This is an author's version published in: **http://oatao.univ-toulouse.fr/24124**

Robust Optical Flow Estimation in Cardiac Ultrasound Images Using a Sparse Representation

Nora Ouzir[iD], *Student Member, IEEE*, Adrian Basarab[iD], *Member, IEEE*,
Olivier Lairez, and Jean-Yves Tourneret[iD], *Senior Member, IEEE*

Abstract—**This paper introduces a robust 2-D cardiac motion estimation method. The problem is formulated as an energy minimization with an optical flow-based data fidelity term and two regularization terms imposing spatial smoothness and the sparsity of the motion field in an appropriate cardiac motion dictionary. 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 spatial and sparse regularizations. The motion fields and the weights are computed jointly using an iteratively re-weighted minimization strategy. The proposed robust approach is evaluated on synthetic data and realistic simulation sequences with available ground-truth by comparing the performance with state-of-the-art algorithms. Finally, the proposed method is validated using two sequences of *in vivo* images. The obtained results show the interest of the proposed approach for 2-D cardiac ultrasound imaging.**

Index Terms—**Cardiac ultrasound, robust motion estimation, optical flow, sparse regularization, dictionary learning.**

I. INTRODUCTION

ULTRASOUND imaging (UI) has been used succesfully in many practical applications because of its low cost, non-ionisation properties and reduced discomfort for the patients. In the context of cardiac ultrasound (US), the high temporal resolution of echocardiography allows the fast and complex motion of the heart to be captured. Cardiac motion estimation from US images is therefore an esssential tool for the diagnosis of cardiovascular diseases [1]–[3].

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

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

O. Lairez is with INSERM, UMR 1048, Institut des Maladies Métaboliques et Cardiovasculaires, CHU de Toulouse, Université Paul Sabatier, Toulouse, France.

Speckle tracking algorithms are among the most widely used techniques for cardiac motion estimation [4], [5]. These methods exploit the image characteristics to match blocks between consecutive frames using a similarity measure, *e.g.,* cross-correlation [6] or the sum of absolute differences [7]. Elastic registration methods use a non-rigid transformation to represent the motion. In particular, B-splines have been widely used as a parametric model for cardiac motion [8]–[10]. Groupwise motion estimation has also been investigated for cardiac US [8], [11]. In contrast with the pairwise approach, groupwise methods make use of the whole image sequence, allowing temporal consistency to be incorporated. In the context of UI, some methods have combined the B-mode data with Doppler imaging using, *e.g.*, single [12] or multiplane images [10]. These methods benefit from the high temporal resolution of Doppler imaging and have been shown to be more resilient to image noise.

Another well-established motion estimation method for UI is optical flow (OF) [6], [13], [14]. 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. OF methods, as many motion estimation strategies, are generally formulated as ill-posed inverse problems. It is therefore 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 [15], [16]. Smoothness can be incorporated into the motion estimation problem by means of explicit regularization constraints [17] or using parametric motion models, *e.g.*, by imposing an affine transformation on the motion vectors [18]. Finally, recent works have investigated sparse regularizations for motion estimation [19]–[22]. In these works, a classical smoothness regularization (*i.e.*, total variation) is combined with a patch-wise sparse regularization. The latter imposes sparsity on all motion patches when decomposed on a dictionary of typical cardiac motion patterns.

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 so-called 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 [23], [24]. 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 [23]–[25] prior to motion estimation. 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 [26] and [27] and an iteratively re-weighted approach has been considered in [28]. Many other strategies have been studied for robust motion estimation, *e.g.*, local OF methods based on the least median of squares [29], [30], the use of multiple images to address the problem of drift between frames [31] or feature extraction techniques to account for illumination variations [31]. Motion discontinuities have been taken into account in [32] for myocardial boundaries. In this work, image segmentation is used to down-weight the epicardial motions. In the context of 3D US, an M-estimator-based OF method using a robust spatial smoothness term is proposed in [33] for brain images. A specific similarity measure accounting for temporal speckle correlation is also considered in [9] for 3D cardiac US. In [34], 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 [35]. In this work, a robust weighting function is introduced within an iteratively re-weighted minimization strategy to deal with uncorrelated radio-frequency data.

This work focuses on 2D echocardiography, which is still more routinely used than emerging 3D UI in clinical practice.

The proposed motion estimation method is based on a recent work that has shown the benefits of a sparsity-based prior for cardiac motion estimation [21]. In addition to exploiting the sparse regularization promoted in [21], this paper investigates a new motion estimation method mitigating the effect of outliers. The problem is formulated within a general OF-based energy minimization framework. As in [21], two 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 regularization terms. Finally, an iterative strategy is used to jointly estimate the motions, the sparse codes and the corresponding weights.

This paper is organized as follows. Section II briefly reviews the theory of M-estimators and weight functions. Details about the proposed robust motion estimation method and the implementation strategy are provided in Section III. In Section IV, synthetic experiments are used to investigate the interest of robustness for the data fidelity and regularization terms. The proposed method is then compared to the previous non-robust cardiac motion estimation method of [21] and to the robust OF method of [26], 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 Sections V and VI.

II. WEIGHTING FUNCTIONS

M-estimators are robust functions that address the issue of outliers by reducing their impact on the estimates. In this work, we use weight functions associated with 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. These functions depend on the first derivative of their corresponding M-estimator, *i.e.*, $w(e_i) = r'(e_i)/e_i$, where w denotes the weight function, r is an appropriate objective function, r' is the derivative of r and 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 paper, 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

$$w_L(e_i) = \frac{1}{1 + (e_i/c\sigma)^2} \quad (1)$$

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 [28]. This function provides a hard rejection of outliers in comparison with the Lorentzian and is defined by

$$w_B(e_i) = \begin{cases} [1 - (e_i/c\sigma)^2]^2, & |e_i| \le c\sigma \\ 0, & |e_i| > c\sigma \end{cases} \quad (2)$$
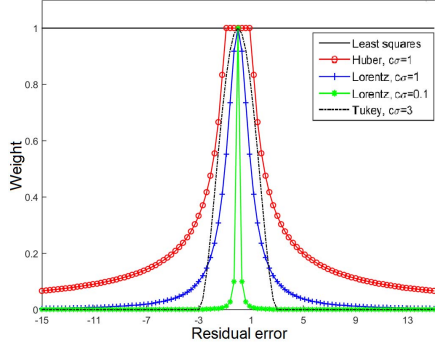
Fig. 1. Weights associated with different M-estimator functions.

