Isotropic Total Variation Regularization of Displacements in Parametric Image Registration

Valeriy Vishnevskiy*, Tobias Gass, Gabor Szekely, Christine Tanner, Orcun Goksel

Computer Vision Laboratory, ETH Zurich, Switzerland

Abstract—**Spatial regularization is essential in image registration, which is an ill-posed problem. Regularization can help to avoid both physically implausible displacement fields and local minima during optimization. Tikhonov regularization (squared ℓ_2-norm) is unable to correctly represent non-smooth displacement fields, that can, for example, occur at sliding interfaces in the thorax and abdomen in image time-series during respiration. In this paper, isotropic Total Variation (TV) regularization is used to enable accurate registration near such interfaces. We further develop the TV-regularization for *parametric* displacement fields and provide an efficient numerical solution scheme using the Alternating Directions Method of Multipliers (ADMM). The proposed method was successfully applied to four clinical databases which capture breathing motion, including CT lung and MR liver images. It provided accurate registration results for the whole volume. A key strength of our proposed method is that it does not depend on organ masks that are conventionally required by many algorithms to avoid errors at sliding interfaces. Furthermore, our method is robust to parameter selection, allowing the use of the same parameters for all tested databases. The average target registration error (TRE) of our method is superior (10% to 40%) to other techniques in the literature. It provides precise motion quantification and sliding detection with sub-pixel accuracy on the publicly available breathing motion databases (mean TREs of 0.95 mm for DIR 4D CT, 0.96 mm for DIR COPDgene, 0.91 mm for POPI databases).**

Index Terms—**Sliding at anatomical interfaces, breathing motion, ADMM, 4DCT, 4DMR**

I. Introduction

Image registration is nowadays an essential part of many applications in medical image analysis, such as image data fusion, image guided interventions, atlas-based segmentation, volumetric image reconstruction from sliced acquisitions, motion tracking, radiation-therapy planning, and tissue elasticity estimation. All these applications rely on correctly estimating spatial anatomical correspondences between images. Such bijective correspondences might not exist in some scenarios, e.g. inter-patient or 2D registrations. However, if the anatomy changes are minimal, for example, in intra-patient registration of time-series 3D (4D) volumes, one-to-one correspondences can be assumed in the overlapping image regions. This is the case for the breathing motion that involves sliding between different parts of anatomy, such as the liver or the lungs and their surrounding (cf. Fig. 1). The resulting motion fields are non-smooth, and therefore difficult to represent and estimate using the commonly-used registration techniques with squared ℓ_2-norm, also known as Tikhonov regularization. A lot of more advanced regularization techniques, such as hyperelastic regularization [1] imply smooth deformation as well.

A simple and popular way to accommodate for discontinuities near sliding interfaces is to use binary masks for objects to be registered, such that optimization is only with respect to the image similarity inside these masks. This helps to dismiss the discontinuities on the mask border and avoids influence of the neighbouring structures on the registration result, but cannot provide meaningful displacements outside the mask. Binary masks have been employed for lung motion estimation in several works, as lungs are easier to segment than abdominal organs. Masks from automatic lung segmentations were used in [2] for extracting the motion inside the lungs. In landmark-based registration, lung masks are also used to restrict the location of the detected lung-specific landmarks, to be further used for matching and motion interpolation [3]. Lung masking allows for simple and effective registration methods based on bending energy regularization and normalized gradient-field image-metric as demonstrated in [4]. A combined displacement field for the thorax was created by simultaneously estimating two separate displacement fields inside and outside, while enforcing physically-valid motions along the mask surface through additional constraints [5].

Nevertheless, motion masking has several disadvantages. It requires an initial segmentation stage, which is a challenging problem in itself for many abdominal structures. Even when such segmentation is viable, for example for lungs in CT, this stage is still error-prone when done automatically and quite cumbersome manually. Furthermore, such segmentation poses a *hard decision* for any subsequent registration, i.e. the success of the latter will depend on the accuracy of the former. This will not only potentially yield errors where segmentation failed, but may also cause error propagation to other locations through smoothing. Additionally, using this approach sliding can obviously be *only* recovered at the given mask interface and neglects sliding outside or inside the mask, for example, between lung lobes when using lung masks [6]. Other specialized lung registration methods utilize temporal regularization, instead of spatial regularization of the displacements [7]. However, these approaches require the acquisition of intermediate breathing phases (not always available from common breath-hold images), which may impose practical limitations.

To allow for sliding motion, there also exist heuristic approaches such as the modified demons [8], [9] or the anisotropic demons [10], where the Gaussian filtering is substituted with anisotropic inhomogeneous filtering. Yet such methods do not involve an explicit regularization term, hence lacking a formal cost definition required for a proper opti-
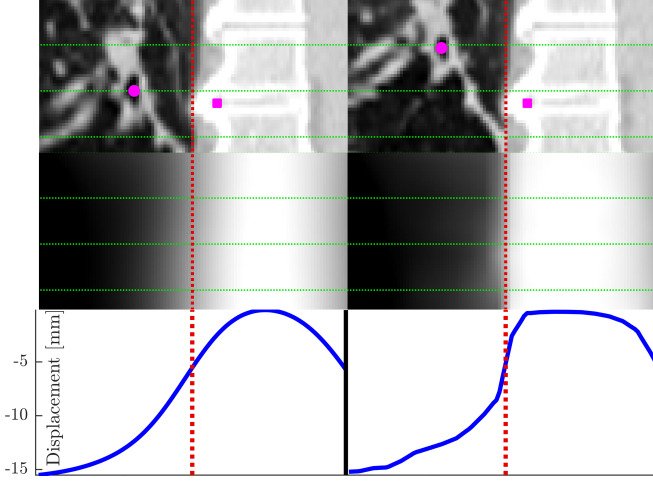
Fig. 1: Sliding motion example. (top) CT lung images of two extreme breathing phases with two landmarks. (middle) Estimated inferior-superior displacement field component in [-16,0] mm for (left) smoothness and (right) TV regularization. (bottom) Horizontal cross-section for inferior-superior displacements at the level of right landmark. This illustrates that TV-regularization does not oversmooth the motion transition.

mization scheme. In contrast, introducing an explicit penalty such as smoothness, curvature, or total variation (TV) [11] functionals allows for a well-defined optimization framework.

