosa, Sandro Queirós, Piet Claus and Jan D'hooge. *Cardiac Motion and Deformation Estimation from Tagged MRI Sequences Using a Temporal Coherent Image Registration Framework*. In Sébastien Ourselin, Daniel Rueckert and Nicolas Smith, editors, Functional Imaging and Modeling of the Heart, pages 316–324, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[Myronenko 2009a] Andriy Myronenko, Xubo Song and DavidJ Sahn. *Maximum Likelihood Motion Estimation in 3D Echocardiography through Non-rigid Registration in Spherical Coordinates*. In Proc. 5th Int. Conf. Functional Imag. and Modeling of the Heart (FIMH'09), volume 5528, pages 427–436, Nice, France, June 2009.

[Myronenko 2009b] Andriy Myronenko, Xubo Song and DavidJ Sahn. Maximum likelihood motion estimation in 3D echocardiography through non-rigid registration in spherical coordinates, volume 5528 of *Lecture Notes in Computer Science*, chapter 46, pages 427–436. 2009.

[Myronenko 2010] Andriy Myronenko. *Non-rigid image registration regularization, algorithms and applications*. PhD thesis, Oregon Health & Science University, 2010.

[Nagata 2015] Yasufumi Nagata, Masaaki Takeuchi, Kei Mizukoshi, Victor Chien-Chia Wu, Fen-Chiung Lin, Kazuaki Negishi, Satoshi Nakatani and Yutaka Otsuji. *Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software*. J. Amer. Soc. Echocardiogr., vol. 28, no. 6, pages 630–641, 2015.

[Odobez 1995] J.M. Odobez and P. Bouthemy. *Robust Multiresolution Estimation of Parametric Motion Models*. Journal of Visual Communication and Image Representation, vol. 6, no. 4, pages 348–365, 1995.

[Oksuz 2015] Ilkay Oksuz, Anirban Mukhopadhyay, Marco Bevilacqua, Rohan Dharmakumar and Sotirios A. Tsaftaris. *Dictionary Learning Based Image Descriptor for Myocardial Registration of CP-BOLD MR*. In Proc. 18th Int. Conf. Med. Image Computing and Comput. Assisted Intervention (MICCAI'15), pages 205–213, Munich, Germany, Oct. 2015.

[Ong 1999] E.P. Ong and M. Spann. *Robust Optical Flow Computation Based on Least-Median-of-Squares Regression*. Int. Journal of Computer Vision, vol. 31, no. 1, pages 51–82, Feb 1999.

[Onofrey 2015] John A. Onofrey, Lawrence H. Staib and Xenophon Papademetris. Segmenting the brain surface from CT images with artifacts using dictionary learning for non-rigid MR-CT registration, pages 662–674. Sabhal Mor Ostaig, Isle of Skye, UK, June 2015.

[Organization 2017] World Health Organization. World health statistics. 2017.

[Otte 1994] M. Otte and H.-H. Nagel. *Optical Flow Estimation: Advances and Comparisons.* In Proc. 3rd Eur. Conf. Comput. Vision (ECCV '94), volume 1, pages 51–60, Stockholm, Sweden, 1994.

[Parajuli 2015] Nripesh Parajuli, Colin B. Compas, Ben A. Lin, Smita Sampath, Matthew O'Donnell, Albert J. Sinusas and James S. Duncan. Sparsity and biomechanics inspired integration of shape and speckle tracking for cardiac deformation analysis, pages 57–64. Maastricht, The Netherlands, June 2015.

[Parajuli 2016] Nripesh Parajuli, Allen Lu, John C. Stendahl, Maria Zontak, Nabil Boutagy, Melissa Eberle, Imran Alkhalil, Matthew O'Donnell, Albert J. Sinusas and James S. Duncan. Integrated dynamic shape tracking and rf speckle tracking for cardiac motion analysis, pages 431–438. Athens, Greece, Oct. 2016.