A. 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 $\boldsymbol{D}$ [21], [22]. The motion field is obtained through the minimization of the resulting cost function

$$\min_{\boldsymbol{\alpha},\boldsymbol{U}} \left\{ E_Q(\boldsymbol{U}, \boldsymbol{I}) + \lambda_s E_S(\boldsymbol{U}) + \lambda_d E_W(\boldsymbol{U}, \boldsymbol{\alpha}) \right\} \qquad (4)$$

where $\boldsymbol{\alpha}$ is the sparse coefficient vector, λ_d 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 [21]. In a second step, the motion of each pair of test images is estimated using the minimization problem (4). In this work, we make use of robust weight functions (see Section II) in order to mitigate the effect of outliers. Further details about the way the data fidelity term and regularizations of (4) are defined are provided in the following sections.

B. Robust Data Fidelity Term

This work considers an OF data fidelity term for motion estimation. OF estimation methods have been investigated in the context of cardiac motion estimation [13], [14], and more recently, for cardiac motion estimation with sparsity regularization [22]. In differential OF methods, the motion is usually estimated by linking the spatial and temporal image intensity variations within a least squares estimation. 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 $\boldsymbol{Q} = \text{diag}[\boldsymbol{q}(1), \ldots, \boldsymbol{q}(N)] \in \mathbb{R}^{N \times N}$. 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 $\nabla \boldsymbol{I}^T = [\text{diag}(\nabla \boldsymbol{I}_x), \text{diag}(\nabla \boldsymbol{I}_y)] \in \mathbb{R}^{N \times 2N}$, with $\nabla \boldsymbol{I}_x$ and $\nabla \boldsymbol{I}_y$ the spatial intensity gradients in both directions and ∇ indicating the gradient operator, the proposed robust data fidelity term is finally defined as follows

$$E_Q(\boldsymbol{U}, \boldsymbol{I}) = \| \boldsymbol{Q}^{1/2}(\partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T \boldsymbol{U}) \|_2^2. \qquad (5)$$

where $c > 0$ is also a constant and $\sigma > 0$ a scale parameter. Fig. 1 shows the shapes of the weight functions (1) and (2) 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.

In this work, the weight functions (1) and (2) will be used to mitigate the effect of outliers for the data fidelity term (5) and the regularizations (6) and (8) that will be considered for motion estimation.

Parameter Estimation

The two weight functions in (1) and (2) 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. The scale parameter σ can be computed jointly with the current estimate [36]. 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) [36] defined as

$$\widehat{\sigma}_{\text{MAD}} = \sigma_0 \underset{i=1,\ldots,N}{\text{median}} [|e_i - \text{median}(\boldsymbol{e})|] \qquad (3)$$

with $\boldsymbol{e} = (e_1, \ldots, e_N)^T$. 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 [28]). 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. 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.

The spatially-variant data weights Q are obtained using a robust weight function (*i.e.*, w_L or w_B) introduced in Section II and the corresponding residual error e_d such that, $q(i) = w_{L,B}(e_d(i)), \forall i$ and $e_d = \partial_t I + \nabla I^T U$. As explained in Section II, the scale parameter σ_d is iteratively and jointly estimated with the motion estimates using (3).

C. Robust Spatial Regularization

The spatial regularization ensures the smoothness of the motion estimates. A classical choice is $E_{\text{spatial}}(U) = \|\nabla U\|_2^2$ [17]. 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 ($s(i)$ close to 0), while still imposing smoothness in homogeneous regions (where $s(i)$ is close to 1). The associated robust spatial regularization term can be written as follows

$$E_S(U) = \|S^{1/2}\nabla U\|_2^2 \qquad (6)$$

where $S = \text{diag}[s(1),\ldots,s(2N)] \in \mathbb{R}^{2N\times 2N}$. The spatial weighting matrix S is computed using the error $e_s = \nabla U$ associated with the magnitude of the motion field gradient such that, $s(i) = w_{L,B}(e_s(i))$ for $i = 1,\ldots,2N$. Higher values of 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 (3). Note that the horizontal and vertical motion fields u and v are weighted separately. Note also that distinct weights could be assigned to the horizontal and vertical gradients, 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.

D. Robust Sparse Regularization

The proposed sparse regularization consists in finding the motion field U that is best described by a few atoms of a dictionary containing typical patterns of cardiac motion. This strategy has been used sucessfully for cardiac motion estimation in [21] and [22]. In these works, the sparse regularization was performed patch-wise, so that each pair of motion patches $P_p U$ is constrained to have a sparse representation with respect to the motion dictionary D, *i.e*,

$$E_{\text{sparse}}(U, \alpha) = \sum_p \|P_p U - D\alpha_p\|_2^2 \qquad (7)$$

where $P_p \in \mathbb{R}^{2n\times 2N}$ is an operator that extracts the pth pair of patches in the horizontal and vertical directions from U,[1]

---

[1] P_p is a block diagonal matrix whose blocks are P_p', which extracts the pth patch in the horizontal or vertical direction.

$\alpha \in \mathbb{R}^{2q\times 2N_p}$ is a sparse coding matrix whose columns are $\alpha_p = (\alpha_{u,p}^T, \alpha_{v,p}^T)^T$, $D \in \mathbb{R}^{2n\times 2q}$ is a block diagonal matrix whose blocks are $D_u \in \mathbb{R}^{n\times q}$ and $D_v \in \mathbb{R}^{n\times q}$, 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 α_p in (7) are constrained to be sparse, *i.e.*, $\|\alpha_{u,p}\|_0 \leq K$ and $\|\alpha_{v,p}\|_0 \leq K$, with K a fixed maximum number of non zero coefficients.