By penalizing TV of the displacement field it is possible to constrain its spatial incoherence without restricting it to be smooth. This property of TV was studied and applied in the computer vision community for estimating optical flow between scenes with independent motion of overlapping objects [12]. Unlike the smooth (Tikhonov) regularization, the TV functional is generally non-differentiable, since ℓ_1-norm is not smooth at zero. This greatly complicates the use of gradient-based continuous optimization methods, which usually rely on the problem smoothness. In [12], a smooth semi-norm approximation to TV was proposed, i.e. $\|x\|_1 \approx \varphi_\varepsilon(x) = \sqrt{x^2 + \varepsilon}$. This approximation was used in [13] to estimate motion masks automatically during the registration process via the segmentation model of Chan *et al.* [14]. Starting with large ε values and potentially-poor-but-smooth ℓ_1 approximations, one can then gradually decrease ε, and initialize in each step a gradient-based optimization with the previous solution. However, since $\partial\varphi_\varepsilon/\partial x = x/\sqrt{x^2 + \varepsilon}$ is undefined when $x=0$ and $\varepsilon=0$, the prescribed gradient-based optimization tends to be unstable in practice [15].

To address the unreliable behaviour of common gradient-based optimization algorithms for small ε values, discrete optimization methods for the registration problem, e.g. Markov random fields, were used in [16] for recovering breathing motion. Yet, fine displacement estimation requires a dense discretization of the search space, for which memory and computational complexity of common message passing and graph-cut based algorithms grow to an impractical level for precise 3D image registration. This issue was addressed in [17] by relaxing the graph structure to a minimum spanning tree, which reduces the degrees of freedom and enables fast (linear) inference. However, the minimum spanning tree was constructed from a graph weighted by the image intensities differences, which implies that the method "expects" sliding motion only at interfaces with high intensity gradient, which is not necessarily always true.

A recent approach for solving ℓ_1-norm problems is based on *duality*, which was applied to *non-parametric* medical image registration earlier in [15], [18] using sum of square differences as image dissimilarity metric. A census cost function was later proposed in [19], yielding improved and validated results for recovering breathing motion from 4D CT lung images. However, such non-parametric pixel-level approaches often lead to physically implausible motion fields, for which the authors proposed a combined median and Gaussian filtering of the displacement fields after each iteration. Such additional heuristic regularization schemes again prevent formal cost definitions and often complicate adaptation to new imaging problems.

In contrast, many registration techniques use parametric deformation models. For example, using B-splines [20] allows for explicit cost function and optimization formalisms, as well as offering more physically plausible displacements through robust schemes with large displacement capture ranges [21]. In this work, *parametric* image registration is formulated as a minimization problem with isotropic TV-regularization (pTV) and linear grid interpolation and the framework is shown to successfully approximate anatomical non-smooth (sliding) motion through a coherent optimization framework. To accommodate TV regularization, we use an efficient solution scheme via the *alternating directions method of multipliers* (ADMM), enabling better convergence in practice than conventional comparable techniques. Initial evaluation of our approach for only three databases, employing a simpler *anisotropic* TV regularization scheme (ℓ_1-norm of the displacement gradient magnitudes) with *ad hoc* parameter tuning, was presented in [22] and is referred herein as aTV.

II. METHODS

Estimating the N-dimensional displacement field $\mathbf{d} : \Omega \to \mathbb{R}^N$ that maps a moving image $\mathbf{f}_m : \Omega \to \mathbb{R}$ to a fixed image $\mathbf{f}_f : \Omega \to \mathbb{R}$ within a discrete image domain Ω is commonly formulated as the following optimization problem:

$$\mathbf{d}^* = \arg\min_{\mathbf{d}} \ \mathcal{F}(\mathbf{d}) = \arg\min_{\mathbf{d}} \ E_\mathrm{D}(\mathbf{d}; \mathbf{f}_f, \mathbf{f}_m) + \lambda E_\mathrm{R}(\mathbf{d}), \ (1)$$

where each N-dimensional pixel (voxel) has physical dimensions $\{\delta_n\}_{n=1}^N$ yielding a pixel volume of $\upsilon = \prod_{n \leq N} \delta_n$ and each image consists of L pixels, i.e. $|\Omega| = L$. Herein, E_D is an image dissimilarity metric and E_R is a spatial displacement regularization term, while λ controls the amount of regularization.

In this work, images $\mathbf{f} \in \mathbb{R}^L$ and displacement fields are treated as vectors; i.e. $\mathbf{d} = (\mathbf{d}_1^\mathsf{T}, \ldots, \mathbf{d}_N^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{LN}$, where $\mathbf{d}_n \in \mathbb{R}^L$ denotes the n-th component of the displacement field $\mathbf{d}$. The operator $\mathbf{f}[\mathbf{x}]$ (or $\mathbf{f}[l]$) is used to interpolate (or index) the image or displacement field at the position $\mathbf{x}$ (or pixel index l), where $\mathbf{d}[\mathbf{x}] = (\mathbf{d}_1[\mathbf{x}], \ldots, \mathbf{d}_N[\mathbf{x}])^\mathsf{T} \in \mathbb{R}^N$. Derivative operator

(which is a sparse matrix in the discrete image space) along the i-th component will be denoted as $\nabla_i \in \mathbb{R}^{L \times L}$, and the entire image gradient as $\nabla = \left(\nabla_1^\mathsf{T}, \ldots, \nabla_N^\mathsf{T}\right)^\mathsf{T} \in \mathbb{R}^{LN \times L}$. For forward differences, each row l_1 of ∇_i will have two non-zero elements (l_1, l_2) and (l_1, l_2') containing values $-1/\delta_i$ and $1/\delta_i$, respectively, if $\mathbf{x}[l_2']$ neighbors pixel $\mathbf{x}[l_2]$ in the i-th direction.

A. Image Dissimilarity Metric

Typically, E_D is a smooth metric, such as the sum of squared differences (SSD), local correlation coefficient (LCC), or normalized gradient field (NGF), to allow for explicit differentiation. We evaluated the following two image metrics:

SSD is the sum of squared intensity-differences between two images computed over all pixels, i.e.:

$$\mathrm{SSD}(\mathbf{f}_1, \mathbf{f}_2) = \sum_{\mathbf{x} \in \Omega} \frac{1}{2} \left(\mathbf{f}_1[\mathbf{x}] - \mathbf{f}_2[\mathbf{x}]\right)^2 \upsilon = \frac{\upsilon}{2}\|\mathbf{f}_1 - \mathbf{f}_2\|_2^2. \quad (2)$$

Then, its gradient vector is

$$\frac{\partial \, \mathrm{SSD}(\mathbf{f}_1, \mathbf{f}_2)}{\partial \mathbf{f}_1} = (\mathbf{f}_1 - \mathbf{f}_2) \, \upsilon, \quad (3)$$

where pixel volume υ is incorporated to keep measures consistent when resampling images in pyramidal image registration.