[Pati 1993] Y. Pati, R. Rezaiifar and P. Krishnaprasad. *Orthogonal Matching Pursuit: recursive function approximation with application to wavelet decomposition.* In Proc. 27th Asilomar Conf. Signals, Syst. and Comput., pages 2391—2398, Pacific Grove, CA, 1993.

[Pereyra 2016] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero and S. McLaughlin. *Tutorial on Stochastic Simulation and Optimization Methods in Signal Processing.* IEEE J. Sel. Topics Signal Process., vol. 10, no. 2, pages 224–241, March 2016.

[Perperidis 2016] Antonios Perperidis. *Post-processing approaches for the improvement of cardiac ultrasound B-mode images: a review.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 63, no. 3, pages 470–485, March 2016.

[Porée 2018] J. Porée, M. Baudet, F. Tournoux, G. Cloutier and D. Garcia. *A Dual Tissue-Doppler Optical-Flow Method for Speckle Tracking Echocardiography at High Frame Rate.* IEEE Trans. Med. Imag., vol. 37, no. 9, pages 2022–2032, Sept 2018.

[Porras 2016] A. R. Porras, M. Alessandrini, O. Mirea, J. D'hooge, A. F. Frangi and G. Piella. *Integration of Multi-Plane Tissue Doppler and B-Mode Echocardiographic Images for Left Ventricular Motion Estimation.* IEEE Trans. Med. Imag., vol. 35, no. 1, pages 89–97, Jan 2016.

[Pratikakis 2001] Ioannis Pratikakis, Christian Barillot and Pierre Hellier. Robust multiscale non-rigid registration of 3d ultrasound images, pages 389–397. Springer, Berlin, Heidelberg, Vancouver, Canada, July 2001.

[Prince 1992] J. L. Prince and E. R. McVeigh. *Motion estimation from tagged MR image sequences.* IEEE Trans. Med. Imag., vol. 11, no. 2, pages 238–249, June 1992.

[Ramani 2008] S. Ramani, T. Blu and M. Unser. *Monte-Carlo SURE: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms.* IEEE Trans. Image Process., vol. 17, no. 9, pages 1540–1554, Sept. 2008.

[Rivaz 2011] H. Rivaz, E. M. Boctor, M. A. Choti and G. D. Hager. *Real-Time Regularized Ultrasound Elastography.* IEEE Trans. Med. Imag., vol. 30, no. 4, pages 928–945, April 2011.

[Rueckert 1999a] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach and D. J. Hawkes. *Nonrigid registration using free-form deformations: application to breast MR images.* IEEE Trans. Med. Imag., vol. 18, no. 8, pages 712–721, Aug 1999.

[Rueckert 1999b] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach and D. J. Hawkes. *Nonrigid registration using free-form deformations: application to breast MR images.* IEEE Trans. Med. Imag., vol. 18, no. 8, pages 712–721, Aug 1999.

[Salles 2015] S. Salles, A. J. Y. Chee, D. Garcia, A. C. H. Yu, D. Vray and H. Liebgott. *2-D arterial wall motion imaging using ultrafast ultrasound and transverse oscillations.* vol. 62, no. 6, pages 1047–1058, June 2015.

[Sariyanidi 2017] E. Sariyanidi, H. Gunes and A. Cavallaro. *Robust Registration of Dynamic Facial Sequences.* IEEE Trans. Image Process., vol. 26, no. 4, pages 1708–1722, April 2017.

[Sederberg 1986] Thomas W. Sederberg and Scott R. Parry. *Free-form Deformation of Solid Geometric Models.* In Proc. 13th Annu. Conf. Comput. Graph. and Interactive Techn., pages 151–160, New York, NY, USA, 1986.

[Shah 2012] Amil M. Shah and Scott D. Solomon. *Myocardial Deformation Imaging: Current Status and Future Directions.* Circulation, vol. 125, no. 2, pages e244–e248, 2012.

[Shen 2007] H. Shen, L. Zhang, B. Huang and P. Li. *A MAP Approach for Joint Motion Estimation, Segmentation, and Super Resolution.* IEEE Trans. Image Process., vol. 16, no. 2, pages 479–490, Feb 2007.