The sparse prior in (7) 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 too far from the patterns of the dictionary. 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 (5) and (6). 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 ($w_p(i)$ close to 0), whereas the patch elements that are sufficiently close to the dictionary have higher weights ($w_p(i)$ close to 1). The robust sparse regularization term considered in this work is formulated as follows

$$E_W(U, \alpha) = \sum_p \|W_p^{1/2}(P_p U - D\alpha_p)\|_2^2 \qquad (8)$$

where $W_p = \text{diag}[w_{u,p}(1),\ldots,w_{u,p}(n), w_{v,p}(1),\ldots, w_{v,p}(n)] \in \mathbb{R}^{2n\times 2n}$ 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 according to $w_{u,p}(i) = w_{L,B}(e_{u,p}(i))$ and $w_{v,p}(i) = w_{L,B}(e_{v,p}(i))$, with $e_{u,p} = P_p'u - D_u\alpha_{u,p}$ and $e_{v,p} = P_p'v - D_v\alpha_{v,p}$ the residual sparse coding errors of the pth patch. Note that the corresponding scale parameters σ_u and σ_v are computed using the global reconstructed errors $e_{u,g} = \sum_p P_p'^T e_{u,p}$ and $e_{v,g} = \sum_p P_p'^T e_{v,p}$ 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 $e_{u,g}$ and $e_{v,g}$ allow a common threshold to be computed for all patches. The patches with high but homogeneous errors are thus discarded.

Finally, the combination of (5), (6) and (8) 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. The next section studies the optimization algorithm that will be used to solve (4).

E. Alternate Minimization

After learning the motion dictionaries offline, the cardiac motion estimation problem (4) can be solved using an alternate

minimization strategy as in [21] and [37]. It is based on an iterative approach, where the optimization alternates with respect to the motion U and the sparse codes α for fixed regularization parameters λ_s and λ_d, before increasing the sparsity parameter λ_d and repeating the process. In this work, this strategy allows us to incorporate an iterative re-weighted minimization of (4), 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 II). 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 below.

1) *Sparse coding and weight estimation:* The motion vectors in U are fixed and the optimization is performed with respect to α. The horizontal sparse vectors at the current iteration are determined by solving

$$\min_{\alpha_u} \sum_p \|\boldsymbol{P'}_p \boldsymbol{u} - \boldsymbol{D_u}\boldsymbol{\alpha}_{u,p}\|_2^2 \quad \text{s.t., } \|\boldsymbol{\alpha}_{u,p}\|_0 \leq K, \forall p$$

This sparse coding problem is NP-hard and is solved using the orthogonal matching pursuit (OMP) algorithm [38]. A similar problem is solved to find the vertical sparse codes α_v. The sparse regularization weights W_p are then updated for each patch as explained in Section III-D.

2) *Motion field, data weight and spatial weight estimation:* Once the sparse codes α have been determined, the motion field U is updated by solving the following minimization problem

$$\min_{U} E_Q(U, I) + \lambda_d E_W(U, \alpha) + \lambda_s E_S(U) \quad (9)$$

where the matrices Q and S have been determined at the previous iteration. The minimization problem (9) is solved using the scaled conjugate gradient algorithm (SCG) [39]. The data and spatial weights Q and S are then computed for the next iteration using the obtained motions U (see Sections III-B and III-C).

A full description of the sparse coding and motion estimation steps with an iterative weighting is provided in Algorithm 1.

IV. EXPERIMENTAL RESULTS

This section evaluates the proposed 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 is evaluated using the endpoint error described in [18]. For each pixel i, this error is defined as $\epsilon(i) = \sqrt{[\boldsymbol{u}(i) - \hat{\boldsymbol{u}}(i)]^2 + [\boldsymbol{v}(i) - \hat{\boldsymbol{v}}(i)]^2}$, where $\boldsymbol{u}(i)$, $\boldsymbol{v}(i)$ and $\hat{\boldsymbol{u}}(i)$, $\hat{\boldsymbol{v}}(i)$ are the true and estimated horizontal and vertical flow fields at pixel i. The cumulative displacement error ϵ_c is also computed, providing a global performance measure for the considered time instants. Finally, a paired t-test [40] is used to evaluate the statistical significance of the errors obtained with the proposed method and two other algorithms.

Algorithm 1: Robust Motion Field Estimation

Input : $I, D, K, \lambda_s, \lambda_d$, OuterSteps, InnerSteps

Initialization: $U = 0$, $Q = 1$, $W_p = 1$, $S = 1$

1 **for** $k = 1$ **to** *OuterSteps* **do**
2   **for** $j = 1$ **to** *InnerSteps* **do**
    %Sparse coding
3     $\alpha \leftarrow \text{OMP}(U, D, K)$;
    %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 $\boldsymbol{e}_{u,g} \leftarrow \sum_p \boldsymbol{P'}_p^T \boldsymbol{e}_{u,p}$, $\boldsymbol{e}_{v,g} \leftarrow \sum_p \boldsymbol{P'}_p^T \boldsymbol{e}_{v,p}$
    • Compute the scales $\sigma_u \leftarrow \sigma_0 \text{MAD}(\boldsymbol{e}_{u,g})$, $\sigma_v \leftarrow \sigma_0 \text{MAD}(\boldsymbol{e}_{v,g})$
    • Update the weights W_p for $p = 1, \ldots, N_p$
    %Motion estimation
4     $U \leftarrow \min_{U} \|\boldsymbol{Q}^{1/2}(\partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T U)\|_2^2 + \lambda_s \|\boldsymbol{S}^{1/2}\nabla U\|_2^2 + \lambda_d \sum_p \|\boldsymbol{W}_p^{1/2}(\boldsymbol{P}_p U - \boldsymbol{D}\boldsymbol{\alpha}_p)\|_2^2$;
    %Data and spatial weight update
    • Compute the residuals $\boldsymbol{e}_d \leftarrow (\partial_t \boldsymbol{I} + \nabla \boldsymbol{I}^T U)$ and $\boldsymbol{e}_s \leftarrow \nabla U$
    • Compute the scales $\sigma_d \leftarrow \sigma_0 \text{MAD}(\boldsymbol{e}_d)$ and $\sigma_s \leftarrow \sigma_0 \text{MAD}(\boldsymbol{e}_s)$
    • Update the weights Q and S
5   **end**
    • Adjust λ_d
6 **end**
Output: Motion U, weights Q, W_p, S and sparse codes α.