LCC is the sum of weighted correlation coefficients of image intensities computed over n-dimensional patches centered at each pixel. As the patch size can be quite large, the explicit computation of the LCC metric and its derivatives becomes impractical. Following [23], [24], we use a spatial-invariant Gaussian weighting kernel H_w with standard deviation (bandwidth) w. LCC is then computed in terms of element-wise arithmetical operations and convolutions. This can be carried out in the Fourier domain, where the convolution's computational complexity does not depend on the patch size.

$$\bar{\mathbf{f}} = H_w \otimes \mathbf{f}, \quad \sigma^2(\mathbf{f}) = \overline{\mathbf{f}^2} - \bar{\mathbf{f}}^2, \quad (4)$$

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle = \overline{\mathbf{f}_1 \circ \mathbf{f}_2} - \bar{\mathbf{f}}_1 \circ \bar{\mathbf{f}}_2, \quad (5)$$

$$\mathrm{LCC}(\mathbf{f}_1, \mathbf{f}_2) = -\sum_{\mathbf{x} \in \Omega} \mathrm{CC}[\mathbf{x}] = -\sum_{\mathbf{x} \in \Omega} \frac{\langle \mathbf{f}_1, \mathbf{f}_2 \rangle}{\sigma(\mathbf{f}_1) \circ \sigma(\mathbf{f}_2)}[\mathbf{x}]\upsilon, \quad (6)$$

$$\frac{\partial \, \mathrm{LCC}(\mathbf{f}_1, \mathbf{f}_2)}{\partial \mathbf{f}_1} \approx -\frac{1}{\sigma(\mathbf{f}_1) \circ \sigma(\mathbf{f}_1)} \circ \left((\mathbf{f}_2 - \bar{\mathbf{f}}_2) - \right.$$
$$\left. - (\mathbf{f}_1 - \bar{\mathbf{f}}_1) \circ \frac{\langle \mathbf{f}_1, \mathbf{f}_2 \rangle}{\sigma^2(\mathbf{f}_1)} \right), \quad (7)$$

where $\otimes$ denotes the convolution operator and $\circ$ denotes element-wise multiplication. Eq. (7) for the metric gradient is approximate by dropping the final convolution as in [24]. We observed this approximation to outperform the exact formula in our initial experiments. Further details can be found in [24].

B. Registration Regularization

The widely used *smoothness* regularization penalizes the squared ℓ_2-norm derivatives of the displacement field components [20]:

$$E_\mathrm{R}^\mathrm{sm}(\mathbf{d}) = \upsilon \sum_{n \leq N} \|\nabla \mathbf{d}_n\|_2^2. \quad (8)$$

Any continuous optimization method can then be applied for minimizing (1) with such a regularization term, since dissimilarity metric E_D and regularizer E_R are smooth and allow for analytical differentiation with respect to the transformation parameters.

C. TV-Regularization

Many methods employ the anisotropic version of total variation measure by treating each gradient directional of each displacement component independently:

$$E_\mathrm{R}^\mathrm{aTV}(\mathbf{d}) = \upsilon \sum_{n \leq N} \|\nabla \mathbf{d}_n\|_1. \quad (9)$$

Note that TV does not "distinguish" a sharp transition from a smooth one. Unlike our previous work [22], herein we use isotropic TV-regularization [25] (referred hereafter as *TV*), which considers all components of the displacement gradients jointly. This is better suited to describe motion that is not aligned with Cartesian axes:

$$E_\mathrm{R}^\mathrm{TV}(\mathbf{d}) = \upsilon \sum_{l \leq L} \sqrt{\sum_{i,j \leq N} (\nabla_i \mathbf{d}_j[l])^2} = \upsilon \|\mathcal{D}(\mathbf{d})\|_{2,1}, \quad (10)$$

$$\mathcal{D}(\mathbf{d}) = \left[\nabla_1 \mathbf{d}_1 \, \nabla_2 \mathbf{d}_1 \ldots \nabla_1 \mathbf{d}_2 \ldots \nabla_N \mathbf{d}_N\right]^\mathsf{T} \in \mathbb{R}^{N^2 \times L}. \quad (11)$$

This functional is no longer smooth, i.e. it is non-differentiable at the spatial locations where transformation gradients vanish, since $\sqrt{x}$ is not smooth at zero. This leads to poor performance when using continuous gradient-based optimization methods. This is addressed in this work by using ADMM optimization [26], [27], described later in Sec. II-F. Next, we introduce our displacement parametrization approach that allows us to impose TV regularization on the displacement field simply and efficiently.

D. Parametrizing Displacement Fields

Reducing the dimensionality via parametrization of the search space is often desired in practical optimization-based algorithms, as it can improve robustness and convergence speed of numerical methods. It is also a form of implicit regularization, since many physically implausible displacement fields can not be represented by the selected parametrization.

As in the free-form deformations (FFD) framework [20], [28], we parametrize the displacement field $\mathbf{d}$ via interpolation of the displacements $\mathbf{k}$ on a regularly spaced (K_n-pixel span in n-th direction) control point grid, i.e. $\mathbf{d} = \mathbf{d}(\mathbf{k})$. Unlike more popular cubic B-splines, which were shown in [20] to be a suitable model for elastic deformations, we use 1st order B-splines (linear interpolation), which guarantees that the interpolated displacement values are strictly bounded by the corresponding control point displacements. 1st order B-splines also help to avoid overshooting effects while approximating sharp transitions (see Fig. 2(a)) and have smaller spatial support.

We impose regularization on the control grid displacements $\mathbf{k}$ instead of the displacement field $\mathbf{d}$ itself. This approximation was used in smooth FFD registration and it was shown in [29]
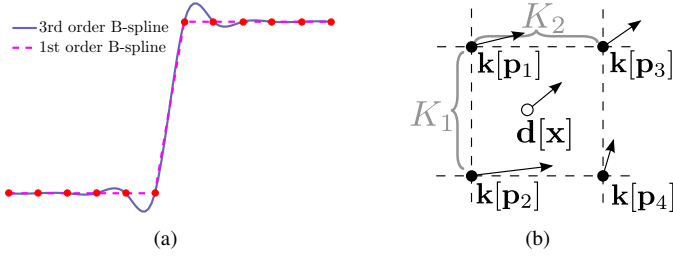
Fig. 2: (a) Linear (1st order) and cubic (3rd order) B-splines approximation of a sharp transition. (b) Illustration of spatial arrangement of the control grid points and parametrized displacement $\mathbf{d}$.

to be sufficient for the regularization. For 1st order B-splines, TV of the control points displacements is an upper bound for TV on the dense displacement field: $E_{\mathrm{R}}^{\mathrm{TV}}(\mathbf{d}(\mathbf{k})) \leq E_{\mathrm{R}}^{\mathrm{TV}}(\mathbf{k})$, see [22] for a proof. This can also be explained by Jensen's inequality using the fact that TV measure is a convex function.