[Shen 2010] X Shen and Y Wu. *Sparsity model for robust optical flow estimation at motion discontinuities.* In Proc. IEEE Conf. Comput. Vision and Pattern Recognition, pages 2456–2463, San Francisco, CA, 2010.

[Shi 2013] Wenzhe Shi, Martin Jantsch, Paul Aljabar, Luis Pizarro, Wenjia Bai, Haiyan Wang, Declan O'Regan, Xiahai Zhuang and Daniel Rueckert. *Temporal sparse free-form deformations.* Medical Image Analysis, vol. 17, no. 7, pages 779–789, 2013. Special Issue on the 2012 Conference on Medical Image Computing and Computer Assisted Intervention.

[Simões 2015] M. Simões, J. Bioucas-Dias, L. B. Almeida and J. Chanussot. *A Convex Formulation for Hyperspectral Image Superresolution via Subspace-Based Regularization.* IEEE Trans. Geosci. and Remote Sens., vol. 53, no. 6, pages 3373–3388, June 2015.

[Somphone 2013] Oudom Somphone, Mathieu De Craene, Roberto Ardon, Benoit Mory, Pascal Allain, Hang Gao, Jan D'Hooge, Stéphanie Marchesseau, Maxime Sermesant and Eric Saloux. *Fast Myocardial Motion and Strain Estimation in 3D Cardiac Ultrasound with Sparse Demons.* In Proc. 10th IEEE Int. Symp. Biomed. Imag. (ISBI'13), pages 1182–1185, San Francisco, United States, April 2013.

[Spottiswoode 2007] B. S. Spottiswoode, X. Zhong, A. T. Hess, C. M. Kramer, E. M. Meintjes, B. M. Mayosi and F. H. Epstein. *Tracking Myocardial Motion From Cine DENSE Images Using Spatiotemporal Phase Unwrapping and Temporal Fitting.* IEEE Trans. Med. Imag., vol. 26, no. 1, pages 15–30, Jan 2007.

[Strintzis 1997] M.G. Strintzis and I. Kokkinidis. *Maximum likelihood motion estimation in ultrasound image sequences.* IEEE Signal Process. Lett., vol. 4, no. 6, pages 156–157, 1997.

[Suhling 2005] M. Suhling, M. Arigovindan, C. Jansen, P. Hunziker and M. Unser. *Myocardial motion analysis from B-mode echocardiograms.* IEEE Trans. Image Process., vol. 14, no. 4, pages 525–536, April 2005.

[Sulam 2015] J. Sulam and M. Elad. *Expected patch log likelihood with a sparse prior.* Energy Minimization Methods in Computer Vision and Pattern Recognition, pages 99—-111, 2015.

[Sutherland 2004] George R. Sutherland, Giovanni Di Salvo, Piet Claus, Jan D'Hooge and Bart Bijnens. *Strain and strain rate imaging: a new clinical approach to quantifying regional myocardial function.* Journal of the American Society of Echocardiography, vol. 17, no. 7, pages 788–802, 2004.

[SZASZ 2016] Teodora SZASZ. *Advanced beamforming techniques in ultrasound imaging and the associated inverse problems.* PhD thesis, University of Toulouse, 2016.

[Tanter 2014] M. Tanter and M. Fink. *Ultrafast imaging in biomedical ultrasound.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 61, no. 1, pages 102–119, January 2014.

[Tavakoli 2014] V. Tavakoli, N. Bhatia, R. A. Longaker, M. F. Stoddard and A. A. Amini. *Tissue Doppler Imaging Optical Flow (TDIOF): A Combined B-Mode and Tissue Doppler Approach for Cardiac Motion Estimation in Echocardiographic Images.* IEEE Trans. Biomed. Eng., vol. 61, no. 8, pages 2264–2277, Aug 2014.

[Tibshirani 1996] Robert Tibshirani. *Regression Shrinkage and Selection via the lasso.* Journal of the Royal Statistical Society, vol. 58, no. 1, pages 267—88, 1996.

