



HAL
open science

Robust cardiac motion estimation with dictionary learning and temporal regularization for ultrasound imaging

Nora Leïla Ouzir, José Bioucas-Dias, Adrian Basarab, Jean-Yves Tournet

► **To cite this version:**

Nora Leïla Ouzir, José Bioucas-Dias, Adrian Basarab, Jean-Yves Tournet. Robust cardiac motion estimation with dictionary learning and temporal regularization for ultrasound imaging. IEEE International Ultrasonics Symposium (IUS 2019), Oct 2019, Glasgow, United Kingdom. pp.2326-2329, 10.1109/ULTSYM.2019.8925936 . hal-02901597

HAL Id: hal-02901597

<https://hal.science/hal-02901597v1>

Submitted on 17 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:

<http://oatao.univ-toulouse.fr/26244>

Official URL

<https://doi.org/10.1109/ULTSYM.2019.8925936>

To cite this version: Ouzir, Nora Leïla and Bioucas-Dias, José and Basarab, Adrian and Tourneret, Jean-Yves *Robust cardiac motion estimation with dictionary learning and temporal regularization for ultrasound imaging*. (2019) In: IEEE International Ultrasonics Symposium (IUS 2019), 6 October 2019 - 9 October 2019 (Glasgow, United Kingdom).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Robust Cardiac Motion Estimation With Dictionary Learning and Temporal Regularization for Ultrasound Imaging

N. Ouzir

Department of Signal
Processing and Acoustics
Aalto University, Finland
Email: nora.ouzir@aalto.fi

J. Biucas-Dias

Instituto de Telecomunicações
Instituto Superior Técnico
Universidade de Lisboa, Portugal
Email: biucas@lx.it.pt

A. Basarab¹ and J.-Y. Tournet²

University of Toulouse
IRIT UMR CNRS 5505
Université Paul Sabatier, ENSEEIHT, France
Email: ¹basarab@irit.fr,

²Jean-Yves.Tournet@enseeiht.fr

Abstract—Estimating the cardiac motion from ultrasound (US) images is an ill-posed problem that requires regularization. In a recent study, it was shown that constraining the cardiac motion fields to be patch-wise sparse in a learnt overcomplete motion dictionary is more accurate than local parametric models (affine) or global functions (B-splines, total variation). In this work, we extend this method by incorporating temporal smoothness in a multi-frame optical-flow (OF) strategy. An efficient optimization strategy using the constrained split augmented Lagrangian shrinkage algorithm (C-SALSA) is proposed. The performance is evaluated on a realistic simulated cardiac dataset with available ground-truth. A comparison with the pairwise approach shows the interest of the proposed temporal regularization and multi-frame strategy in terms of accuracy and computational time.

Index Terms—Optical flow, multi-frame motion estimation, temporal regularization, ultrasound, echocardiography.

I. INTRODUCTION

Cardiac motion estimation from US images is a well-established tool for the diagnosis of cardiovascular diseases [1]. This is mainly due to the practical advantages of US imaging, such as low cost and safety, but also to its relatively high frame rates that allow the fast motions of the tissues to be captured. The most classical cardiac motion estimation techniques fall into the category of speckle tracking methods [2]. These methods seek to match blocks between images using a similarity measure, *e.g.*, the sum-of-squared-differences [3]. Other widely used approaches are based on parametric models, such as affine transformations [4], or non-rigid deformation models, typically B-splines parametrizations [5]. Differential OF methods have also been successfully used in the context of echocardiography [6]. The basic assumption of OF, is that the intensities are constant across consecutive frames. The flow is then estimated using the spatial and temporal image intensity variations.

Motion estimation is an ill-posed problem that requires regularization. One of the most classical regularization strategies enforces spatial smoothness, usually involving constraints on the gradient of the motion field [7]. Parametric models also enable regularization by restricting the number of possible dis-

placements [8]. Recently, sparsity-based regularizations have been used successfully in various signal and image processing problems. The key idea is that an image patch can be sparsely represented in an appropriate learnt dictionary. In the context of echocardiography, a sparsity-based regularization has been successfully combined with spatial smoothness constraints to regularize the cardiac motion estimation problem [9], [10]. In these works, a dictionary of typical cardiac displacement patterns was learnt using patches of simulated realistic cardiac motions.