Finally, we can formulate parametric TV-regularized image registration as the following optimization problem:

$$\begin{aligned}
\mathbf{k}^* &= \arg\min_{\mathbf{k}} \mathcal{F}(\mathbf{d}(\mathbf{k})) \\
&= \arg\min_{\mathbf{k}} E_{\mathrm{D}}(\mathbf{d}(\mathbf{k}); \mathbf{f}_{\mathrm{f}}, \mathbf{f}_{\mathrm{m}}) + \lambda\eta\|\mathcal{D}(\mathbf{k})\|_{2,1},
\end{aligned} \quad (12)$$

where the volume of the control grid cell $\eta = \prod_{i\leq N}\delta_i K_i$.

E. Full Image Metric Gradient

To speed-up any energy-based registration, it is beneficial to compute *analytical* (rather than numerical) gradients of the energy $\mathcal{F} = E_{\mathrm{D}} + \lambda E_{\mathrm{R}}$ that should be minimized with respect to the free parameters (here grid displacements). As seen later, in our algorithm we only need to compute the gradient of E_{D}, which can be done via the chain rule:

$$\begin{aligned}
\frac{\partial E_{\mathrm{D}}(\mathbf{f}_{\mathrm{f}}, \mathbf{f}_{\mathrm{m}}(\mathbf{d}(\mathbf{k})))}{\partial \mathbf{k}_i[\mathbf{p}]} &= \sum_{\mathbf{x}\in\Omega} \frac{\partial E_{\mathrm{D}}(\mathbf{f}_{\mathrm{f}}, \mathbf{f}_{\mathrm{m}}(\mathbf{d}(\mathbf{k})))}{\partial \mathbf{f}_{\mathrm{m}}(\mathbf{d}(\mathbf{k}))}[\mathbf{x}]\cdot \\
&\nabla_i \mathbf{f}_{\mathrm{m}}(\mathbf{d}(\mathbf{k}))[\mathbf{x}] \cdot \frac{\partial \mathbf{d}(\mathbf{k})}{\partial \mathbf{k}_i[\mathbf{p}]}[\mathbf{x}],
\end{aligned} \quad (13)$$

where the derivative is taken with respect to the i-th displacement component of the control grid at pixel $\mathbf{p}$. The first term is the image metric derivative introduced in Eqs. (3) and (7). The second term is the deformed image gradient in i-th direction computed at position $\mathbf{x}$. The third term is the Jacobian of the displacement parametrization. Since trilinear parametrization is a 1st order B-spline with finite spatial support, its Jacobian also has finite support. Eventually, the image metric derivative (13) with respect to the i-th displacement component of the control grid point at position $\mathbf{p}$ can be computed as the weighted sum of the element-wise multiplications of the first two terms [29]:

$$\frac{\partial \mathbf{d}(\mathbf{k})}{\partial \mathbf{k}_i[\mathbf{p}]}[\mathbf{x}] = \prod_{n\leq N}\left(1 - \frac{|p^n - x^n|}{K_n}\right)_+. \quad (14)$$

Here, x^n and p^n denote the n-th component of the corresponding pixel and control point locations $\mathbf{x}$ and $\mathbf{p}$ in terms of *pixel* spacing, and $(x)_+ = \max\{x,0\}$.

F. Numerical Scheme: ADMM

As was mentioned above, it is challenging to solve (12), as it involves simultaneous minimization of the smooth non-convex term E_{D} and non-differentiable but convex term $E_{\mathrm{R}}^{\mathrm{TV}}$. The gradient descent methods (GD) show a poor performance (Fig. 3(a)), mostly because the target function gradient $\nabla\mathcal{F}$ does not always exist. Among various duality-based approaches for optimization proposed in the literature, we use the alternative direction method of multipliers (ADMM) for its flexibility, generality, and ease of implementation. We split the parametrizations of the data and regularization terms, yielding affine equality constrained composite optimization (AECCO), which is solved by ADMM. To perform this separation, we introduce a redundant variable Z to (12), which leads to the following equivalent problem:

$$\begin{aligned}
(\mathbf{k}^*, \mathbf{Z}^*) = \arg\min_{\mathbf{k}, \mathbf{Z}} \quad &E_{\mathrm{D}}(\mathbf{d}(\mathbf{k}); \mathbf{f}_{\mathrm{f}}, \mathbf{f}_{\mathrm{m}}) + \lambda\eta\|\mathbf{Z}\|_{2,1}, \\
\text{s.t.} \quad &\mathcal{D}(\mathbf{k}) = \mathbf{Z}.
\end{aligned} \quad (15)$$

Using the augmented scaled Lagrangian formalism, the ADMM is a variant of the dual descent method [26], which can be explicitly written as the following iterative scheme:

$$\mathbf{k}^{j+1} = \arg\min_{\mathbf{k}} E_{\mathrm{D}}(\mathbf{d}(\mathbf{k}); \mathbf{f}_{\mathrm{f}}, \mathbf{f}_{\mathrm{m}}) + \frac{\rho}{2}\|\mathcal{D}(\mathbf{k}) - (\mathbf{Z}^j - \mathbf{U}^j)\|_F^2, \quad (16a)$$

$$\mathbf{Z}^{j+1} = \arg\min_{\mathbf{Z}} \|\mathbf{Z}\|_{2,1} + \frac{\rho}{2\lambda\eta}\|Z - (\mathcal{D}(\mathbf{k}^{j+1}) + \mathbf{U}^j)\|_F^2, \quad (16b)$$

$$\mathbf{U}^{j+1} = \mathbf{U}^j + \mathcal{D}(\mathbf{k}^{j+1}) - \mathbf{Z}^{j+1}, \quad (16c)$$

where $\|.\|_F$ denotes the Frobenius matrix norm, Eq. (16a) is the smooth optimization problem, (16b) is the convex proximal mapping and (16c) is the dual update. Note that none of these steps involve simultaneous optimization over $\mathbf{k}$ and $\mathbf{Z}$ or of E_{D} and $E_{\mathrm{R}}^{\mathrm{TV}}$.

The $\mathbf{k}$-update step (16a) is computationally the most intensive part, as it optimizes over all control grid displacements. It is a smooth optimization problem that can be solved using gradient-based minimization techniques by adding the gradient of $\frac{\rho}{2}\|\mathcal{D}(\mathbf{k}) - (\mathbf{Z}^j - \mathbf{U}^j)\|_F^2$, which equals to $\rho\mathcal{D}^*(\mathcal{D}(\mathbf{k}) - (\mathbf{Z}^j - \mathbf{U}^j))$, to the gradient of the dissimilarity metric in (13). Here $\mathcal{D}^*$ denotes conjugate finite difference operator such that

$$\mathcal{D}^*(\mathcal{D}(\mathbf{k})) = \left[\boldsymbol{\theta}_1^\top \dots \boldsymbol{\theta}_N^\top\right]^\top, \text{ with } \boldsymbol{\theta}_i = \sum_{j\leq N}\nabla_j^\top\nabla_j\mathbf{k}_i. \quad (17)$$

