Motion Estimation in Echocardiography Using Sparse Representation and Dictionary Learning

Nora Ouzir, *Student Member, IEEE*, Adrian Basarab, *Member, IEEE*, Hervé Liebgott, *Member, IEEE*, Brahim Harbaoui, and Jean-Yves Tourneret, *Senior Member, IEEE*

Abstract—This paper introduces a new method for cardiac motion estimation in 2-D ultrasound images. 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. The proposed method is evaluated on one data set with available ground-truth, including four sequences of highly realistic simulations. The approach is also validated on both healthy and pathological sequences of *in vivo* data. We evaluate the method in terms of motion estimation accuracy and strain errors and compare the performance with state-of-the-art algorithms. The results show that the proposed method gives competitive results for the considered data. Furthermore, the *in vivo* strain analysis demonstrates that meaningful clinical interpretation can be obtained from the estimated motion vectors.

Index Terms—Cardiac ultrasound, dictionary learning, motion estimation, sparse representations.

I. Introduction

CARDIOVASCULAR diseases have become the top cause of death in industrialized countries. They are responsible for up to 48% of the total number of deaths, and are expected to reach about 25 millions a year by 2030 [1]. It is therefore of critical importance to improve techniques of cardiac function assessment, thus facilitating diagnosis and treatment of these diseases. There are a variety of methods used to evaluate the health of the heart. Among the noninvasive techniques, medical imaging is used to assess its mechanical action by means of various modalities such as magnetic resonance imaging (MRI) and ultrasound imaging (UI). However, because of its relatively high temporal resolution, UI is more adapted to the rapid motion of the heart. In addition, it presents advantages such as low budget requirements and reduced discomfort for the patient. This makes UI, particularly echocardiography, the most widely used modality in cardiology. Furthermore, the acquired ultrasound (US) images provide information that is essential for cardiac function evaluation. US images can be exploited either through direct visualization or using post-processing methods that extract valuable qualitative and measurable features. In this context, 2D automatic cardiac motion estimation as well as the associated strain measurements have been proved to be efficient tools for the diagnosis of cardiovascular diseases [2]–[6].

Due to the well established clinical feasibility and advantages of 2D echocardiography, and despite the arrival of new technologies such as 3D imagery [7]–[9] (still regarded as an experimental method [10], [11]), the development of new 2D cardiac motion estimation methods is still an active area of research [12]–[15]. Also, note that 2D echocardiography is still more routinely used than 3D in clinical practice. 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 strain measurements. Furthermore, the assessment of the cardiac function is still limited to global measurements [11] and undergoes great amounts of smoothing, causing loss of clinically valuable local information [10]. 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 [10].

Most of the techniques used for cardiac motion estimation fall into three main categories: optical flow, speckle tracking and elastic image registration. Optical flow algorithms are based on differential methods and use the over time pixel intensity constancy assumption [16]–[18]. They provide a dense estimation of the motion field by means of the entire image [18] or by a local window analysis of spatial and

N. Ouzir and J.-Y. Tourneret are with the University of Toulouse, IRIT/INP-ENSEEIHT/TéSA, 31071 Toulouse, France (e-mail: nora.ouzir@enseeiht.fr; jean-yves.tourneret@enseeiht.fr).

H. Liebgott and B. Harbaoui are with the University of Lyon, INSALyon, Claude Bernard University Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, 69100 LYON, France (e-mail: liebgott@creatis.insass-lyon.fr; brahim.harbaoui@creatis.insa-lyon.fr).

B. Harbaoui is with the Radiology Department, Hôpital de la Croix-Rousse, Hospices Civils de Lyon, F-69004 Lyon, France.

A. Basarab is with the University of Toulouse, IRIT, CNRS UMR 5505, F-31062 Toulouse, France (e-mail: adrian.basarab@irit.fr).

temporal gradients [19]. However, because of the intensity preservation assumption and their local nature, these methods are sensitive to noise and fail at estimating large displacements. Consequently, the use of a multi-resolution scheme or a block-matching initialization is necessary for the methods relying on optical flow [19]–[21]. The second category of cardiac motion estimation methods consists in matching blocks of two consecutive images using a similarity measure. These so-called speckle tracking algorithms use the image noise as a spatial marker for motion and rely on its statistical distribution to build a similarity measure [22]. The most common similarity measures include cross-correlation (CC) [23], [24], sum of absolute differences (SAD) [25] and sum of squared differences (SSD) [26]. Finally, in elastic registration, the images are related through a non-rigid geometric transformation, which can be parametric, *e.g.*, represented on a B-spline basis [27], [28], or discrete [29]. In contrast with classical optical flow algorithms, elastic image registration directly leads to a similarity measure adapted to the image characteristics.

Motion estimation is a difficult task because of its ill-posed nature, *i.e.*, it generally does not have a unique solution. In order to tackle this issue, additional constraints are often added to regularize this problem. These constraints represent *a priori* knowledge about the behaviour of motion and force the estimation to narrow down to the type of displacements expressed by this information. In multiple motion estimation schemes, the ill-posedness is overcome by introducing a prior in the form of spatial or temporal smoothness [30], [31]. Regularization can also be achieved by using a parametric motion model. Depending on the problem at hand, these models limit the motion to be, *e.g.*, rigid, or affine [32]. In cardiac motion estimation, a common regularization is the B-spline parametrization [27], [28]. In this approach, the transformation is limited by the presence of control points and smoothness is added due to the B-spline basis function. Regularizations based on radial basis functions have also been used in [33] and [34]. In many other works [19], [21], motion patterns of the cardiac muscle are described by an affine transformation accounting for translation, rotation, expansion, compression, and shear.

Other types of regularizations, which have gained an increasing interest over the last years, are based on sparse representations. In recent works [35]–[39], sparsity has been used successfully to regularize a wide variety of problems by exploiting the sparse nature of signals or images. In particular, one way of exploiting sparsity consists in expressing the unknown signal/image as a linear combination of a few elements 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), and Fourier decompositions. However, recent works have shown that dictionaries learned from the data itself can outperform the predefined ones [40]. The methods used to learn such data-driven dictionaries are called dictionary learning (DL) methods. Aside from various signal processing applications, a few recent attempts to use sparse representations and DL for motion estimation have been investigated in the literature.

Shen and Wu [41] added a sparsity prior to an optical flow estimation problem and used the wavelet basis for the sparse coding step. This approach was also considered in [42], where the wavelet basis was replaced by a learned motion dictionary and a multi-resolution scheme was adopted. In cardiac UI, appearance dictionaries can also be learned from two distinct cardiac regions for contour tracking [43]. A similar strategy was considered in [44] for brain segmentation in CT images. Finally, a dictionary of features was learned in [45] and an SSD measure was used for the registration of cardiac MRI.

In this work, we present a new method for cardiac motion estimation in 2D ultrasound 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. The data fidelity term considered in this work is based on a multiplicative Rayleigh noise model [46].[1] The spatial smoothness is ensured by a regularization based on the gradient of the motion vector. Moreover, we promote the use of a regularization exploiting a sparse representation of motion based on DL using patterns of cardiac motion. In the sparse coding step associated with motion estimation, the dictionaries are learned using the ground-truth displacements of realistic simulations.

The paper is organized as follows. Section 2 briefly summarizes the theory related to sparse representations and DL. Section 3 formulates the cardiac motion estimation problem and introduces the proposed estimation strategy based on a sparse regularization. Some additional implementation details are provided in Section 4. Simulation results are presented and discussed in Section 5. We compare the performance of the proposed method to results obtained with three state-of-the-art algorithms: (i) the conventional block-matching method [47] using the normalized cross-correlation (NCC), (ii) an approach based on a monogenic phase similarity measure and an affine motion model [21] and (iii) an elastic registration method with a similarity measure based on a Rayleigh noise assumption and a B-spline parameterization [28]. Strain curves resulting from highly realistic simulations are also computed and strains of healthy and pathological subjects are compared using *in vivo* images. Concluding remarks are finally reported in Section 6.