Most of the before-mentioned techniques use only pairs of consecutive images. One drawback of these *pairwise* approaches is that they fail to exploit the temporal information embedded in the whole image sequence. Nonetheless, the temporal aspect plays an essential part in motion analysis. This is especially true in the case of cyclic displacement patterns, such as cardiac motions. In the context of echocardiography, some works have sought to incorporate this temporal aspect, typically, by extending the B-splines basis into the time domain [5], [11]. This also allows temporal smoothness to be enforced, which is useful because of the presence of speckle decorrelation due to, *e.g.*, out-of-plane motions or artefacts. However, these methods still rely on the choice of a single reference frame, resulting in a decrease of accuracy for farther frames in the sequence.

In this work, a new multi-frame OF method for cardiac motion estimation is introduced. In Section II-A, the temporal aspect is incorporated by defining a cost function that allows the motions of an entire image sequence to be estimated simultaneously. One advantage of the proposed approach is that it does not depend on the choice of a single reference frame. Motivated by the success of the sparsity-based regularization studied in [9], the motions are constrained to be sparse in an appropriate cardiac motion dictionary. In addition, in Section II-B, we incorporate *a priori* knowledge about the temporal evolution of motions by assuming piecewise smooth trajectories. A robust weighting technique is employed to relax this constraint for large temporal shifts. In Section II-C, the resulting problem is solved using an efficient optimization

strategy based on C-SALSA. Finally, experiments using simulated data and concluding remarks are reported in Sections III and IV.

II. MULTI-FRAME MOTION ESTIMATION

A. Proposed cost function

Let $\mathbf{I} \in \mathbb{R}^{NM}$ denote a sequence of 2D cardiac images, with N the number of pixels and M the total number of frames. The 2D motion fields $\mathbf{X}_t \in \mathbb{R}^{2N}$ to be estimated between consecutive frames \mathbf{I}_t and \mathbf{I}_{t+1} are concatenated in a vector $\mathbf{X} \in \mathbb{R}^{2MN}$. The proposed motion estimation method is formulated as an energy minimization with an OF-based data fidelity term E_{OF} . The proposed regularization strategy makes use of three different constraints: (i) a temporal regularization term denoted as E_{T} is introduced in order to exploit the temporal smoothness of the motions, (ii) a total-variation regularization E_{TV} is used to ensure spatially smooth motion fields, and (iii) a sparse regularization term E_{P} exploits the patch-wise sparsity of \mathbf{X} when decomposed on a learnt dictionary of cardiac motions $\mathbf{D} \in \mathbb{R}^{2n \times 2q}$ [9], with $2n$ denoting the patch size and $2q$ the number of atoms in the dictionary. The motion fields are finally obtained through the minimization of the resulting energy function as follows

$$\min_{\Omega, \mathbf{X}} E_{\text{OF}}(\mathbf{I}, \mathbf{X}) + \lambda_{\text{TV}} E_{\text{TV}}(\mathbf{X}) + \lambda_{\text{P}} E_{\text{P}}(\mathbf{X}, \Omega) + \lambda_{\text{T}} E_{\text{T}}(\mathbf{X}) + \lambda_1 \|\Omega\|_1 \quad (1)$$

where $\Omega \in \mathbb{R}^{2q \times p}$ contains the sparse codes associated with all the patches in the sequence. Note that in (1), the sparsity of the vectors in Ω is enforced using the ℓ_1 -norm. The regularization parameters allowing the influence of each energy term to be controlled are denoted as λ_{TV} , λ_{P} , λ_{T} and λ_1 .