This test requires to define a significance level, which was set to $\alpha = 0.05$ and an asterisk (*) is used to indicate the statistical significance of the reported values in comparison with the other two methods.

A. Dictionary Learning Parameters

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 [41].[2] The dictionaries were learnt for patches of size 16×16, resulting in a dictionary size of 256×384, where 384 is the number of dictionary atoms. Note that the selected patch size allowed a compromise between the number of patches and the size of the dictionary. The sparsity parameter K representing the maximum number of non-zero coefficients used to represent a patch was fixed to $K = 5$ using cross-validation. This parameter is usually much smaller than the number of training atoms and is related to the noise level [42]. Note finally that the dictionaries were learnt separately for the horizontal and vertical directions and fixed before motion estimation, *i.e.,* learnt offline.

---

[2]More details about the data generation process can be found at https://team.inria.fr/ asclepios/data/straus/.
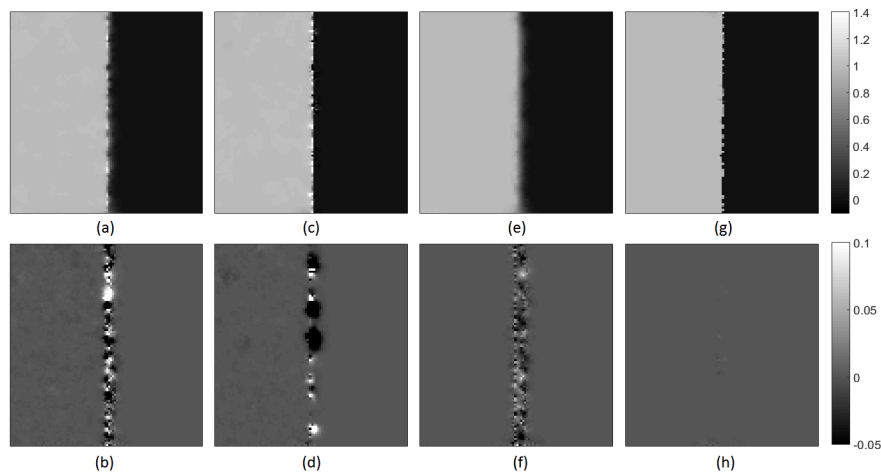
Fig. 2. Estimated horizontal (top) and vertical (bottom) motion fields (in pixels) using the non-robust (a,b), robust spatial regularization only (c,d), robust data fidelity only (e,f) and fully robust (g,h) formulations.

B. 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 [43]. In this section, we use the Tukey Biweight function (2). This choice was motivated by the fact that the synthetic images present abrupt transitions with clear boundaries (see Section II). 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 $\lambda_s = 0.1$ for the spatial regularization parameter.

1) Robust Data Fidelity and Spatial Regularization: This section first considers the example of a pair of images with a simple translation as in [26]. 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 $\lambda_d = 0$. The motion field was then estimated using a non-robust formulation of (4) for which the weights are not updated (*i.e.*, $q(i) = s(i) = 1, \forall i$). The weights were then updated separately for the spatial and data fidelity terms and finally the motion was estimated using the fully robust formulation (4). Fig. 2 shows the resulting horizontal (top row) and vertical (bottom row) motion fields. The non-robust formulation associated with Fig. 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. 2 (c,d) is noisy. The use of a robust data fidelity term allows this noise to

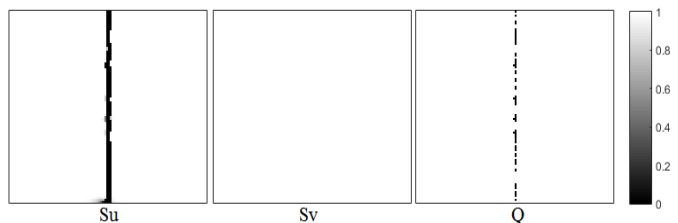| Method	Non-robust	Robust spatial	Robust data	Fully robust
Error	0.020±0.10	0.018±0.14	0.023±0.10	**0.004±0.05**

Fig. 3. Estimated weights S and Q for the images subjected to translations.

be mitigated by relaxing the brightness constancy assumption. The resulting motion displayed in Fig. 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. 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 I. The fully robust formulation provides competitive results in terms of mean and standard deviation (std) of the endpoint error. The final weights Q and S associated with the fully robust formulation are shown in Fig. 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 S_v do not outline any spatial outliers in Fig. 3 (middle).

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
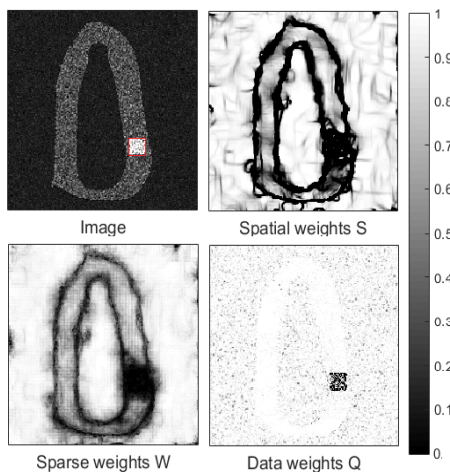
Fig. 4. Synthetic cardiac image and its sparse weights W, spatial weights S and data weights Q. The colorbar indicates the weight values.

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

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 [41] (see Section IV-C). 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).

Fig. 4 shows the estimated weights for the two images. For brevity, the horizontal and vertical sparse weights were merged such that $W = \text{diag}^{-1}(W_u W_v)$. The spatial weights S_u and 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 III, 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 II compares the estimation accuracy obtained with robust and non-robust sparse regularizations. These results are in favour of the fully robust formulation for these images.