II. Sparse Representations - Dictionary Learning

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 [35]–[39]. Motivated by this success, it is argued in [41] 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 the case of cardiac motion, sparsity still holds since 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

---

[1]Note that other similarity measures based on optical flow or on different noise models could be considered as well.

[21], [32]). Motion estimation can be further improved by using learned sparse models as explained in [42]. These results 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 [41], which is based on a predefined wavelet basis.

A. Sparse Representations

The goal of a sparse representation is to express a signal as a linear combination of a few elements chosen from a collection of training signals. The underlying assumption is the redundancy and self-similarity properties of the signal of interest. Due to the usual dimension of motion fields, this paper proposes to use a sparse representation for motion patches, as done for natural images [35], [48], [49]. More precisely, given an appropriate *dictionary* $D \in \mathbb{R}^{n \times q}$, *i.e.*, a set of training elements called *atoms*, which encode patterns of real motion, a displacement patch $u_p \in \mathbb{R}^n$ is represented by a weighted linear combination of few training atoms $D\alpha_p$, where $\alpha_p \in \mathbb{R}^q$ is a sparse coefficient vector with few non-zero entries. After considering the noise affecting real data, the motion vector for a given patch can be written

$$u_p = D\alpha_p + e_p \qquad (1)$$

where $e_p \in \mathbb{R}^n$ is the additive noise. A classical way of exploiting the sparsity of the unknown vector α_p is to look for a solution of (1) with the minimum number of non-zero coefficients[2] [50], *i.e.*, the solution of the problem

$$\min_{\alpha_p} \|\alpha_p\|_0 \quad \text{subject to } u_p = D\alpha_p + e_p \text{ and } \|e_p\|_2^2 < \epsilon \quad (2)$$

where $\|.\|_0$ is the l_0 *pseudo-norm*, which counts the number of non-zero elements of a vector, and the constraint $\|e_p\|_2^2 < \epsilon$ is motivated by the fact that the noise has bounded energy.

The solution of (2) 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 (such as the matching pursuit (MP) [51] and orthogonal matching pursuit (OMP) [52] algorithms) and convex relaxation methods, which relax the l_0-minimization problem to an l_1-minimization (such as the least absolute shrinkage and selection operator (LASSO) [53]). The quality of the approximations provided by these algorithms does not only depend on the signal itself but also on the chosen set of atoms, *i.e.*, the dictionary [50]. 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, *i.e.*, to cardiac motion in this paper.

B. Dictionary Learning

Learning a motion dictionary $D \in \mathbb{R}^{n \times q}$ consists of a joint optimization problem with respect to the dictionary D and the sparse coefficient vectors α_p. In practice, the dictionary is learned patch-wise, and thus, constructed as a sum over all patches of a global motion field $u \in \mathbb{R}^N$.

[2]When $n < q$, the dictionary is *overcomplete*. In this case, the solution of the motion estimation problem is not unique, requiring the definition of additional constraints.

Assuming that the noise $e_p \in \mathbb{R}^n$ is Gaussian zero-mean (see [38], [48], [50] for motivations), the DL problem can be fomulated as follows

$$\min_{D, \alpha_p} \sum_p \|P_p u - D\alpha_p\|_2^2 \quad \text{subject to } \forall p, \|\alpha_p\|_0 \leq K \quad (3)$$

where $P_p \in \mathbb{R}^{n \times N}$ is a binary operator that extracts the pth patch of size n from u and α_p is the corresponding sparse code, with K its maximum number of non-zero coefficients. The optimization problem (3) is usually solved by iterating between two steps. The first one is a sparse coding step, where the dictionary D is fixed and the optimization is performed with respect to the coefficients α_p. The second step is the dictionary update step (or learning step), where the sparse coefficient vectors α_p are fixed and the minimization is conducted with respect to the dictionary D.

Numerous algorithms designed for solving the DL problem (3) have been investigated in the literature. These algorithms include the K-SVD [48] and the online DL (ODL) [40]. 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 [35]. 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.

Learning Strategies: The dictionary can be either fixed in advance, *i.e.*, learned offline from a set of ground-truth motion 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.

1) Offline dictionary learning: Given a set of ground-truth data, the motion dictionary is learned and fixed before the estimation process. This strategy was used for image denoising in [35], face recognition in [37] and texture segmentation in [38]. The offline learning supposes that the training set is adapted to the problem and spans the different types of motions 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.

2) Adaptive dictionary learning: This strategy, also known as online learning, is commonly used in the area of natural image denoising [35], [36], [39]. Online learning consists in extracting training patches from the noisy image itself. Note that the initial dictionary can either be an offline learned dictionary, a predefined dictionary, or simply a set of random patches extracted from an initial motion estimate u_0, *i.e.*, the result of a first rough motion estimation. Since the dictionary is jointly estimated with the motion field, the adaptive learning strategy is more time consuming. However, it remains the most appropriate approach in the absence of a suitable or sufficient training set.

This paper formulates the motion estimation problem in a general form based on an offline dictionary. However, it would

be also possible to update the dictionary jointly with the motion field in situations where the set of training data is not sufficiently rich.

III. MOTION REGULARIZATION BASED ON A SPARSE REPRESENTATION

A. Motion Estimation Using the ML Method

We consider the estimation of a 2D displacement field $U_k = (u_{x,k}, u_{y,k})^T \in \mathbb{R}^N \times \mathbb{R}^N$, along the x and y axes, between a pair of consecutive frames $(r_k, r_{k+1}) \in \mathbb{R}^N \times \mathbb{R}^N$. In this paper, the problem is written independently for the horizontal and vertical components of U_k and the subscript k is omitted for simplicity, so that the considered displacement vector is equal to $u = u_{x,k}$ or $u = u_{y,k}$. The maximum likelihood (ML) approach is a common framework for the motion estimation problem [28], [54]. 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 approach, the estimation of the motion u between two images r_k and r_{k+1} is achieved by maximizing the conditional probability of the observation r_{k+1} given r_k and u denoted as $p(r_{k+1}|r_k, u)$. However, for convenience, the problem is usually reformulated in the negative log-domain, where the maximization of the likelihood term is equivalent to the minimization problem

$$\min_{u} \; -\ln[p(r_{k+1}|r_k, u)]. \tag{4}$$

The term in (4) is commonly referred to as the data fidelity term, which expresses the similarity between the displaced image r_{k+1} and the reference image r_k. In the following, details about the observation model used to construct the likelihood, and thus the data fidelity term, are provided.

Observation Model: We consider the problem of cardiac motion estimation from US B-mode images. These images are obtained by performing a series of transformations on the original beamformed radio frequency (RF) data. First, envelop detection is achieved by the demodulation of RF signals. In a second step, log-compression is used to give the final B-mode version of the images. In this paper, we first express the observation model based on known properties of US envelop images, then derive the model for B-mode data. Starting from envelop images, we use the Rayleigh multiplicative noise model that is widely accepted in US image processing [46], [55]. For a given sequence of images, motion is introduced for any pair of consecutive frames using the following observation model

$$\begin{cases} r_k(n) = a_k(n)s(n) \\ r_{k+1}(n) = a_{k+1}(n)s(n - u(n)) \end{cases} \tag{5}$$

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

- $r_k(n)$ and $r_{k+1}(n)$ are the envelop image amplitudes at pixel n for the frames k and $(k + 1)$
- $s_k(n)$ is the unknown noise-free signal corresponding to the nth pixel of the kth frame

- $a_k(n)$ and $a_{k+1}(n)$ are the Rayleigh multiplicative noises corrupting the nth pixel of the kth and $(k + 1)$th frames
- $u(n)$ is the unknown (horizontal or vertical) displacement value of pixel n between the kth and $(k + 1)$th frames.