As in [9], the dictionary \mathbf{D} is learnt beforehand from a set of training cardiac motion fields (see Section III-A). Following [9], the online dictionary learning method [12] is used for learning the dictionary. The motion fields are then estimated by minimizing (1). The data fidelity term and the proposed regularizations are detailed in the following.

1) *OF-based data fidelity*: The data fidelity term is based on the OF method [7]. In this work, we propose a multi-frame formulation using all the images in the sequence simultaneously, *i.e.*,

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

where $\|\cdot\|_F$ denotes the Frobenius norm and $\mathbf{Y} \in \mathbb{R}^{NM}$ is a vector containing the temporal derivatives of the intensities $\partial_t \mathbf{I}_t \in \mathbb{R}^N$ with $t = 1, \dots, M$, $\mathbf{A} \in \mathbb{R}^{MN \times 2MN}$ is a block diagonal matrix whose blocks are $\nabla \mathbf{I}_t^T \in \mathbb{R}^{N \times 2N}$ where ∇ indicates the spatial gradient operator. Note that the term (2) incorporates the OF data fidelity terms of all consecutive frames and thus does not require the choice of a single reference frame.

2) *Sparse and spatial regularizations*: The terms E_{TV} and E_{P} enforce spatial smoothness and sparsity for all the motions

in \mathbf{X} simultaneously. The spatial regularization is based on a classical total-variation defined using the following form

$$E_{\text{TV}}(\mathbf{X}) = \|\nabla_h \mathbf{X}\|_F^2 + \|\nabla_v \mathbf{X}\|_F^2 \quad (3)$$

where ∇_h and ∇_v indicate the horizontal and vertical gradient components. The pairwise sparse regularization used in [9], is also reformulated for the entire sequence, *i.e.*,

$$E_{\text{P}}(\mathbf{X}, \Omega) = \|\mathcal{P}(\mathbf{X}) - \mathbf{D}\Omega\|_F^2 \quad (4)$$

where $\mathcal{P}(\cdot) : \mathbb{R}^{MN} \mapsto \mathbb{R}^{2n \times p}$ is an operator that extracts p overlapping patches of size $2n$ from \mathbf{X} .

B. Temporal regularization

The aim of the temporal regularization term E_{T} is to enforce piecewise smooth trajectories for the pixels inside the myocardium. This is achieved by constraining the displacement of each pixel at a time instant t to be close to the displacement of the corresponding pixel in adjacent frames, while allowing some temporal discontinuities (due, for example, to frame rate limitations). In this work, we propose to enforce temporal smoothness using the previous and forward frames simultaneously. This choice allows us to avoid biased estimates towards the end (or beginning) of the sequence.

Let \mathbf{X}_b and \mathbf{X}_f be the time shifted versions of \mathbf{X} containing all the backward and forward motions in the sequence. One way of enforcing smooth trajectories, consists in imposing a constant velocity. More precisely, one can use a finite difference approximation of the second derivatives of the motions [13], leading to

$$E_{\text{T}}(\mathbf{X}) = \sum_i [2\mathbf{X}(i') - \mathbf{X}_b(i) - \mathbf{X}_f(i'')]^2 \quad (5)$$

where the indexes i' and i'' correspond to the positions of the pixel i in the frames t and $t+1$, *i.e.*, $i' = i + \mathbf{X}_b(i)$ and $i'' = i' + \mathbf{X}(i')$. This change in the pixel positions causes a non linearity, which is bypassed by using an approximation based on the 1st order Taylor expansion of the term $\mathbf{X}_{i'}$. Furthermore, we propose to cope with temporal discontinuities by using a robust weighting approach. Specifically, a weight matrix $\mathbf{W} \in \mathbb{R}^{MN \times MN}$ is introduced in order to assign lower weights to the pixels with large shifts in the displacements. The temporal smoothness constraint is thus relaxed for large temporal discontinuities, *e.g.*, at end-systole. More specifically, the weights are computed for the current residual using the Lorentzian weight function, which is a redescending M-estimator that allows the influence of the discontinuities to be decreased to zero. After applying the Taylor expansion and introducing \mathbf{W} , the regularization term is reformulated as follows