The penalty parameter ρ introduced by the Lagrangian augmentation is usually initialized with a value between 1 and 1000 and then updated heuristically to accelerate convergence [26], as is described in Algorithm 1. $\mathbf{U}$ is a scaled dual variable. The $\mathbf{Z}$-update step (16b) can be decomposed into per-pixel Euclidean proximal operators on the rows of matrix $\mathbf{A} = \mathcal{D}(\mathbf{k}^{j+1}) + \mathbf{U}^j$:

$$\mathbf{Z}_{l,*}^{j+1} = \arg\min_{\mathbf{z}\in\mathbb{R}^{n^2}} \|\mathbf{z}\|_2 + \frac{\rho}{2\lambda\eta}\|\mathbf{z} - \mathbf{A}_{l,*}\|_2^2, \quad (18)$$

where $\mathbf{A}_{l,*}$ denotes l-th row of the matrix $\mathbf{A}$. The Euclidean proximal mapping has an explicit closed-form solution sometimes called block soft thresholding [26]:

$$\left(1 - \frac{\lambda\eta}{\rho\|\mathbf{A}_{l,*}\|_2}\right)_+ \mathbf{A}_{l,*}. \quad (19)$$

ALGORITHM. 1. The pTV registration algorithm.

$\mathbf{f}_m$, $\mathbf{f}_f$ — images; λ — regularization; w — LCC kernel bandwith; M_{pyr} — # of image pyramid levels; M_{ref} — # of grid refinements; M_{iter} — max. # of ADMM iterations; M_{LBFGS} — max. # of LBFGS iterations ; ε_{tol} — optimization argument tolerance (10^{-2});

ADMM internal parameters $\rho_0 = 5$, $\mu = 10$, $\tau = 1.3$.

Result: Displacement field $\mathbf{d}$

$\mathbf{k}^{1,0} \leftarrow \mathbf{0}$ or initialize

for $j \leftarrow 1$ $\;$ to $\;$ $M_{pyr} + M_{ref}$ **do**

$\quad$ **if** $j \leq M_{pyr}$ **then**

$\quad\quad$ $\mathbf{f}_m^j \leftarrow (M_{pyr} - j)$-th Gaussian pyramid level of $\mathbf{f}_m$

$\quad\quad$ $\mathbf{f}_f^j \leftarrow (M_{pyr} - j)$-th Gaussian pyramid level of $\mathbf{f}_f$

$\quad$ **else**

$\quad\quad$ $\mathbf{f}_m^j \leftarrow \mathbf{f}_m^{j-1}$; $\quad$ $\mathbf{f}_f^j \leftarrow \mathbf{f}_f^{j-1}$

$\quad$ **end**

$\quad$ $Z^{j,0} \leftarrow \mathcal{D}(\mathbf{k}^{j,0})$; $\quad$ $U^{j,0} \leftarrow 0$; $\quad$ $\rho \leftarrow \rho_0$

$\quad$ **while** $i \leq M_{iter}$ **and** $\|\mathbf{k}^{j,i} - \mathbf{k}^{j,i-1}\|_\infty \geq \varepsilon_{tol}$ **do**

$\quad\quad$ $\mathbf{k}^{j,i+1} \leftarrow$ solution of (16a) with $M_{LBFGS}, w, \mathbf{f}_m^j, \mathbf{f}_f^j$

$\quad\quad$ $Z^{j,i+1} \leftarrow$ solution of (16b)

$\quad\quad$ $U^{j,i+1} \leftarrow$ result of (16c)

$\quad\quad$ $r \leftarrow \mathcal{D}(\mathbf{k}^{j,i+1}) - Z^{j,i+1}$; $s \leftarrow \rho\mathcal{D}^*(Z^{j,i+1} - Z^{j,i})$

$\quad\quad$ **if** $\|r\|_2 \geq \mu\|s\|_2$ **then**

$\quad\quad\quad$ $\rho \leftarrow \tau\rho$; $\quad$ $U^{j,i+1} \leftarrow U^{j,i+1}/\tau$

$\quad\quad$ **end**

$\quad\quad$ **if** $\|s\|_2 \geq \mu\|r\|_2$ **then**

$\quad\quad\quad$ $\rho \leftarrow \rho/\tau$; $\quad$ $U^{j,i+1} \leftarrow \tau U^{j,i+1}$

$\quad\quad$ **end**

$\quad\quad$ $i \leftarrow i + 1$

$\quad$ **end**

$\quad$ $\mathbf{k}^{j+1,0} \leftarrow$ linearly upsample $\mathbf{k}^{j,i}$

end

return $\mathbf{d} = \mathbf{d}(\mathbf{k}^{j,i})$

The foregoing numerical scheme is the core of our registration algorithm, which is formally described in Algorithm 1. The method has three major parameters (λ, w, K) that affect the registration quality and should be tuned.

Regularization parameter λ should be tuned for a given database, taking into account the imaging modality, noise level, and expected amount of motion. The first-order necessary optimality condition for problem (1) can be written as $\nabla E_D = -\lambda \nabla E_R$. Since correlation is inversely proportional to the standard deviation of noise, s, the LCC metric gradient (7) can be considered approximately to be proportional to $\frac{1}{s}$. Consequently, if λ_0^L is the best performing regularization parameter for a pair of images with noise level s_0, then a good initial point for images with noise level s_1 is $\lambda_1^L = \lambda_0^L \frac{s_0}{s_1}$. Similar reasoning can be used for the SSD metric, yielding the following tuning rule: $\lambda_1^S = \lambda_0^S \frac{s_1}{s_0}$. Hence, λ needs to be tuned only for different motion types.

The weighting kernel bandwidth w (for LCC) depends purely on the image modality and pixel resolution. To have sufficient statistics we should restrict the kernel's full-width at half-maximum (FWHM) to at least 1 pix. Also, setting w to large values disables the LCC's ability to handle local intensity variations. As shown later, the registration results are in practice insensitive to this parameter. For all experiments and all databases we used $w=2.5$ px, which implies (using 3-sigma rule) an effective patch size of ≈ 15 px edge-length. Control grid spacing K controls the flexibility of the displacement field. Small values of K imply finer grid resolutions and more degrees of freedom in the transformation, at a cost of higher computational complexity.

1) Implementation Details: Both fixed and moving image intensities are converted to double precision and then mapped to the [0, 1] interval as described in Sec. III. We use trilinear interpolation for image warping.

Gaussian pyramids are used to downsample images, keeping pixel spacing at each pyramid level as isotropic as possible. The number of pyramid levels M_{pyr} is set to the largest value such that the coarsest image level has at least 8 pixels along each dimension: $M_{pyr} = \lceil \log_2 \frac{m}{8} \rceil$, where m is the image size along the direction with the minimum size after resampling to isotropic resolution. Registration at each pyramid level is initialized by linearly upsampling from the previous level's control grid displacements, keeping the same control grid spacing in terms of pixels. When the finest level of the image pyramid is reached, we start to subdivide the control grid (M_{ref} times), bisecting the pixel span of each grid cell. This coarse-to-fine approach helps to avoid local minima during optimization process and is finalized at the original image resolution, where no image smoothing is applied. Blurring is always kept below grid spacing to minimize its influence on the registration of sliding interfaces.