In the case of B-mode images, the pixel intensities of a given image r_b are expressed as the log-compressed values of the envelope image, $r_b = b \log(r) + g$, where b and g are the dynamic range and linear gain constants [28]. The corresponding observation model is

$$\begin{cases} r_{b,k}(n) = b[a_{b,k}(n) + s_b(n)] + g \\ r_{b,k+1}(n) = b[a_{b,k+1}(n) + s_b(n - u(n))] + g \end{cases} \tag{6}$$

where $a_{b,k}(n) = \log[a_k(n)]$ and $s_b(n) = \log[s(n)]$.

Data Fidelity Term: Using (6) we obtain a likelihood term based on the Rayleigh noise model leading to the following data fidelity term (see [28], [56] for more details)

$$E_{\text{data}}(u) = -2d(u) + 2\log[e^{2d(u)} + 1] + \text{cst} \tag{7}$$

where $d(u) = \frac{1}{b} \sum_{n=1}^{N} [r_{b,k+1}(n + u(n)) - r_{b,k}(n)]$ and $\text{cst} = -\log(2\sigma^4/b)$ is a constant.

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 [54]. 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 [28], where a similarity measure MS_2 has been proposed as a more realistic extension of CD_2.

B. Sparse and Spatial Regularizations

As seen in Section I, motion estimation is an ill-posed inverse problem that requires prior knowledge about the motion field. In this work, this *a priori* information is incorporated into the motion estimation problem through a regularization term E_{reg}. Therefore, the problem of estimating the motion vector u takes the form

$$\min_{u} \; [E_{\text{data}}(u) + E_{\text{reg}}(u)]. \tag{8}$$

In this paper, we seek to exploit the sparse properties of cardiac motion fields. Thus, we introduce a term E_{reg} that combines two different types of regularization. The first one expresses the sparsity of the motion field while the second one exploits a more traditional spatial coherence. More precisely, the regularization term is formulated as follows

$$E_{\text{reg}}(u) = \lambda_d E_{\text{sparse}}(u) + \lambda_s E_{\text{spatial}}(u) \tag{9}$$

where $(\lambda_d, \lambda_s) \in \mathbb{R}^2$ balances the effects of the data fidelity and regularization terms.

Sparse Regularization: The sparse regularization consists in finding the motion u that is best described by a few atoms of a dictionary that contains patterns of either horizontal or vertical training motions. As mentioned in Section II, this approach is based on the assumption of redundancy and self-similarity properties of the motion field. Moreover, the sparse regularization is performed patch-wise, so that each patch of motion

$P_p u$ is constrained to have a sparse representation with respect to the motion dictionary D, *i.e*,

$$E_{\text{sparse}}(u) = \sum_p \| P_p u - D\alpha_p \|_2^2 \qquad (10)$$

where α_p is the sparse coefficient vector associated with the pth patch. Note that the interest of the prior (10) has been previously shown in [42].

Spatial Regularization: The spatial regularization term expresses the smoothness of the motion estimate. It usually takes the form $E_{\text{spatial}}(u) = \phi(\nabla u)$, where ϕ is a penalty function and ∇ is the gradient operator. In this work, we use $\phi(.) = \| . \|_2^2$, which enforces weak spatial gradients on the two motion components and thus ensures a smooth variation of the motion field. This is a very common choice that leads to the following first-order spatial regularization term [57]

$$E_{\text{spatial}}(u) = \|\nabla u\|_2^2. \qquad (11)$$

Finally the global regularization term combining the spatial and sparse priors is

$$E_{\text{reg}}(u) = \lambda_d \sum_p \| P_p u - D\alpha_p \|_2^2 + \lambda_s \|\nabla u\|_2^2. \qquad (12)$$

Note that the above term is computed independently for the horizontal and vertical motion components u_x and u_y. The combination of (7) and (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 the cardiac motions with the spatial smoothness and sparsity constraints (10) and (11).

IV. MOTION ESTIMATION

This section describes the methods proposed for learning the dictionary before and during motion estimation for the offline and adaptive learning strategies. The different steps allowing the final motion field to be estimated are also presented.

A. Offline Dictionary Learning

We start by learning an offline dictionary (see Section II-B) from a set of training motion fields denoted as $u = u_t$. The dictionary D is learned from patches of the training set u_t by solving

$$\min_{D, \alpha_p} \sum_p \| P_p u_t - D\alpha_p \|_2^2 \quad \text{subject to } \forall p, \ \|\alpha_p\|_0 \leq K. \qquad (13)$$

In this work, the DL problem (13) is solved using the ODL algorithm with OMP for the sparse coding step. This choice is motivated by the fact that it has been shown to give more accurate estimation results than the K-SVD algorithm and to be more efficient from a computational point of view [58]. Once the dictionary D has been learned, it is fixed and used for the motion estimation process described below.

B. Motion Field Estimation

Using the data fidelity and regularization terms detailed in Section III, the cardiac motion estimation can be formulated as the following optimization problem

$$\min_{\alpha_p, u} \left\{ E_{\text{data}}(u) + \lambda_d \sum_p \| P_p u - D\alpha_p \|_2^2 + \lambda_s \|\nabla u\|_2^2 \right\}$$
$$\text{subject to } \forall p, \ \|\alpha_p\|_0 \leq K \qquad (14)$$

where E_{data} has been introduced in (7) and the dictionary has been determined using the method described in Section IV-A. Since (14) is hard to solve directly, we adopt an alternate minimization scheme, similar to the half quadratic splitting strategy employed in [42] and [49]. For fixed values of λ_d and λ_s, we alternate optimizations with respect to α_p and u. This process is repeated during a few iterations (typically 4 or 5 [49]) after which the sparsity parameter λ_d is increased. Note that when increasing λ_d, more importance is attached to the distance $\| P_p u - D\alpha_p \|_2^2$, 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 u and D, the sparse coding problem is solved using the OMP algorithm. For all patches, the corresponding sparse vectors are found by solving

$$\min_{\alpha_p} \sum_p \| P_p u - D\alpha_p \|_2^2 \quad \text{subject to } \forall p, \ \|\alpha_p\|_0 \leq K \qquad (15)$$

where $p = 1, \ldots, N_p$, with N_p the total number of patches. Recall that the parameter K indicates the maximum number of non-zero coefficients of α_p. The choice of K is briefly discussed in Section V-C.

2) Motion Field Estimation: Once the sparse codes and the dictionary have been determined, the motion field u is updated (starting from a first initialization $u_0 = 0$) by solving the following minimization problem

$$\min_u \left\{ E_{\text{data}}(u) + \lambda_d \sum_p \| P_p u - D\alpha_p \|_2^2 + \lambda_s \|\nabla u\|_2^2 \right\} \qquad (16)$$

where $E_{\text{data}}(u)$ is given in (7). The solution to (16) can be found by equating the gradient to zero leading to

$$\nabla E_{\text{data}}(u) + \lambda_d \nabla E_{\text{sparse}}(u) + \lambda_s \nabla E_{\text{spatial}}(u) = 0. \qquad (17)$$

Following the optimization approach studied in [28], we use the implicit Euler time marching method [48] to solve (17). At each iteration, the displacement u is estimated as

$$u = (Id + \gamma \lambda_s \Delta)^{-1} [u - \gamma (\nabla E_{\text{data}}(u) + \lambda_d \nabla E_{\text{sparse}}(u))] \qquad (18)$$

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

$$\nabla E_{\text{sparse}}(u) \propto \sum_p P_p^T P_p u - \sum_p P_p^T D\alpha_p \qquad (19)$$

where $\propto$ means proportional to. Note also that the first term $\sum_p P_p^T P_p$ does not depend on u and needs to be computed only once, and that the second term $\sum_p P_p^T D\alpha_p$ represents the reconstruction of the motion field from the current sparse representation.