$$E_{\text{T}}(\mathbf{X}) = \|\mathbf{W}^{1/2} [2\mathbf{X} - (\mathbf{X}_b - 2\mathbf{X}_b^T \mathbf{X}_b + \mathbf{X}_f)]\|_F^2 \quad (6)$$

where the pixel index i'' has now been dropped from \mathbf{X}_f . Note that \mathbf{W} is computed iteratively using the previous residual value, *i.e.*, $\mathbf{e} = 2\mathbf{X} - (\mathbf{X}_b - 2\mathbf{L}\mathbf{X}_b^T \mathbf{X}_b + \mathbf{X}_f)$, such that $\mathbf{W} = \text{diag}([w(e_1), \dots, w(e_{MN})])$ where w denotes the Lorentzian weight function.

C. Proposed optimization strategy

Problem (1) is solved using the C-SALSA algorithm [14]. C-SALSA is based on the alternating direction method of multipliers (ADMM), which iterates between simple optimization sub-problems. This approach provides a flexible framework that allows us to deal with the different regularization terms considered in this work. Furthermore, it allows us to cope with the high dimensionality of the unknown matrices \mathbf{X} and $\mathbf{\Omega}$, as well as the presence of the non-quadratic term $\|\mathbf{\Omega}\|_1$.

The sub-problems are formulated by introducing 7 auxiliary variables that stand for the motion fields in \mathbf{X} and the sparse codes $\mathbf{\Omega}$. The optimization problem (1) is reformulated as follows

$$\begin{aligned} \min_{\mathbf{V}_1-\mathbf{V}_5} \quad & \frac{1}{2}\|\mathbf{Y} + \mathbf{A}\mathbf{V}_1\|_F^2 + \lambda_{\text{TV}}(\|\nabla_h \mathbf{V}_{21}\|_F^2 + \|\nabla_v \mathbf{V}_{22}\|_F^2) \\ & + \lambda_P \|\mathcal{P}(\mathbf{V}_{31}) - \mathbf{D}\mathbf{V}_{32}\|_F^2 + \lambda \|\mathbf{V}_4\|_1 + \lambda_T E_T(\mathbf{V}_5) \end{aligned} \quad (7)$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{V}_1 = \mathbf{X} \\ & \mathbf{V}_{21} = \nabla_h \mathbf{X} \quad \mathbf{V}_{32} = \mathbf{\Omega} \\ & \mathbf{V}_{22} = \nabla_v \mathbf{X} \quad \text{and} \quad \mathbf{V}_4 = \mathbf{\Omega}. \\ & \mathbf{V}_{31} = \mathbf{X} \\ & \mathbf{V}_5 = \mathbf{X} \end{aligned} \quad (8)$$

The augmented Lagrangian associated with the problem (7) is defined as follows

$$\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{V}, \mathbf{G}) = E(\mathbf{V}) + \frac{\mu}{2} \|\mathbf{H}\tilde{\mathbf{X}} - \mathbf{V} - \mathbf{G}\|_F^2 \quad (9)$$

where the auxiliary variables are concatenated in \mathbf{V} , $\tilde{\mathbf{X}}$ contains \mathbf{X} and $\mathbf{\Omega}$, the matrix \mathbf{H} stands for the corresponding identity or gradient operators, \mathbf{G} contains the Lagrange multipliers corresponding to \mathbf{V} , $\mu > 0$ is the penalty parameter, and $E(\mathbf{V})$ is the cost function associated with the problem (7). Finally, the minimization is carried out by iterating between optimizations with respect to each variable (or auxiliary variable) and updating the Lagrange multipliers in \mathbf{G} . The different iterations of the C-SALSA scheme are summarized in Algorithm 1.