C. Realistic Simulations

This section evaluates the proposed method using realistic simulated US images with ground-truth motions [41]. The simulation sequences span an entire cardiac cycle of 34 frames with a frame rate of 22 Hz. The original 3D images were of size $224 \times 176 \times 208$, with a voxel size of $0.7 \times 0.9 \times 0.6$ mm^3. The middle slice of the considered

imaging plane was extracted for each sequence in order to create 2D images. The considered sequences include one healthy sequence (*i.e.,* Normal), two ischemic cases with occlusions of the proximal and distal parts of the left anterior descending coronary artery (*i.e.*, LADprox and LADdist) and one case 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. 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 IV-C.2). In a second step, synthetic artefacts were introduced in order to corrupt the LADprox sequence (see Section IV-C.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) [26]. The BA method uses a robust OF-based data fidelity term with a robust smoothness constraint based on the gradient of the flow. The second non-robust method (referred to as NR for non-robust) was studied in [21] and is based on an energy minimization framework. Its energy is defined using 1) a data fidelity term based on the assumption of multiplicative Rayleigh noise, and 2) a sparse prior based on dictionary learning.

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 sequence and $\lambda_s = 0.2$ for all other sequences. For each outer iteration, the sparse regularization parameter λ_d was logarithmically increased from 10^{-4} to 10 (see Section III-E) in 4 iterations. 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 IV-B. 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 III. 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

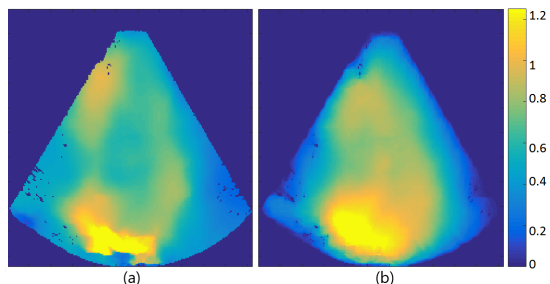| Sequence	Method		
	NR	BA	Proposed
LADprox	0.1410 ± 0.112 (4.48)	0.175 ± 0.128 (5.57)	**0.101 ± 0.089 (3.22)***
LADprox SAX	0.222 ± 0.147 (2.87)	0.164 ± **0.120** (2.12)	**0.163** ± 0.124 (**2.11**)
Normal	0.240 ± **0.161** (7.38)	0.363 ± 0.269 (11.14)	**0.237** ± 0.167 (**7.3**)*
LCX	0.150 ± 0.151 (3.65)	0.169 ± 0.142 (4.11)	**0.125 ± 0.128 (3.05)***

Fig. 5. Estimated motions (in pixels) for the 4th frame of the LADprox sequence for (a) the proposed robust method and (b) the NR method of [21].

Fig. 6. First and second frames of the *Corrupted1* simulation sequence. The red boxes indicate the regions with attenuation and clutter artefacts.

Sequence	Method		
	NR	BA	Proposed
Corrupted1	0.196 ± 0.147 (6)	0.176 ± 0.124 (5.34)	**0.117 ± 0.116 (3.6)***
Corrupted2	0.222 ± 0.190 (6.8)	0.176 ± **0.121** (5.4)	**0.122** ± 0.132 (**3.73**)*

sequence and $\lambda_{NR} = 0.5$ for the data with artefacts. The dictionary learning and sparse regularization parameters were adjusted as in [21].

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 III. This table shows that the proposed method provided smaller mean errors for the considered sequences. tIn the case of the LADprox SAX sequence, the statistical significance of the differences between the errors of the proposed and BA methods was not found and the corresponding p-value was 0.497, whereas this p-value was less than 0.001 for the other sequences. Fig. 7 (a,c) shows the mean errors for each frame of the cardiac cycle for the LADprox sequence. An improved performance can be observed, specifically for the large displacements in the beginning of the sequences (systole). These results show that robustness can also be beneficial for motion estimation accuracy in the absence of UI artefacts. In order to show the impact of the proposed robust approach on motion estimates, Fig. 5 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. 7 for the corresponding error maps).

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 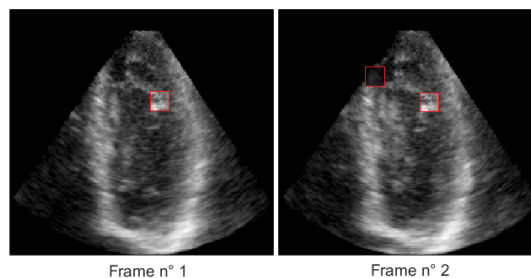to simulate moving reverberations, reflections or brighter speckles. Two sequences *Corrupted1* and *Corrupted2* were created with different attenuation and reflection magnitudes. For each pair of consecutive images, only one frame was attenuated in a region of size 15×15 pixels, with attenuations of 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. 6 shows the first two frames of the *Corrupted1* sequence.

Table IV summarizes the results obtained for the two corrupted sequences in terms of the global endpoint and the cumulative displacement errors. Note that the errors were computed for the uncorrupted regions only[3]. The results show that the proposed method provides a competitive performance in terms of the error means and stds, as well as the cumulative errors computed considering the entire sequence. 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. 7 shows the time evolution of the mean endpoint errors for the entire cardiac cycle. Note that 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. 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.

In order to understand the impact of robustness, Fig. 8 shows the error maps for one frame (peak-systole) 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