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

Algorithm 1: Motion Field Estimation for a Pair of Images Using Dictionary Learning

---

Input : $r_{b,1}$, $r_{b,2}$, K, λ_s, λ_d, OuterSteps, InnerSteps

Initialization: $u_0 = 0$, $D_0 =$ offline dictionary

1 $u = u_0$;
2 **for** $k = 1$ **to** *OuterSteps* **do**
3   **for** $j = 1$ **to** *InnerSteps* **do**
        %Dictionary update (optional step)
4          $D_j \leftarrow$ ODL(u,D_{j-1});
        %If not updated: $D_j = D_0$
5   %Sparse coding
6          $\alpha_j \leftarrow$ OMP(D_j,u,K);
        %Motion estimation
7          $u \leftarrow \min\limits_{u} E_{\text{data}}(r_{b,1}, r_{b,2}, u) + \lambda_s \|\nabla u\|_2^2 +$
        $\lambda_d(k) \sum_p \|P_p u - D_j \alpha_{p,j}\|_2^2$;
8   **end**
9 **end**
Output : u (estimated motion field).

Remark: Adaptive Dictionary Learning. When there are missing motion patterns in the training set, e.g., specific to a given cardiac pathology, it might be advantageous to use an adaptive dictionary learning strategy such as the one in Section II-B. Depending on the selected learning strategy, the estimation process consists either of a sparse coding step alone, i.e., offline learning, or of a joint sparse coding and dictionary update step, i.e., adaptive learning. When the dictionary needs to be updated, we start by an offline initialization of the dictionary denoted as D_0. The dictionary D is then updated[3] at each iteration from patches of the current estimate of the motion field u by replacing (15) with (13). In the case of adaptive DL, the motion estimation problem is formulated as follows

$$\min\limits_{D,\alpha_p,u} \left\{ E_{data}(u) + \lambda_d \sum_p \|P_p u - D\alpha_p\|_2^2 + \lambda_s \|\nabla u\|_2^2 \right\}$$

subject to $\forall p$, $\|\alpha_p\|_0 \leq K$.   (20)

V. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method for cardiac motion estimation using a sparse representation and DL. The first experiments are conducted on a dataset of highly realistic simulations for which ground-truth measurements are available. Three sequences of *in vivo* US B-mode images, representing healthy and pathological cases, are then considered.

A. State-of-the-Art Methods

We compare the proposed approach with three state-of-the-art motion estimation methods.

- Block-matching: we consider the block-matching algorithm using the NCC similarity measure [47]. For each patch, a full-grid search is conducted in a defined searching window. Moreover, spatial regularization is induced in

---

[3]As with the offline DL strategy (Section IV-A), the dictionary is updated using the ODL algorithm.

post-processing by the cubic interpolation used to derive sub-pixel valued displacements and dense motion fields. Note that block-matching algorithms are also referred to as speckle tracking methods in the US literature.

- B-spline: in order to evaluate the performance of the sparsity-based regularization term, we consider the method studied in [28]. The algorithm of [28] uses the same similarity measure (CD_2) and spatial regularization as the proposed method. However, a free form deformation model [59], [60] is considered to parameterize the motion. The displacement is 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 [32] for which the intensity-based similarity measure has been replaced by a spatial phase-based metric. Note that the monogenic signal method of [21] has been shown to improve motion estimation with respect to [32].

B. Performance Measures

Endpoint Error: The first performance measure is the endpoint error [21], [61] that can be used for datasets with an available ground-truth, *i.e.*, for the highly realistic simulations. For each pixel $n = 1, \ldots, N$, the endpoint error is the l_2 point distance between the estimate and ground-truth and is defined as $e_n = \sqrt{[u_x(n) - \hat{u}_x(n)]^2 + [u_y(n) - \hat{u}_y(n)]^2}$, where $u_x(n)$, $u_y(n)$ and $\hat{u}_x(n)$, $\hat{u}_y(n)$ are the true and estimated horizontal and vertical displacements at pixel #n.

Error Mean and Standard Deviation: For the kth frame, the mean and standard deviation of the error are also computed to evaluate the estimation performance. These parameters denoted as $m_e(k)$ and $s_e(k)$ are defined as follows

$$m_e(k) = \frac{1}{N} \sum_{n=1}^{N} e_n(k)$$

$$s_e(k) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} [e_n(k) - m_e(k)]^2}. \quad (21)$$

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. For the kth frame, we compute the radial and longitudinal (respectively circumferential) strains for the apical four-chambers and short-axis views. Following the method in [62], we measure along each direction the distance d_k between adjacent points in the kth frame of the cardiac cycle. Strain values are then obtained relatively to the first frame as

$$s_k = \frac{d_k}{d_0} - 1 \quad (22)$$

where d_0 is the distance between points in the first frame. The myocardium is then automatically segmented into 6 regions of
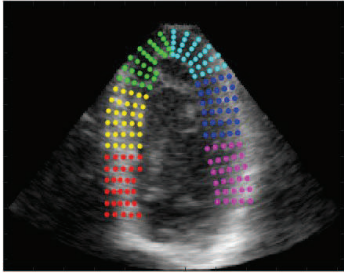
Fig. 1. Initial points used for strain computation for the LADprox sequence considered in Section V-D.

interest for which an average strain value is computed [62]. 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. Through an example of simulated data (see Section V-D), Fig. 1 displays the initial points and segmentation of the myocardium.

C. Dictionary Learning and Regularization Parameters

The choice of the parameters used for DL and for the proposed motion regularization are detailed in this section. The parameters used for the three state-of-the-art algorithms are also provided for each dataset in the following sections.

Dictionary Learning: DL parameters were selected empirically and fixed for all tests. For simplicity, the horizontal and vertical motion dictionaries D_x and D_y were learned separately on patches of size $w = 16 \times 16$ with 1.5 redundancy. Therefore, the number of atoms was set to $n_a = 384$ and the dictionaries were of size $D_{x,y} \in \mathbb{R}^{256 \times 384}$. Different patch sizes classically used in the literature [48], [50] 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 $K = 5$. Generally, the sparsity parameter K is much smaller than the number of atoms in the dictionary $K \ll n_a$ [58] and is related to the noise level [35]. 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. However, we have not observed significant improvement with this adaptive learning strategy for motion estimation (see [56] for more details).

For all tests, the motion dictionaries were learned on one pathological sequence (*i.e.*, LADdist) considered in Section V-D. 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. 2 shows the resulting atoms for the obtained horizontal dictionary D_x with $N_t = 517225$ patches used for learning. This figure allows us to appreciate how the dictionary captures the spatial properties of motion. Specifically, we can observe the presence of atoms containing two distinct regions separated

Fig. 2. Atoms of the dictionary learnt from the displacements of the LADdist sequence (Section V-D), patch size $w = 16 \times 16$, maximum number of non-zero coefficients $K = 5$.

TABLE I
VALUES OF λ_s (SPATIAL REGULARIZATION) USED FOR THE
REALISTIC SIMULATIONS AND *In Vivo* SEQUENCES
(SEE SECTIONS V-D AND V-E)

dataset	LADprox	Normal	Rca	In vivo
λ_s	0.75	1	0.25	0.1

by clear edges. These types of motion patches are well suited to the abrupt changes frequently observed in displacement values, which usually occur in motion boundaries. Other atoms correspond to more or less varying motions that belong to the same region.