Displacement field derivatives in (11) are calculated with the forward difference scheme to allow for spatial discontinuities. Image gradients $\nabla_i \mathbf{f}$ in (13) are computed with the central difference scheme.

Our method was implemented in `Matlab` with the time-consuming procedures in `C++`. Computation of the LCC metric and its gradients was done in Fourier domain using GPUs, when available. The $\mathbf{k}$-update step in (16a) is solved with the quasi-Newton limited-memory BFGS distributed in the `minFunc`[1] package with $M_{LBFGS}=5$ and the number of Hessian corrections set to 100 or 20, depending on available memory. The LBFGS iterations were initialized with the solution and Hessian approximation from the previous ADMM iteration. The experiments were conducted on a 6-core Intel Xeon 2.4 GHz processor. The peak memory consumption for registering a pair of $256{\times}256{\times}100$ voxel images was 1.2 GB.

III. Results and Discussion

We evaluated our *parametric TV* registration (pTV) method on four three-dimensional ($N{=}3$) abdominal time-series databases that involve breathing motion: three lung CT databases and one liver MRI database. All databases had manually annotated landmarks. The target registration error (TRE) is defined as the Euclidean distance between two corresponding landmarks after registration. The mean TRE was used to summarize registration accuracy. To keep the evaluation methodology in accordance with the literature, we used the snap-to-voxel TRE evaluation strategy (the landmark position

---

[1] http://www.di.ens.fr/~mschmidt/Software/minFunc.html

TABLE I: Mean TRE [mm] over 40 landmarks for the 9 POPI registrations. The lowest value per column is marked in bold.

Method	TRE for each image pair									Mean TRE
	#1	#2	#3	#4	#5	#6	#7	#8	#9	
Demons	1.28	1.38	1.39	1.22	1.24	1.25	1.29	1.12	1.11	1.25
FFD	0.79	0.80	1.13	1.11	1.10	1.20	1.20	0.88	0.92	1.01
aTV [22]	0.72	0.71	1.12	1.01	1.11	**1.03**	**1.06**	0.84	**0.81**	0.93
pTV	**0.66**	**0.66**	**1.11**	**0.98**	**1.08**	1.06	1.08	**0.76**	0.82	**0.91**

from the registration is snapped to the nearest voxel center) for 4D-CT DIR and COPDgene databases (Secs. III-B, III-C). For other datasets TREs were computed at (linearly interpolated) sub-voxel positions, without snapping to voxels.

Following [6], [30] we estimate and visualize the local amount of sliding of the computed displacement fields by the maximum shear stretch γ extracted from the eigendecomposition of the Jacobian matrix $J[\mathbf{x}]$ of the local transformation:

$$\gamma[\mathbf{x}] = \frac{\sqrt{\lambda_N(J_\mathbf{x}^\mathsf{T} J_\mathbf{x})} - \sqrt{\lambda_1(J_\mathbf{x}^\mathsf{T} J_\mathbf{x})}}{2}, \qquad (20)$$
$$J_\mathbf{x}(i,j) = \delta_{ij} + \nabla_j \mathbf{d}_i[\mathbf{x}], \quad J_\mathbf{x} \in \mathbb{R}^{N \times N},$$

where δ_{ij} is the Kronecker's delta, since $\mathbf{d}$ is the *displacement* field. The largest and the smallest eigenvalues of the corresponding matrix are denoted as λ_N and λ_1, respectively. The sliding amount $\gamma[\mathbf{x}]$ is a non-negative quantity, where higher values indicate more prominent sliding at $\mathbf{x}$.

To show the robustness of our method to parameter tuning, we used a single parameter configuration for all datatsets and all modalities, namely $\lambda = 0.1$, LCC image metric with $w = 2.5$ mm, and control point grid spacing K set to $4 \times 4 \times 4$ voxels. The parameter tuning process is described in Sec. III-E. For comparison, registration with smoothness regularization (squared ℓ_2-norm of gradients) was also performed using the same image metric and grid configuration. The regularization parameter of smooth registration was tuned via grid search for its best performance in terms of mean TRE for all cases of a particular database.

To test the proposed numerical optimization scheme, we compared the described ADMM-based (pTV) optimization of (12) with regular minimization algorithms: limited-memory BFGS (LBFGS) and gradient descent (GD). Fig. 3(a) compares algorithms in terms of minimized energy $\mathcal{F}(\mathbf{d}(\mathbf{k}))$ and mean TRE for one pair of 3D liver MR images. Unlike ADMM, the LBFGS and GD implementations accept only steps that decrease the target energy, which results in monotonically decreasing energy. However, the pTV was able to converge to a better optimum (lower energy) with fewer function evaluations. In the presented case, lower energy minimum value *generally* led to better TRE. This statement also turned out to be correct in our experiments on intra-patient CT and MR image sequences capturing breathing data, with fixed and *sufficient* amount of regularization λ. However, generally, residual image metric or optimized target value should not be treated as the direct measure of the non-rigid registration quality [31].

A. 4D-CT POPI Database

The POPI database [32], which is freely available[2], consists of 10 3D CT reconstructions of different phases of a single breathing cycle. Each breathing phase image has a resolution of $0.98 \times 0.98 \times 2$ mm^3 and is complemented with a lung mask and 40 corresponding anatomical landmarks. Following the organizers' approach, we resized all images to an isotropic $2 \times 2 \times 2$ mm^3 resolution. Then, the first image was registered to others, yielding 9 image pairs for evaluating the method. Mean run-time of our proposed registration method was 80 seconds per image-pair.

For this database the organizers also publicly distribute the displacement fields produced by the FFD and demons registrations, which we used for comparison in our evaluation. Even though these methods already provide mean subpixel registration accuracy, our method achieves an average TRE improvement of 10% over FFD, as can be seen in Table I. This is partially due to the replacement of the squared ℓ_2-regularization used by the FFD method, which is unable to capture sliding motion near the lung wall.

B. 4D-CT DIR Database

Similarly to the POPI, the DIR[3] database provides 4D-CT breathing sequences with annotated landmarks. It consists of 10 different sequences with an average voxel resolution of $1 \times 1 \times 1$ mm^3 and typical size $256 \times 256 \times 100$ voxels (see Fig. 4, 5). For this database the average TRE is calculated for 300 anatomical landmarks between images of extreme inhale and exhale phases. In our experiments we clipped the image intensities between 50 and 1200 HU, and then scaled this range to $[0, 1]$. All images were then resampled to an isotropic $1 \times 1 \times 1$ mm^3 voxel resolution, leading to an average pTV run-time of 3 minutes.