Algorithm 1: Motion estimation using C-SALSA

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

```

1 for  $k = 1, \dots, k_{\text{max}}$  do
  %Motions and sparse codes
2    $\tilde{\mathbf{X}}^k \in \underset{\tilde{\mathbf{X}}}{\text{argmin}} \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{V}^{k-1}, \mathbf{G}^{k-1});$ 
  %Proximal computations
3    $\mathbf{V}^k \in \underset{\mathbf{V}}{\text{argmin}} \mathcal{L}(\tilde{\mathbf{X}}^k, \mathbf{V}, \mathbf{G}^{k-1});$ 
  %Lagrange multipliers
4    $\mathbf{G}^k = \mathbf{G}^{k-1} - (\mathbf{H}\tilde{\mathbf{X}}^k - \mathbf{V}^k);$ 
5 end
```

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

III. EXPERIMENTS

In this section, the proposed multi-frame approach is tested using a realistic simulated cardiac US sequence. We focus on the comparison of the proposed approach with the pairwise method studied in [9]. In addition, the impact of the temporal regularization term is investigated. The comparison is based on the endpoint error [15] between the estimated motion fields and the ground-truth. The different regularization parameters involved in all tested methods were tuned to provide the best performance.

A. Dataset

In this work we use a realistic cardiac US dataset with available ground-truth¹. The true motion fields provide the possibility of training the motion dictionaries as well as evaluating the motion estimation accuracy. In this work, the dictionaries are learnt using the LADdist sequence, and the tests are conducted on the LADprox sequence of the same dataset. The sequences contain 34 frames of size 224×208 (pixel size $0.7 \times 0.6\text{mm}$) with a frame rate of 22Hz (see Fig. 2-a for an example).

B. Motion estimation results

The errors obtained for the proposed multi-frame and pairwise approaches are reported in Table. I. The proposed multi-frame method with temporal regularization is referred to as Multi-frame TR, whereas the proposed method without temporal regularization is denoted as Multi-frame. One can see from this table that the proposed approaches lead to smaller global errors for this sequence. In addition, the use of the temporal regularization yields the best performance. Notice that the multi-frame approaches required significantly shorter execution times when compared to the pairwise method.

Method	Pairwise	Multi-frame	Multi-frame TR
Errors	0.14±0.11	0.12±0.11	0.11±0.10
Time (s)	52.33	10.27	12.54

TABLE I: Means \pm stds of the endpoint errors for the LADprox sequence and the average execution times per frame.

Fig. 1 shows the temporal evolution of the mean errors for the LADprox sequence. This figure shows that the proposed multi-frame approaches (with and without temporal regularization) outperform the pairwise method over the entire cardiac cycle (excluding the first frame). A relatively larger improvement can be observed in the systole phase, where the displacements have higher magnitudes. Overall, the multi-frame approach with temporal regularization leads to the best performance.

The systolic motion field obtained using the Multi-frame TR approach is displayed in Fig. 2-a. One can see from this figure that the estimated displacements are consistent with the inward systolic motions of the myocardium. In order to have more

¹The data and related papers can be found at <https://team.inria.fr/asclepios/data/strauss/>

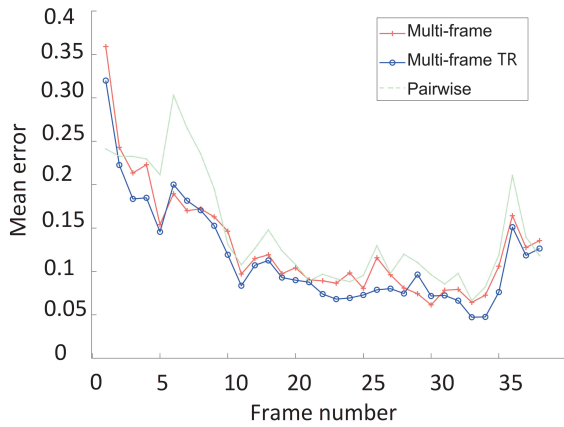
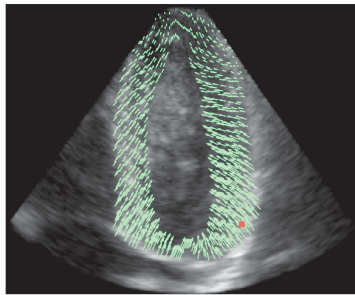
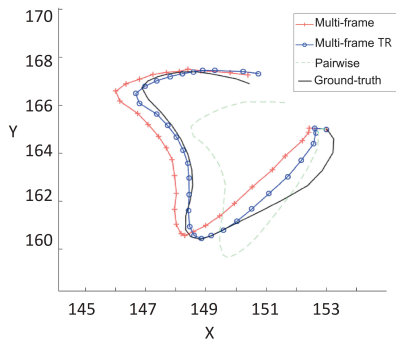