Regularization Parameters: The first-order spatial regularization parameter λ_s was manually varied such that $0.01 \leq \lambda_s \leq 10$. The optimal value of λ_s was selected as the value providing the smallest average error for the sequences with available ground-truth. Conversely, the value of λ_s was chosen based on visualization results for the *in vivo* data. The corresponding values of this regularization parameter are reported in Table I. At this point, it is worth mentioning that it would be interesting to consider approaches based on Bayesian inference [63] or on the Stein's unbiased risk estimate (SURE) [64] to estimate this regularization parameter. However, theses approaches are out of the scope of this paper.

For each outer iteration of the proposed method, the sparse regularization parameter λ_d was logarithmically increased from 10^{-3} to 10^2 (see Section IV) in 6 iterations. The same procedure was applied for the parameters of the three state-of-the-art methods considered in this paper. The parameter values returning the smallest average error were selected for all the experiments.

D. Realistic Simulations

This section evaluates the performance of the proposed method on a set of highly realistic simulations of B-mode US data [62]. The images and motion fields were generated using real data combined with synthetic biomechanical and US models. More precisely, the images were generated by varying
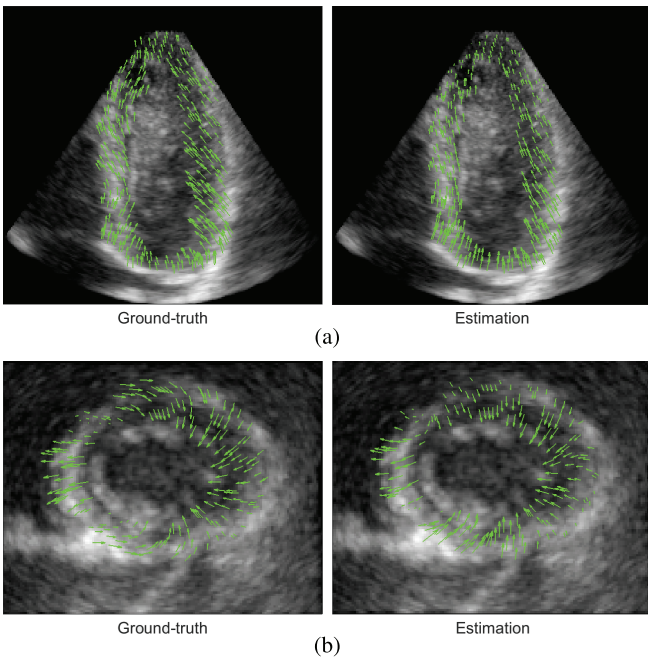
Fig. 3. Ground-truth and estimated meshes of the 5th frame of the LADprox sequence: (a) apical four chambers view, (b) short axis view.

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
Time	66.17	4.55	2.53	**0.20**

the parameters of a highly realistic E/M model, resulting in 8 sequences of different pathophysiological conditions [62]. To the best of our knowledge, it is the most realistic simulation method available in the recent literature. The reader is invited to read [11], [62] for more details about the data generation process. A ground-truth was available for all measurements. We conducted tests on 4 sequences (LADdist, LADprox, Normal and Rca) of 3D images (of size $224 \times 176 \times 208$ voxel^3, voxel size $0.7 \times 0.9 \times 0.6$ mm^3, frame rate 21-23 Hz [62]) from which we extracted 2D slices for apical four chambers and short-axis views. All sequences contained 34 images that span a full cardiac cycle and represent either healthy heart motion (*i.e.*, Normal), or ischemic cases (*i.e.*, LADdist, LADprox and Rca), which correspond to distal and proximal occlusions of the left anterior descending artery and occlusions of right coronary artery, respectively [62].[4]

The motion dictionaries were learned on one pathological sequence (*i.e.*, LADdist). The method was then tested on the remaining sequences containing (i) one healthy sequence, (ii) one sequence with a pathology that is similar to the training sequence and (iii) one sequence with a distinct pathology. Note that four other sequences (LBBB, LBBBsmall, LCX and sync) were available in the dataset and were also tested. However, the results obtained with these sequences were not significantly different from the ones presented in this paper (see [56] for more details and additional results conducted on synthetic data with controlled ground-truth).

For all algorithms, the parameters giving the best mean endpoint error were selected. These parameters are given in Table I for the proposed method. For the B-spline algo-

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

rithm, the mesh window size between the B-spline control points was $w_{\text{B-spline}} = 15 \times 15$ and the 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 and normal sequences and to $\lambda_0 = 0.1$ for the Rca sequence. The number of refinement steps was fixed to $N_p = 4$. The window size for the block-matching algorithm was set to $w_{\text{BM}} = 16 \times 16$. 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 block-matching methods. Fig. 3 displays a typical example of estimated motions using the proposed method compared to the corresponding true displacement meshes. The estimated motion field is clearly consistent with the ground-truth.

1) Endpoint Error: The results obtained for the three sequences LADprox, Normal and Rca in terms of mean endpoint error are summarized in Table II. 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 standard deviation for the endpoint error calculated using the entire sequence. The average execution times associated with the different methods are also reported in Table II. The proposed method is 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.

In order to appreciate the performance of the methods during the cardiac cycle, Fig. 5 shows the time evolution of the estimates for all the sequences. The difference between the performance of the methods is less pronounced at the end of the sequence, where the displacements are small. It is also clear that the block-matching 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 frames of the Rca sequence. However, for the normal sequence, the proposed method is outperformed
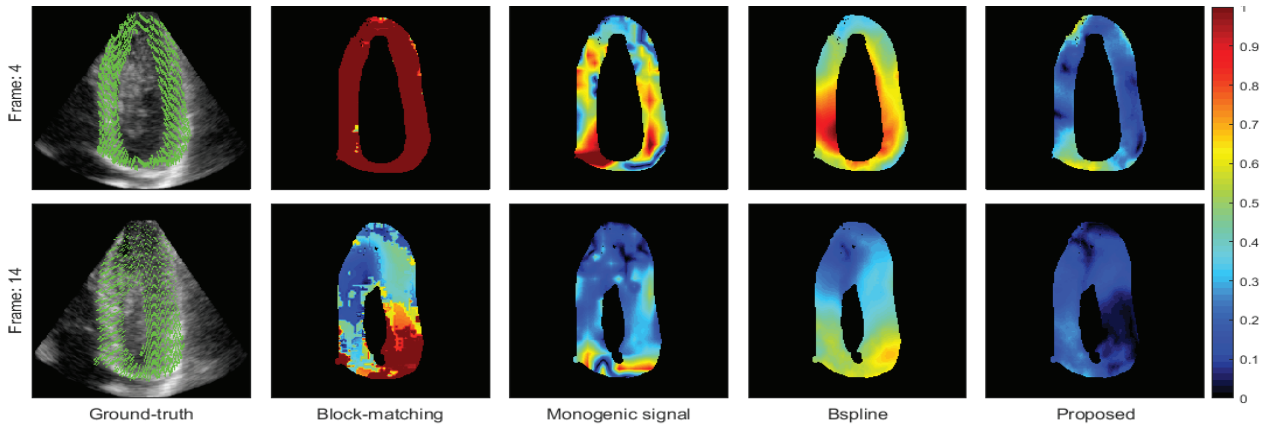
Fig. 4. Error map and ground-truth meshes of the 4th and 14th frames of the LADprox sequence.
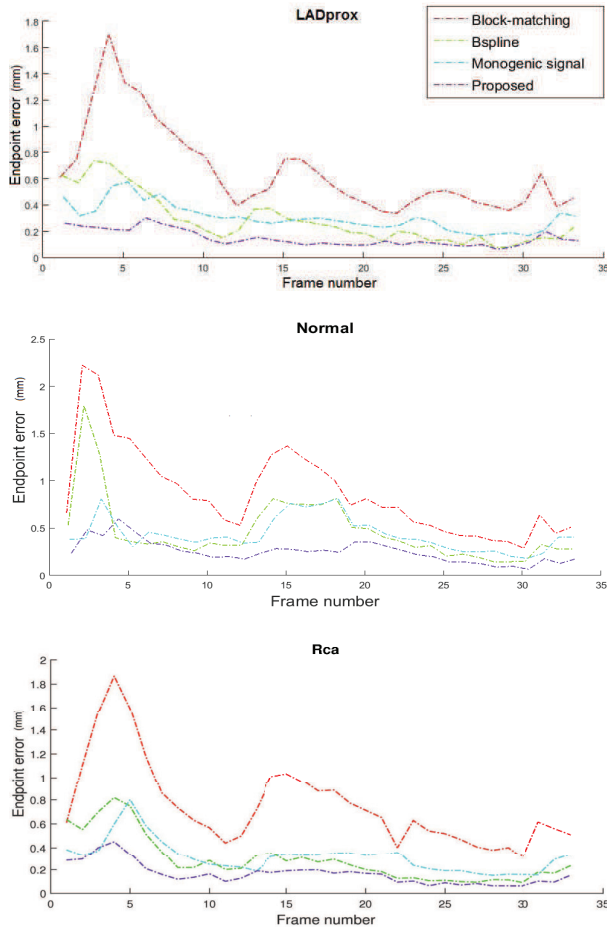