[Tosic 2011] Ivana Tosic and Pascal Frossard. *Dictionary Learning.* IEEE Signal Process. Mag., vol. 28, no. 2, pages 27–38, 2011.

[Turek 2015] Javier S. Turek, Michael Elad and Irad Yavneh. *Clutter Mitigation in Echocardiography Using Sparse Signal Separation.* Int. Journal of Biomed. Imag., Article ID 958963, Jan 2015.

[Vemuri 2003] B.C. Vemuri, J. Ye, Y. Chen and C.M. Leonard. *Image registration via level-set motion: Applications to atlas-based segmentation.* Medical Image Analysis, vol. 7, no. 1, pages 1–20, 2003.

[Villarreal 1988] F J Villarreal, L K Waldman and W Y Lew. *Technique for measuring regional two-dimensional finite strains in canine left ventricle.* Circulation Research, vol. 62, no. 4, pages 711–721, 1988.

[Viola 2003] F. Viola and W. F. Walker. *A comparison of the performance of time-delay estimators in medical ultrasound.* IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 50, no. 4, pages 392–401, April 2003.

[Volz 2011] S. Volz, A. Bruhn, L. Valgaerts and H. Zimmer. *Modeling temporal coherence for optical flow.* In Int. Conf. Computer Vision, pages 1116–1123, Barcelona, Spain, Nov. 2011.

[Wei 2015] Q. Wei, J. Bioucas-Dias, N. Dobigeon and J. Y. Tourneret. *Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation.* IEEE Trans. Geosci. and Remote Sens., vol. 53, no. 7, pages 3658–3668, July 2015.

[Wong 2013] E. Y. Wong, M. O'Donnell, K. Thiele, C. B. Compas, X. Huang, S. Sampath, B. A. Lin, P. Pal, X. Papademetris, D. Dione, L. Staib, A. J. Sinusas and J. S. Duncan. *4-D echocardiography assessment of local myocardial strain using 3-D speckle tracking combined with shape tracking.* In Proc. IEEE Int. Ultrasonics Symp. (IUS'13), pages 100–103, Prague, Czech Republic, July 2013.

[Woo 2009] Jonghye Woo, Byung-Woo Hong, Chang-Hong Hu, K. Kirk Shung, C. C. Jay Kuo and Piotr J. Slomka. *Non-Rigid Ultrasound Image Registration Based on Intensity and Local Phase Information.* Journal of Signal Processing Systems, vol. 54, no. 1-3, pages 33–43, 2009.

[Yeung 1998] F. Yeung, S. F. Levinson and K. J. Parker. *Multilevel and Motion Model-Based Ultrasonic Speckle Tracking Algorithms.* Ultrasound in Medicine & Biology, vol. 24, no. 3, pages 427–442, 1998.

[Yu 2006] W. Yu, P. Yan, A.J. Sinusas, K. Thiele and J.S. Duncan. *Towards pointwise motion tracking in echocardiographic image sequences - Comparing the reliability of different features for speckle tracking.* Medical Image Analysis, vol. 10, no. 4, pages 495–508, 2006.

[Zhang 2010] Q. Zhang and B. Li. *Discriminative K-SVD for Dictionary Learning in Face Recognition.* In Proc. IEEE Conf. Comput. Vision and Pattern Recognition, pages 2691–2698, San Francisco, CA, June 2010.

[Zhang 2011] Zhijun Zhang, Xubo Song and David J. Sahn. *Cardiac Motion Estimation from 3D Echocardiography with Spatiotemporal Regularization.* In Dimitris N. Metaxas and Leon Axel, editors, Functional Imaging and Modeling of the Heart, pages 350–358, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[Zhijun 2014] Zhang Zhijun, Ashraf Muhammad, Sahn David J. and Song Xubo. *Temporally diffeomorphic cardiac motion estimation from three-dimensional echocardiography by minimization of intensity consistency error.* Medical Physics, vol. 41, no. 5, page 052902, 2014.