[3]The proposed robust and BA methods may result in different outliers. However, in this paper, 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.
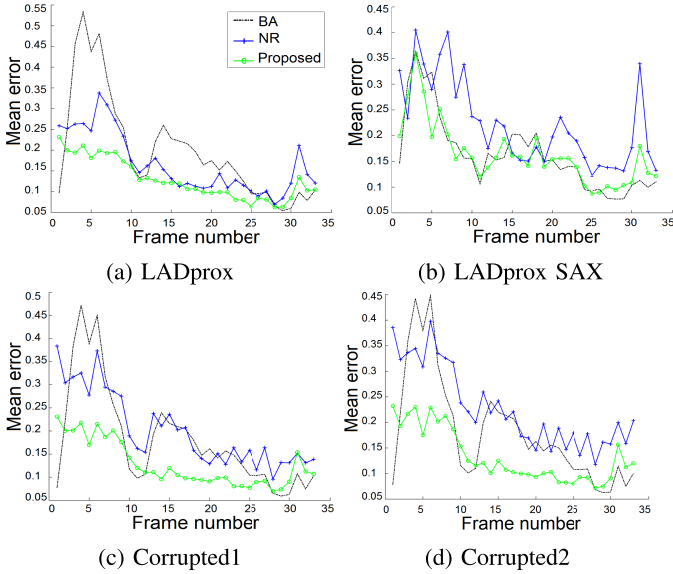
Fig. 7. Mean endpoint error for (a,b) the uncorrupted LADprox sequence and (c,d) the two corrupted sequences.
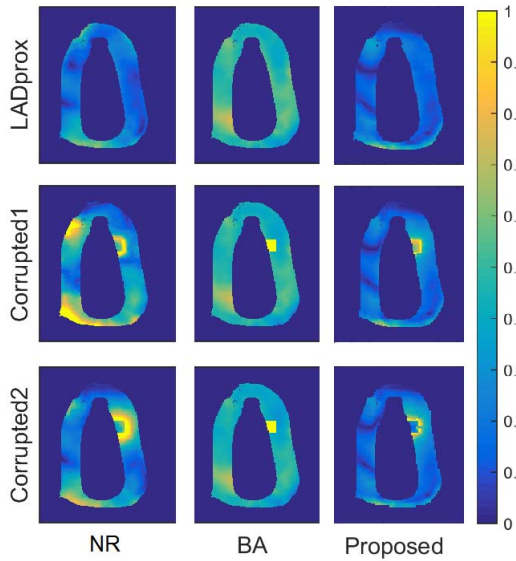

Fig. 8. Error maps (in pixels) of the 4th frame (maximum displacement) for the 3 simulated sequences.

for the clutter artefact and data errors for the attenuation. Also, large errors can be observed in some uncorrupted portions of the 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.

D. In Vivo

This section evaluates the proposed robust motion estimation method using two sequences of real US cardiac
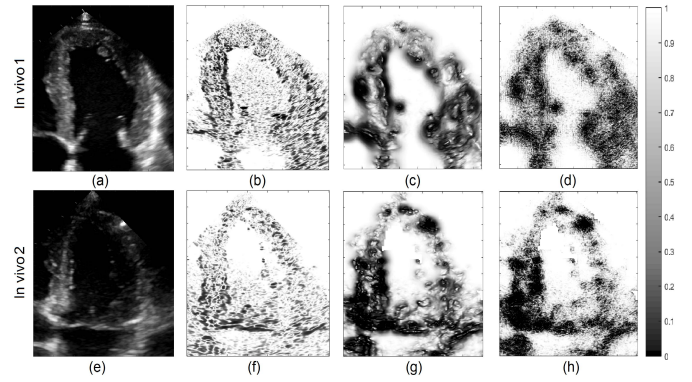

Fig. 9. Weights for two examples of *in vivo* images obtained for two different patients. Images are displayed in (a) and (e) whereas the data, spatial and sparse weights are shown in (b)(c)(d) and (f)(g)(h).

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 a tsector M5Sc-D 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 disease (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 twith motion dictionaries of size 400×600 learned using the LADdist sequence (see Section IV-A). The spatial regularization parameters were tuned to give the best visual tracking results (see Fig. 12), leading to $\lambda_s = \lambda_{\mathrm{NR}} = 0.2$ for the proposed approach and the NR method and $\lambda_{\mathrm{BA}} = 1$, while the other parameters were the same as in Section IV-C.

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. 9 shows two images and the corresponding data, spatial and sparse weights using the Lorentzian function. 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. 9 (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. 9 is provided in Fig. 10. 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.
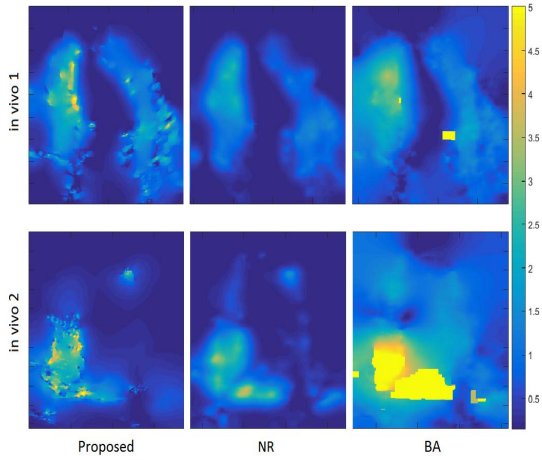
Fig. 10. Displacement maps (in pixels) obtained for the proposed, NR and BA methods for the *in vivo* images in Fig. 9.

(a) Horizontal displacements   (b) Vertical displacements

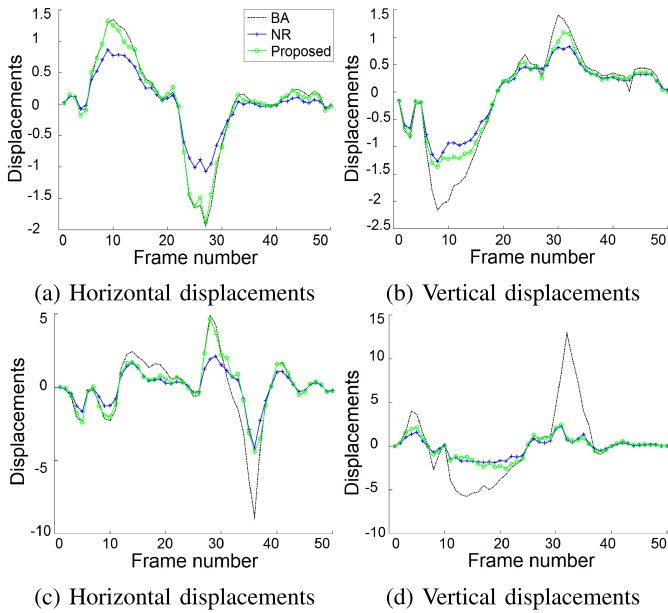(c) Horizontal displacements   (d) Vertical displacements

Fig. 11. Mean horizontal and vertical displacements (in pixels) in the basal segment of the septal wall for (a,b) the *in vivo 1* and (c,d) the *in vivo 2* sequences.