Registration accuracy results of DIR evaluation are published and updated online, which allows for a broad comparison of methods. Fig. 3(c) and Table II show the mean TRE for each 4D-CT case, comparing pTV to the best performing published methods. The results on this database indicate that pTV outperform all mask-free methods (excluding our initial results [22]) at least by 0.36 mm (27%), with a mean TRE of 0.95 mm. This is merely 0.01 mm (1%) worse than the best masked method (namely, NGF), which estimates the displacement field only inside the lungs. The non-parametric method with anisotropic TV regularization (cTVL1), which employs a similar optimization approach to ours, is outperformed by our parametric approach by 29% when no lung masks are used. Compared to the masked cTVL1 version, our mask-free method still shows a 4% reduction in TRE.

An illustration of the estimated displacement fields can be seen in Figs. 4 and 5. The landmark errors visualized in Figs. 5 (b,c), show that large errors occur in the vicinity of sliding interfaces, especially at lung borders when smoothness-regularized registration is used. Fig. 3(b) illustrates the ability of pTV to capture sliding motion near lung walls.

After registration we compute the amount of sliding motion using Eq. (20). Overlaying the estimated amount of sliding

---

[2]http://www.creatis.insa-lyon.fr/rio/popi-model
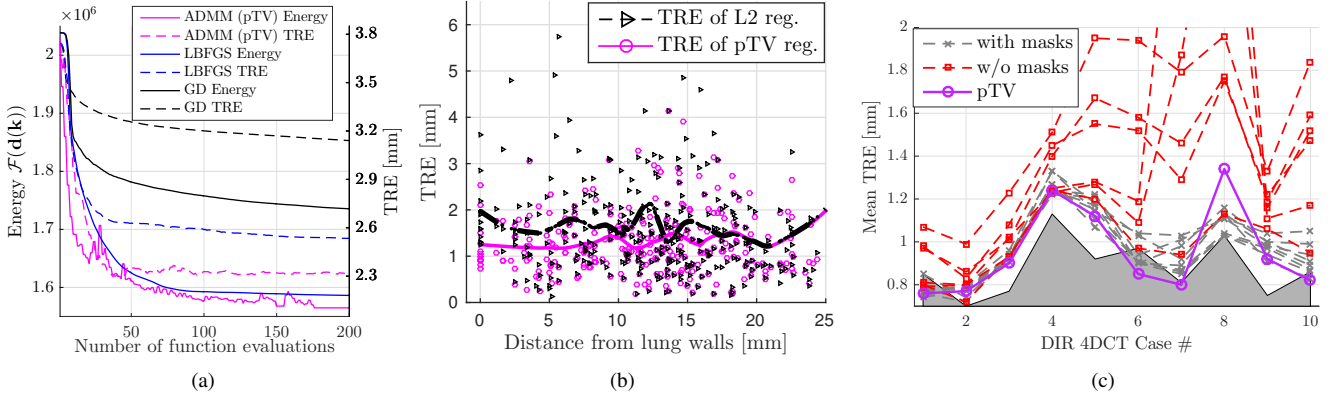[3]http://www.dir-lab.com

Fig. 3: (a) Image registration energy (1) and TRE for the same problem formulation. (b) Registration accuracy with respect to landmark distance from the lung wall, which was calculated from the discrete approximate lung mask. (c) Proposed method pTV compared to the best results from DIR website in Dec. 2015. Interobserver variability depicted with the gray region.
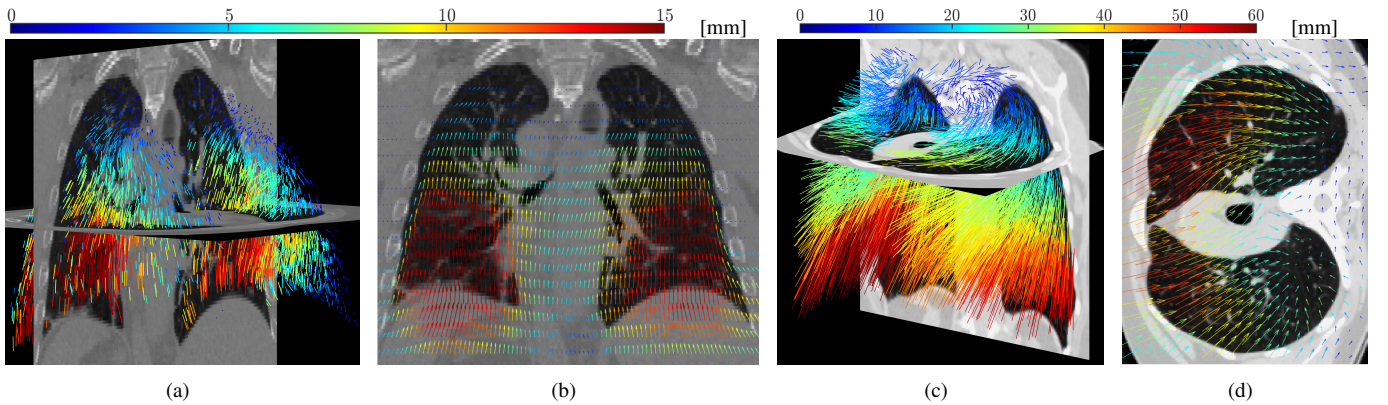

Fig. 4: Displacement fields estimated by the pTV method for the DIR 4D CT case # 4 (a,b) and COPDgene case #4 (c,d) color-coded with the displacement magnitude. (a) 3D vector field visualization; (b) coronal image slice with displacement field overlaid; (c) 3D vector field visualization; (d) axial image slice with displacement field overlaid. Displacement fields in (a,c) are masked with coarse lung masks for 3D visualization purposes.

$\gamma[\mathbf{x}]$ on the dorsal and ventral coronal slices (Fig. 6), prominent sliding can be seen on the lung border, especially in the inferior parts. Sliding motion *inside* the lungs (between its lobes) is also observed, and demonstrates an anatomically-viable pattern similarly to those shown in [6].

C. 3D Breath-Hold CT DIR COPDgene Database

In contrast to the two previous databases that were acquired with respiratory-correlated 4D reconstruction imaging, the COPDgene study [41] provides 3D breath-hold CT lung images. It consists of 10 cases, each case including a pair of images at end-inhale and end-exhale, together with 300 publicly-available landmarks per image. The average image resolution is $0.6 \times 0.6 \times 2.5\,\mathrm{mm}^3$ and average image size is $512 \times 512 \times 120$ voxels (see Fig. 4). Larger motion amplitude than the 4D-CT images (12–31 mm vs. 4-15 mm) makes the COPDgene a greater challenge for registration.