Fig. 5. Mean endpoint error (in mm) over the LADprox, normal and Rca sequences.

the cardiac cycle are less pronounced, which is an interesting property of the proposed method.

In order to examine the local behavior of the algorithms, error maps of the displacement estimates are displayed in Fig. 4. The error maps match the 4th and 14th frames of the LADprox sequence, which respectively correspond to the maximum average displacement value in systole and diastole phases (additional results, including the 12th and 20th frames are available in [56]). Overall, all other tested sequences present a similar behavior. As shown in previous results, the block-matching algorithm is outperformed by all methods. This can be explained by the resulting integer valued 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 data attachment term used in this paper. The B-spline 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. The monogenic signal algorithm does not suffer from over-regularization, but still performs poorly for the 4th frame of the sequence. This is mainly due to the fact that the phase constancy assumption considered for this method holds less for large displacements. On the other hand, the proposed method presents less errors even for large displacements. 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 probably due to the influence of patches including estimates from outside this region.

Finally, we have investigated the possibility of using the proposed spatial and sparse regularizations jointly with other data fidelity terms. More precisely, Table III shows the average mean and standard deviations of the endpoint errors for three different similarity measures, i.e., mutual information (MI), SSD and the data fidelity term in (7) (CD$_2$), for the

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 also that the differences in estimation accuracy between the beginning (large displacements) and the end (small displacements) of

CD2	MI	SSD
0.147±0.088	0.157±0.091	0.173±0.105

Dimention	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
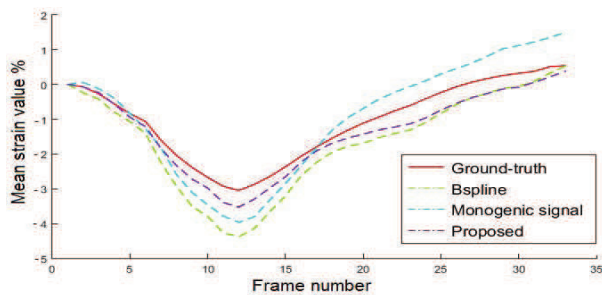
Fig. 6. Mean longitudinal strain values (LADprox sequence).

Fig. 7. Examples of estimated motion fields for a systolic frame in the healthy *in vivo* sequence.

LADprox sequence. (Note that the regularization settings were the same for all tests.) 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 II.

2) Strain Accuracy: The average means and standard deviations of strain errors for the LADprox sequence are summarized in Table IV. The proposed method provides the most accurate estimates in terms of endpoint error. Fig. 6 shows the average longitudinal strains for the LADprox sequence. We can observe that the strain estimated with the proposed method is closer to the ground-truth.

E. 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 the 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 ultrasound machine equipped with a 3Sc-RS 1.5-4.0 MHz transducer array (center frequency 2.75 MHz).
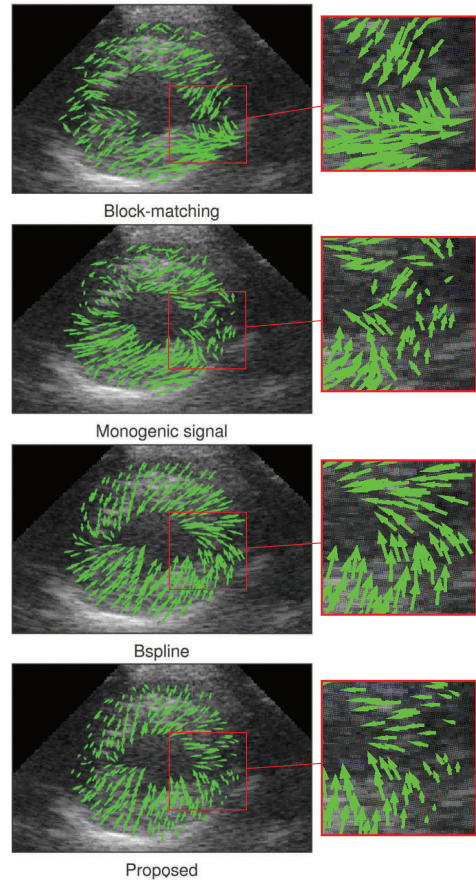
The healthy reference sequence was acquired from a 37 years old male volunteer with no cardiovascular disease, whereas the second sequence (*i.e*, patient $n°1$) corresponds to an 88 years old female who has undergone non ST elevation myocardial infarction in the left descending artery territory. Finally, 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.

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 were $\lambda_s = 0.1$ and $\lambda_{\text{B-spline}} = 0.1$ for the regularization of the proposed and B-spline methods 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 block-matching method. For all other parameters, we used the values from the tests conducted on the highly realistic simulation dataset. The dictionaries D_x and D_y were learned using the LADdist sequence. Fig. 7 shows the
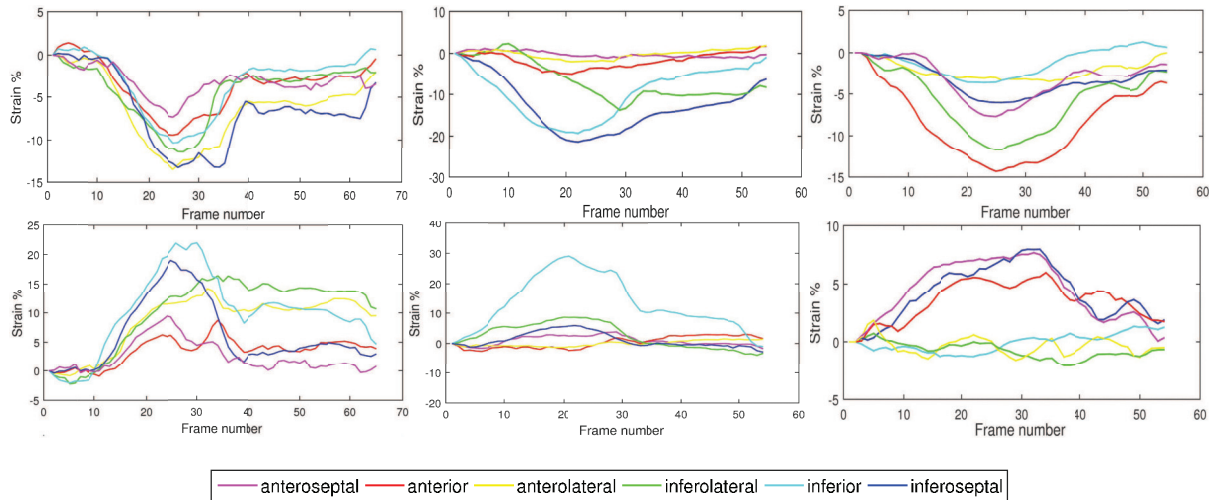
Fig. 8. Strain curves of *in vivo* sequences: circumferential (top row) and radial (bottom row) strains of (a) the healthy sequence, (b) the patient $n°1$ and (c) the patient $n°2$.