The time evolutions of the mean horizontal and vertical displacements in the basal segment of the septal wall are shown in Fig. 11 for the considered algorithms. This figure shows that for both sequences the horizontal displacements have positive values in the systole phase while these displacements are negative in the diastole phase. Overall, the displacements associated with the diseased patient (*i.e., in vivo 1*) have smaller magnitudes in comparison with the healthy case (*i.e., in vivo 2*).

Fig. 12 shows the displacements obtained for 6 landmarks in comparison with a manual tracking (red circles). The landmarks were located on the endocardium 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 manual tracking of the 10 consecutive frames was also used to compute the corresponding motion

estimation errors provided in Fig. 12 and Table V. This table shows that the smallest error means and stds were obtained for both sequences using the proposed method. 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. In contrast, the motions obtained using the proposed and BA methods were closer to the manual tracking, with smaller errors for the proposed method.

V. DISCUSSION AND PERSPECTIVES

This paper presented a new motion estimation method for robust 2D cardiac US images. The main objective of this method was to robustify the cardiac motion estimation algorithm of [21] (based on spatial and sparse regularizations) in order to mitigate the effects of outliers. 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. 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 I). For example, in [32] 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 [32], 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.

At this point, we would like to mention some issues that would deserved to be considered in future work. The proposed method relies strongly on the tuning of several parameters, *i.e.,* the regularization, dictionary learning and robust weight parameters. The scale parameter σ was estimated jointly with the motions using (3). However, the parameter c still needs to be fixed *a priori*. We think that methods based on the Stein's unbiased risk estimate (SURE) [44] or Bayesian inference [45] would deserve to be investigated for the estimation of these parameters. Finally, the average execution time of the proposed method for a pair of realistic simulation images was 87.32s (compared to 52.33s for the NR method and 11.31s for BA), which may be high for some applications. As mentioned in [21], a more efficient implementation can be achieved

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

(b) Frame 16 (*in vivo 2*)   Frame 20: $\epsilon = 3.41$, $\epsilon_{NR} = 7.77$,   Frame 22: $\epsilon = 4.95$, $\epsilon_{NR} = 10.48$,   Frame 25: $\epsilon = 5.17$, $\epsilon_{NR} = 12.28$, $\epsilon_{BA} = 9.38$ $\epsilon_{BA} = 12.53$ $\epsilon_{BA} = 14.43$
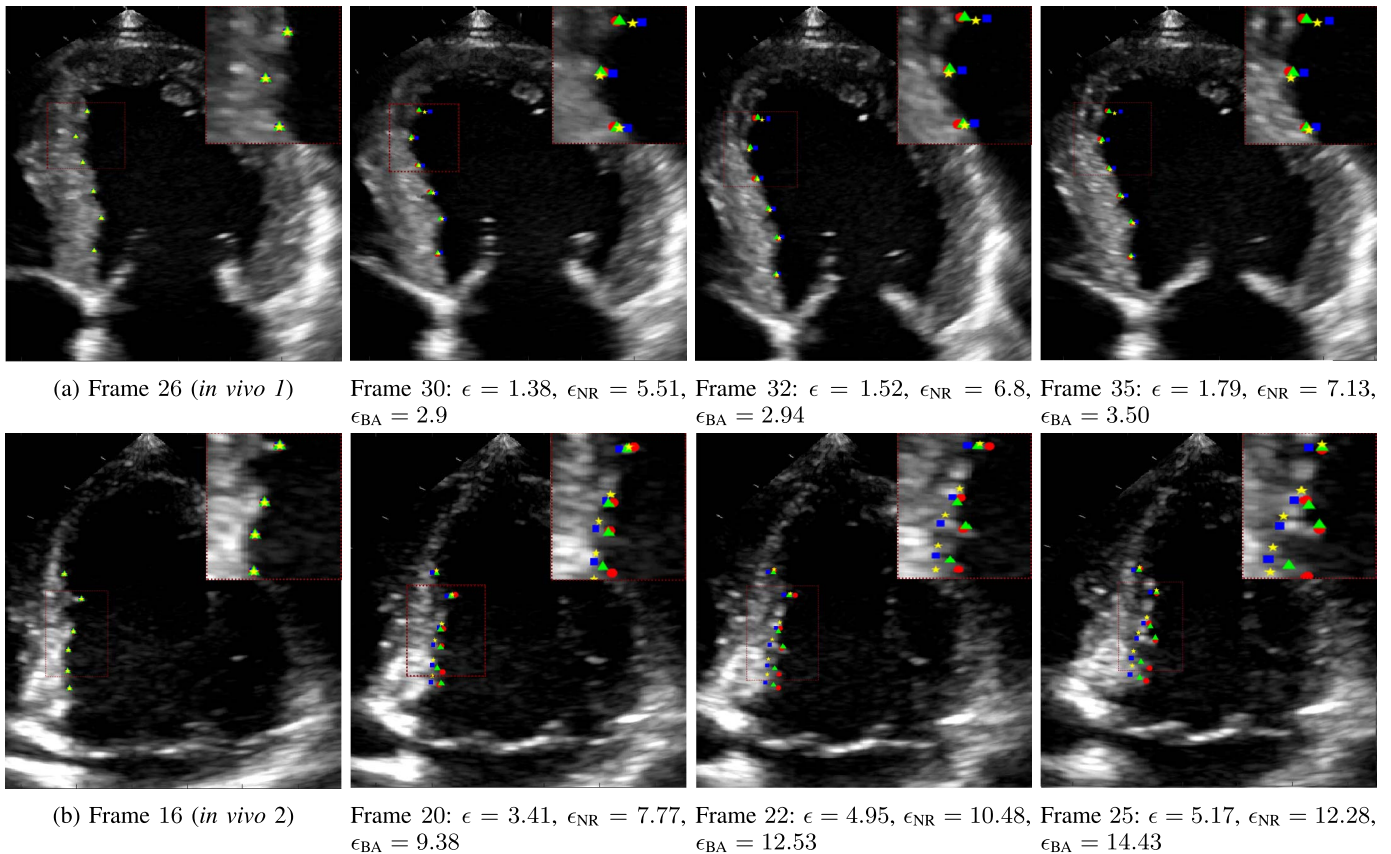
Fig. 12. Tracking results for 6 landmarks on the endocardial wall (magnified in the top-right part of the images) for (a) the *in vivo 1* and (b) *in vivo 2* sequences. The average errors for each frame in comparison with the manual tracking (red circles) are provided for the proposed (green triangles), NR (blue squares) and BA (yellow stars) methods.

using parallel computing for the horizontal and vertical sparse coding and sparse regularization terms.