Fig. 1: Comparison of the mean errors vs time for the proposed multi-frame method with temporal regularization (TR), without temporal regularization and the pairwise method of [9].



(a)



(b)

Fig. 2: (a) Estimated motion field for a systolic frame. (b) The estimated trajectories of a point in the basal segment of the myocardium vs the ground-truth. The position of the point is indicated by a red square in (a).

insight into the effect of the temporal regularization, Fig. 2-b also shows the obtained trajectories for a point in the basal segment of the myocardium. This figure shows that the Multi-frame TR approach yields a trajectory that is significantly closer to the ground-truth when compared to those obtained with the other approaches. One can also see that the pairwise method provides the least accurate trajectory for this point.

IV. CONCLUSIONS

This paper introduced a new multi-frame approach with sparse and temporal regularizations for cardiac motion estimation in US images. The proposed approach used a regularization imposing sparsity in a learnt dictionary of typical cardiac motions. A weighted temporal regularization term enforcing piecewise smooth trajectories was also proposed. Finally, temporal consistency was incorporated by estimating the motions of an entire sequence at once without relying on a single reference frame. An efficient optimization strategy using the C-SALSA scheme was designed to solve the optimization problem resulting from the different regularizations. The proposed method was tested using a dataset of realistic cardiac US simulations. Experimental results showed an improvement in the motion estimation accuracy with a reduced computational time in comparison with a pairwise method also employing a dictionary learning-based regularization.

REFERENCES

- [1] C. Cottrell and J. N. Kirkpatrick, "Echocardiographic strain imaging and its use in the clinical setting," *Expert Review of Cardiovascular Therapy*, vol. 8, no. 1, pp. 93–102, 2010.
- [2] D. Boukerroui, J. A. Noble, and M. Brady, "Velocity estimation in ultrasound images: A block matching approach," in *Proc. Int. Conf. Inform. Process. Med. Imaging*, Ambleside, UK, July 2003, pp. 586–598.
- [3] F. Yeung, S. F. Levinson, and K. J. Parker, "Multilevel and motion model-based ultrasonic speckle tracking algorithms," *Ultrasound in Medicine & Biology*, vol. 24, no. 3, pp. 427–442, 1998.
- [4] M. Suhling et al., "Myocardial motion analysis from B-mode echocardiograms," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 525–536, 2005.
- [5] M. Ledesma-Carbayo et al., "Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1113–1126, 2005.
- [6] 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, 1996.
- [7] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intell.*, vol. 17, no. 1, pp. 185–203, 1981.
- [8] 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, 2008.
- [9] 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.
- [10] N. Ouzir, A. Basarab, O. Lairez, and J.-Y. Tourneret, "Robust optical flow estimation in cardiac ultrasound images using a sparse representation," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 741–752, March 2019.
- [11] M. De Craene et al., "Temporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography," *Med. Image Analysis*, vol. 16, no. 2, pp. 427–450, 2012.
- [12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Int. Conf. Mach. Learning (ICML '09)*, Montreal, Quebec, Canada, 2009, pp. 689–696.
- [13] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer, "Modeling temporal coherence for optical flow," in *Proc. Int. Conf. Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 1116–1123.
- [14] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, March 2011.
- [15] 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, 2013.