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 displacements. 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. Finally, the proposed method provides a good compromise in terms of smoothness and local motion estimation. Overall, these observations confirm the outcome of the tests conducted on the highly realistic simulations dataset in Section V-D.

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. 8 shows the results of the segmental strain values. 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 region 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 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 findings related to the visual examination of this sequence by the cardiologist. A more quantitative analysis of these *in vivo* images is beyond the scope of this paper. However, these final results support the suitability of the method regarding clinical assessment with real cardiac ultrasound data.

F. Adaptive Dictionary Learning

This section briefly explores the results of cardiac motion estimation with adaptive DL, *i.e.*, updated dictionary (see Section II-B 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 IV. In order to illustrate the possible differences between the offline and adaptive DL strategies, we use the healthy *in vivo* sequence introduced above. Note that in contrast with the offline strategy, the updated dictionary is initialized using random patches of the estimated motion field.

For a systolic frame in the healthy sequence, the adaptive and offline DL strategies result in close estimates, with small relative differences (*i.e*, mean and standard deviation: $9.115\% \pm 6.757\%$). For the rest of the sequence, the relative differences remain overall negligible. (For additional results see [56]). 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.

VI. CONCLUSIONS

This paper introduced a method for estimating the cardiac motion of 2D US images with a regularization based on a

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 using a regularization smoothing term based on the gradient of the motion field and by exploiting a sparse motion prior based on DL. Our results showed the effectiveness of these regularizations for cardiac motion estimation. In terms of motion and strain accuracy, the results obtained with highly 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 a 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. In this work, we have addressed the problem of 2D motion estimation, which can present some shortcomings, such as out-of-plane motion and limited geometrical information, that could be overcome in 3D. Nevertheless, it should be pointed out that in contrast with 2D imagery, 3D UI is affected by the problems of frame rate and image spatial resolution in the azimuthal direction and thus, does not necessarily provide better motion estimation results. Furthermore, it is worth mentioning that the data fidelity and regularization terms used in the actual formulation are not inherently limited to 2D and could be extended to 3D. In the same way, the dictionaries could be learned separately for each direction or jointly for the 3 dimensions. The differences between these two strategies of learning the dictionary have not been investigated in this paper, but would also deserve consideration in future work. Another research prospect would be to study the interest of adaptive dictionary learning techniques for applications in which the training database is updated periodically. Furthermore, the proposed approach has not exploited the temporal properties of cardiac motion. Integrating this aspect could be performed by using more than two consecutive frames or by learning motion dictionaries that take into account the sparsity of the motion versus time. Another possible prospect concerns the problem of outliers. Considering potential model deviations or violations of smoothness assumptions (*e.g.*, motion boundaries) in the current approach for robust motion estimation is clearly an interesting prospect.

The dictionary and the resulting sparse codes could also be used to gain additional information about the motion estimates. 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.

VII. Acknowledgments

References

[1] W. H. Organization. (2012). *World Health Statistics*. [Online]. Available: http://www.who.int/healthinfo/statistics/

[2] T. P. Abraham, V. L. Dimaano, and H.-Y. Liang, "Role of tissue Doppler and strain echocardiography in current clinical practice," *Circulation*, vol. 116, no. 22, pp. 2597–2609, 2007.

[3] C. Cottrell and J. N. Kirkpatrick, "Echocardiographic strain imaging and its use in the clinical setting," *Expert Rev. Cardiovascular Therapy*, vol. 8, no. 1, pp. 93–102, 2010.

[4] J. D'hooge *et al.*, "Two-dimensional ultrasonic strain rate measurement of the human heart *in vivo*," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 49, no. 2, pp. 281–286, Feb. 2002.

[5] A. M. Shah and S. D. Solomon, "Myocardial deformation imaging: Current status and future directions," *Circulation*, vol. 125, no. 2, pp. e244–e248, 2012.

[6] G. R. Sutherland, G. Di Salvo, P. Claus, J. D'Hooge, and B. Bijnens, "Strain and strain rate imaging: A new clinical approach to quantifying regional myocardial function," *J. Amer. Soc. Echocardiography*, vol. 17, no. 7, pp. 788–802, 2004.

[7] C. B. Compas *et al.*, "Radial basis functions for combining shape and speckle tracking in 4D echocardiography," *IEEE Trans. Med. Imag.*, vol. 33, no. 6, pp. 1275–1289, Jun. 2014.

[8] C. B. Compas *et al.*, "A combined shape tracking and speckle tracking approach for 4D deformation analysis in echocardiography," in *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, Barcelona, Spain, May 2012, pp. 458–461.

[9] E. Y. Wong *et al.*, "4-D echocardiography assessment of local myocardial strain using 3-D speckle tracking combined with shape tracking," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Jul. 2013, pp. 100–103.

[10] O. Mirea, J. Duchenne, and J. Voigt, "Recent advances in echocardiography: Strain and strain rate imaging," *Research*, vol. 5, no. 787, p. F1000, 2016.

[11] M. Alessandrini *et al.*, "A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database," *IEEE Trans. Med. Imag.*, vol. 34, no. 8, pp. 1436–1451, Jul. 2015.

[12] Y. Nagata *et al.*, "Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software," *J. Amer. Soc. Echocardiograph.*, vol. 28, no. 6, pp. 630–641, 2015.

[13] J.-U. Voigt *et al.*, "Definitions for a common standard for 2D speckle tracking echocardiography: Consensus document of the EACVI/ASE/industry task force to standardize deformation imaging," *Eur. Heart J.-Cardiovascular Imag.*, vol. 16, no. 1, p. 1, Dec. 2014.

[14] J. D'hooge, *et al.*, "Two-dimensional speckle tracking echocardiography: Standardization efforts based on synthetic ultrasound data," *Eur. Heart J.-Cardiovascular Imag.*, vol. 17, no. 6, pp. 693–701, Aug. 2015.

[15] K. Farsalinos, A. Daraban, S. Ünlü, J. Thomas, L. Badano, and J. Voigt, "Head-to-head comparison of global longitudinal strain measurements among nine different vendors: The EACVI/ASE inter-vendor comparison study," *J. Amer. Soc. Echocardiograph.*, vol. 28, no. 10, pp. 1171–1181, 2015.

[16] P. Baraldi, A. Sarti, C. Lamberti, A. Prandini, and F. Sgallari, "Evaluation of differential optical flow techniques on synthesized echo images," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 3, pp. 259–272, Mar. 1996.

[17] G. E. Mailloux, A. Bleau, M. Bertrand, and R. Petitclerc, "Computer analysis of heart motion from two-dimensional echocardiograms," *IEEE Trans. Biomed. Eng.*, vol. BME-34, no. 5, pp. 356–364, May 1987.

[18] G. 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. MI-8, no. 2, pp. 143–153, Jun. 1989.

[19] 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, pp. 525–536, Apr. 2005.

[20] 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," *Ultrasonic*, vol. 41, no. 9, pp. 743–753, 2004.

[21] 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, pp. 1084–1095, Mar. 2013.

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

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

[24] 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)*, Los Alamitos, CA, USA, Apr. 2007, pp. 944–947.

[25] C. Kontogeorgakis, M. Strintzis, N. Maglaveras, and I. Kokkinidis, "Tumor detection in ultrasound B-mode images through motion estimation using a texture detection algorithm," in *Proc. Comput. Cardiol.*, Sep. 1994, pp. 117–120.