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. For example, this is the case when the training set contains patterns far from typical cardiac motions. If the dictionary is learnt in an adaptive way, *i.e.,* using the estimated motions themselves, a robust learning approach would allow us to discard erroneous motion estimates. Furthermore, it would be worth to take advantage of the sparse codes, the dictionary atoms and the robust weights that are obtained using the proposed method. For example, a joint robust motion estimation and segmentation could be obtained by combining the information provided by the weights of the spatial and sparse regularizations. Taking into account the increased use of 3D UI, it is also worth mentioning that the proposed method could be extended to 3D. However, the limitations of frame rate and spatial resolution in the azimuthal direction imply that the use of 3D US images does not lead necessarily to a more accurate estimation when compared to 2D UI.

VI. CONCLUSIONS

This paper introduced a robust cardiac motion estimation method for 2D ultrasound images. The proposed approach exploited the sparse properties of motion when decom-posed on a dedicated dictionary, while being robust to outliers. The motion estimation problem was formulated as a weighted energy minimization in an optical flow framework with combined spatial and sparse regularizations. Robustness was introduced using weight functions derived from robust M-estimators. In order to ensure a fully robust estimation, the weighting was applied jointly to the data fidelity term and the spatial and sparsity regularizations. 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.

REFERENCES

[1] 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.

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

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

[4] D. Boukerroui, J. A. Noble, and M. Brady, *Velocity Estimation in Ultrasound Images: A Block Matching Approach*. Berlin, Germany: Springer, 2003, pp. 586–598.

[5] 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., Ferroelectr., Freq. Control*, vol. 48, no. 4, pp. 1111–1123, Jul. 2001.

[6] 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.

[7] 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.

[8] 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.

[9] 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. Funct. Imag. Modeling Heart (FIMH)*, Nice, France, vol. 5528, Jun. 2009, pp. 427–436.

[10] 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, pp. 89–97, Jan. 2016.

[11] M. D. Craene *et al.*, "Temporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography," *Med. Image Anal.*, vol. 16, no. 2, pp. 427–450, 2012.

[12] 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, pp. 2264–2277, Aug. 2014.

[13] 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.

[14] M. Alessandrini, H. Liebgott, D. Barbosa, and O. Bernard, "Monogenic phase based optical flow computation for myocardial motion analysis in 3D echocardiography," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart (STACOM)*, Nice, France, vol. 7746, Oct. 2013, pp. 159–168.

[15] A. Elen *et al.*, "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.

[16] K. McLeod, A. Prakosa, T. Mansi, M. Sermesant, and X. Pennec, "An incompressible log-domain demons algorithm for tracking heart tissue," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, O. Camara, E. Konukoglu, M. Pop, K. Rhode, M. Sermesant, and A. Young, Eds. Berlin, Germany: Springer, 2012, pp. 55–67.

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

[18] 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.

[19] 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.

[20] 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.

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

[22] N. Ouzir, J.-Y. Tourneret, and A. Basarab, "Optical flow estimation in ultrasound images using a sparse representation," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Curacao, The Netherlands, Dec. 2017, pp. 1–5.

[23] J. S. Turek, M. Elad, and I. Yavneh, "Clutter mitigation in echocardiography using sparse signal separation," *Int. J. Biomed. Imag.*, vol. 2015, Jan. 2015, Art. no. 958963.

[24] A. Perperidis, "Postprocessing approaches for the improvement of cardiac ultrasound B-mode images: A review," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 3, pp. 470–485, Mar. 2016.

[25] 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, pp. 204–216, Feb. 2002.

[26] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comput. Vis. Image Understand.*, vol. 63, no. 1, pp. 75–104, 1996.

[27] P. Héas, C. Herzet, and E. Mémin, "Bayesian inference of models and hyperparameters for robust optical-flow estimation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1437–1451, Apr. 2012.

[28] J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *J. Vis. Commun. Image Represent.*, vol. 6, no. 4, pp. 348–365, Dec. 1995.

[29] E. P. Ong and M. Spann, "Robust optical flow computation based on least-median-of-squares regression," *Int. J. Comput. Vis.*, vol. 31, no. 1, pp. 51–82, Feb. 1999.

[30] Y.-H. 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 Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1418–1435, Sep. 2006.

[31] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Robust registration of dynamic facial sequences," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1708–1722, Apr. 2017.

[32] 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, pp. 1562–1575, Jul. 2015.

[33] I. Pratikakis, C. Barillot, and P. Hellier, *Robust Multi-Scale Non-Rigid Registration of 3D Ultrasound Images*. Vancouver, BC, Canada: Springer, Jul. 2001, pp. 389–397.

[34] A. I. Aviles, T. Widlak, A. Casals, M. M. Nillesen, and H. Ammari, "Robust cardiac motion estimation using ultrafast ultrasound data: A low-rank topology-preserving approach," *Phys. Med. Biol.*, vol. 62, no. 12, pp. 4831–4851, 2017.

[35] 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, pp. 928–945, Apr. 2011.

[36] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "Robust regression methods for computer vision: A review," *Int. J. Comput. Vis.*, vol. 6, no. 1, pp. 59–70, 1991.

[37] J. Sulam and M. Elad, "Expected patch log likelihood with a sparse prior," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Cham, Switzerland: Springer, 2015, pp. 99–111.

[38] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.

[39] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, Nov. 1993.

[40] B. L. Welch, "The generalization of 'student's' problem when several different population variances are involved," *Biometrika*, vol. 34, nos. 1–2, pp. 28–35, 1947.

[41] 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. 7, pp. 1436–1451, Jul. 2015.

[42] 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.

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

[44] 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.

[45] 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.