[26] F. Yeung, S. F. Levinson, and K. J. Parker, "Multilevel and motion model-based ultrasonic speckle tracking algorithms," *Ultrasound Med. Biol.*, vol. 24, no. 3, pp. 427–442, 1998.

[27] M. J. Ledesma-Carbayo *et al.*, "Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1113–1126, Sep. 2005.

[28] A. Myronenko, X. Song, and D. J. Sahn, "Maximum likelihood motion estimation in 3D echocardiography through non-rigid registration in spherical coordinates," in *Proc. 5th Int. Conf. Functional Imag. Modeling Heart (FIMH)*, Nice, France, Jun. 2009, pp. 427–436.

[29] J. Woo, B.-W. Hong, C.-H. Hu, K. K. Shung, C. C. J. Kuo, and P. Slomka, "Non-rigid ultrasound image registration based on intensity and local phase information," *J. Signal Process. Syst.*, vol. 54, nos. 1–3, pp. 33–43.

[30] A. Elen *et al.*, "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, pp. 1580–1591, Nov. 2008.

[31] K. McLeod, A. Prakosa, T. Mansi, M. Sermesant, and X. Pennec, *An Incompressible Log-Domain Demons Algorithm for Tracking Heart Tissue*. Berlin, Germany: Springer, 2012, ch. 6, pp. 55–67.

[32] 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, pp. 525–536, Apr. 2005.

[33] N. Parajuli *et al.*, *Sparsity and Biomechanics Inspired Integration of Shape and Speckle Tracking for Cardiac Deformation Analysis*. Maastricht, The Netherlands: Springer, Jun. 2015, pp. 57–64.

[34] N. Parajuli *et al.*, *Integrated Dynamic Shape Tracking and RF Speckle Tracking for Cardiac Motion Analysis*. Athens, Greece: Springer, Oct. 2016, pp. 431–438.

[35] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[36] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 457–464.

[37] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2691–2698.

[38] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[39] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Model. Simul.*, vol. 7, no. 1, pp. 214–241, 2008.

[40] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, Montreal, QC, Canada, 2009, pp. 689–696.

[41] X. Shen and Y. Wu, "Sparsity model for robust optical flow estimation at motion discontinuities," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 2456–2463.

[42] K. Jia, X. Wang, and X. Tang, "Optical flow estimation using learned sparse model," in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 2391–2398.

[43] X. Huang *et al.*, "Contour tracking in echocardiographic sequences via sparse representation and dictionary learning," *Med. Image Anal.*, vol. 18, no. 2, pp. 253–271, 2014.

[44] J. A. Onofrey, L. H. Staib, and X. Papademetris, *Segmenting the Brain Surface from CT Images with Artifacts Using Dictionary Learning for Non-rigid MR-CT Registration*. Isle Skye, U.K.: Sabhal Mor Ostaig, Jun. 2015, pp. 662–674.

[45] I. Oksuz, A. Mukhopadhyay, M. Bevilacqua, R. Dharmakumar, and S. A. Tsaftaris, "Dictionary learning based image descriptor for myocardial registration of CP-BOLD MR," in *Proc. 18th Int. Conf. Med. Image Comput. Comput. Assist. Intervent (MICCAI)*, Munich, Germany, Oct. 2015, pp. 205–213.

[46] J. Goodman, *Speckle Phenomena in Optics: Theory and Applications*. Winterville, NC, USA: Roberts & Company, 2007.

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

[48] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[49] J. Sulam and M. Elad, *Expected Patch Log Likelihood With a Sparse Prior*. Cham, Switzerland: Springer, 2015, pp. 99–111.

[50] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[51] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[52] P. K. Y. Pati and R. Rezaiifar, "Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 2391–2398.

[53] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc.*, vol. 58, no. 1, pp. 267–88, 1996.

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

[55] 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, pp. 65–77, Jan. 1994.

[56] N. Ouzir, A. Basarab, H. Liebgott, B. Harbaoui, and J.-Y. Tourneret, "Motion estimation in echocardiography using sparse representation and dictionary learning," Dept. IRIT/INP-ENSEEIHT, Univ. Toulouse, Toulouse, France, Tech. Rep. IRIT-ENSEEIHT, Feb. 2017.

[57] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[58] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[59] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," in *Proc. 13th Annu. Conf. Comput. Graph. Interaction Techn.*, New York, NY, USA, 1986, pp. 151–160.

[60] 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, pp. 712–721, Aug. 1999.

[61] M. Otte and H.-H. Nagel, "Optical flow estimation: Advances and comparisons," in *Proc. 3rd Eur. Conf. Comput. Vis. (ECCV)*, vol. 1. Stockholm, Sweden, 1994, pp. 51–60.

[62] M. Alessandrini *et al.*, "Detailed evaluation of five 3D speckle tracking algorithms using synthetic echocardiographic recordings," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1915–1926, Aug. 2016.

[63] M. Pereyra *et al.*, "Tutorial on stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, Mar. 2016.

[64] 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, pp. 1540–1554, Sep. 2008.

Nora Ouzir received the degree from the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications and the M.Sc. degree in signal processing from the University of Toulouse in 2015. She is currently pursuing the Ph.D. degree with the Image Processing Group, IRIT Laboratory. Her research interests are focused on motion estimation and dictionary learning for ultrasound images.

Adrian Basarab received the M.S. and Ph.D. degrees in signal and image processing from the National Institute for Applied Sciences of Lyon, France, in 2005 and 2008. Since 2009, he has been an Associate Professor with the University Paul Sabatier Toulouse 3 and a member of the IRIT Laboratory (UMR CNRS 5505). His research interests include medical imaging, and more particularly motion estimation, inverse problems and ultrasound image formation. He is currently an Associate Editor of digital signal processing.

Hervé Liebgott received the Ph.D. degree in 2006. During his Ph.D. degree, he visited the Technical University of Denmark for six months. In 2014, he was an Invited Researcher with the University of Montreal, Canada, for six months. He is currently a Professor with the University of Lyon1. His research focuses on image/signal processing applied to medical ultrasound. From 2014 to 2016, he served as an Associate Member of the IEEE-BISP. In 2014, he became a member of the IEEE IUS Technical Program Committee. He is a Junior Member of the French Institut Universitaire de France. He is the Head of the Ultrasound Imaging Team, CREATIS. He has been an Associate Editor of the IEEE UFFC Transactions since 2015.

Brahim Harbaoui received the degree from the University of Medicine Lyon 1 and the M.Sc. degree from University Lyon 1 in 2015, where he is currently pursuing the Ph.D. degree with the Research Group CREATIS. On top of being an interventional cardiologist, he is involved in research in the fields of acute cardiac care, coronary, and valvular heart diseases. He has contributed to several fundamental works in signal processing and treatment by bringing his clinical point of view.

Jean-Yves Tourneret (SM'08) received the Ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT), Toulouse, in 1989, and the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992. He is currently a Professor with ENSEEIHT, University of Toulouse, and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo methods. He has been involved in the organization of several conferences, including the European conference on signal processing EUSIPCO'02 (Program Chair), the International Conference ICASSP'06 (plenaries), the Statistical Signal Processing Workshop SSP'12 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2013 (local arrangements), the Statistical Signal Processing Workshop SSP'2014 (special sessions), the workshop on machine learning for signal processing MLSP'2014 (special sessions). He has been the General Chair of the CIMI Workshop on Optimization and Statistics in Image Processing hold in Toulouse in 2013 (with F. Malgouyres and D. Kouamé) and the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2015 (with P. Djuric). He has been a member of different technical committees, including the Signal Processing Theory and Methods committee of the IEEE Signal Processing Society from 2001 to 2007 and from 2010 to 2015. He has been serving as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2011 and since 2015 and the *EURASIP Journal on Signal Processing* since 2013.