We resampled the images to $1 \times 1 \times 1\,\mathrm{mm}^3$ voxel spacing using Gaussian filter (with $1.5 \times 1.5 \times 0.5$ voxels kernel width) to avoid aliasing artifacts, and we clipped image intensities to $[50, 1200]$ HU interval. Image registration for each case took 3–6 minutes.

Much fewer registration algorithms were validated on the more challenging COPDgene database. These methods all use lung masks or specialized keypoint detectors tuned for lung CT imaging. Even then, Table II shows that our method is superior to alternatives regarding TRE for 8 out of 10 cases, showing poor performance only for case #2. Overall mean TRE is 0.96 mm, which is 0.12 mm (11%) better compared to the state-of-the-art method [3], which is specialized for this database.

D. 4D-MRI Liver Sequences

Since CT provides poor contrast for internal liver structures, MRI is often used to observe its motion. The 4D-MR liver database from [10] consists of 8 breathing sequences recorded from 4 subjects. Unlike the previous three lung CTs, this database provides 32 landmarks inside and 20 landmarks outside the moving object (liver) at both extreme breathing phases. The images have $1.37 \times 1.37 \times 4\,\mathrm{mm}^3$ resolution and a typical size of $164 \times 189 \times 23$ voxels. In accordance with previous experiments, we resampled all images to an isotropic $1 \times 1 \times 1\,\mathrm{mm}^3$ voxel spacing. pTV run-time on this database was 30–60 s.

The results reported in Table III indicate an accuracy improvement with pTV registration by 38% inside and 45% outside the liver in comparison to the results reported in [10]. Moreover, pTV was 22% more accurate than aTV. As in

TABLE II: Mean (snap-to-voxel) TRE on DIR database for masked and mask free methods as reported of the DIR webpage (09.12.2015). Per column the best TRE is given in bold, and the best TRE without masks is highlighted in green.

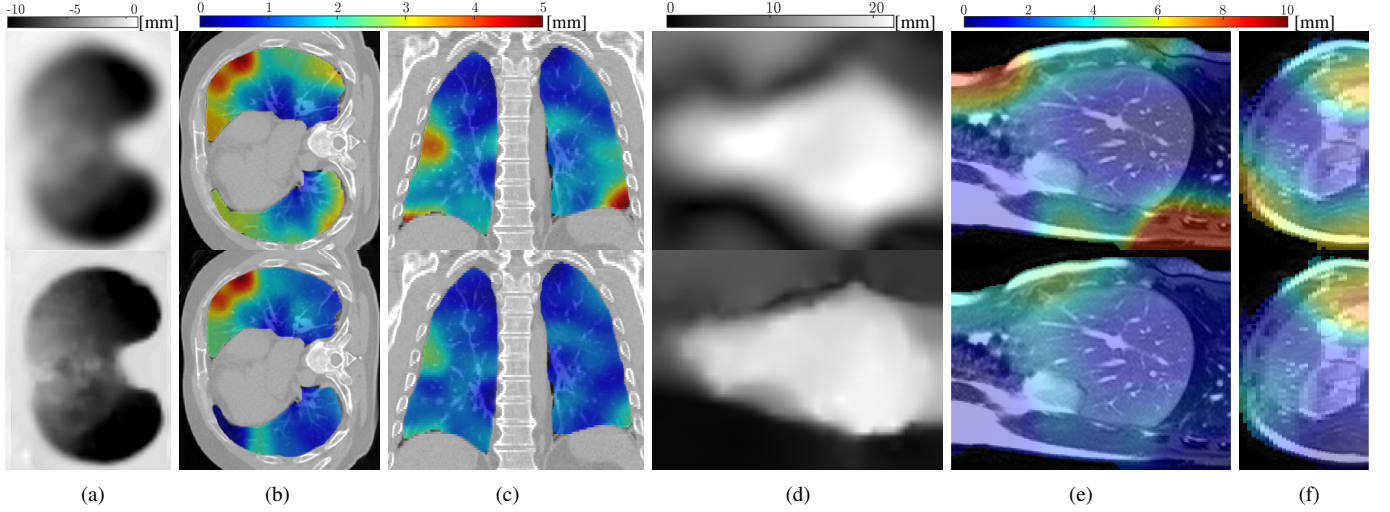|	Method	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean
		TRE [mm] for each 4D CT case											TRE [mm] for each COPDgene case										
	No reg.	4.01	4.65	6.73	9.42	7.10	11.10	11.59	15.16	7.82	7.63	8.52	25.90	21.77	12.29	30.90	30.90	28.32	21.66	25.57	14.84	22.48	23.46
With masks	cEPE [33]	0.80	0.77	0.92	**1.22**	1.21	0.90	0.98	1.16	1.00	0.99	1.00											
	NLR [34]	0.77	0.78	0.93	1.27	1.11	0.91	0.86	**1.03**	0.97	0.87	0.95	1.33	2.34	1.12	1.54	1.39	2.08	1.10	1.57	0.99	1.42	1.49
	LMP [35]	**0.74**	0.78	0.91	1.24	1.17	0.90	0.87	1.04	0.98	0.89	0.95	1.21	1.97	1.06	1.64	1.46	1.34	1.16	1.54	0.99	1.39	1.38
	SGM3D [36]	0.76	**0.72**	0.94	1.24	1.15	0.90	0.89	1.13	**0.91**	0.83	0.95	1.22	2.48	1.01	2.42	1.93	1.45	1.05	1.16	0.81	1.28	1.48
	NGF(a) [4]	0.78	0.79	0.93	1.27	**1.07**	0.90	0.85	**1.03**	0.94	0.86	**0.94**											
	NGF(b) [4]	0.76	0.80	0.96	1.33	1.18	1.03	0.92	1.13	1.00	0.91	1.00											
	LFC [37]	0.85	0.74	0.93	1.33	1.14	1.04	1.03	1.11	1.04	1.05	1.03											
	cTVL1 [19]	0.78	0.78	0.93	1.24	1.22	0.94	1.01	1.11	0.98	0.94	0.99											
	MILO [38]													0.93	1.77	0.99	1.14	1.02	0.99	1.03	1.31	0.86	1.23	1.13
	MRF [3]											0.97	1.00	**1.62**	1.00	1.08	0.96	1.01	1.05	1.08	**0.79**	1.18	1.08
Without masks	CPP [39]	1.07	0.99	1.23	1.51	1.95	1.94	1.79	1.96	1.33	1.84	1.56											
	4DLTM [39]	0.97	0.86	1.01	1.40	1.67	1.58	1.46	1.77	1.19	1.59	1.35											
	cTVL1 [19]	0.79	0.80	1.02	1.23	1.27	1.09	1.87	3.01	1.11	1.17	1.34											
	cEPE [33]	0.81	0.80	0.93	1.25	1.28	1.19	3.03	3.52	1.16	1.52	1.55											
	ALK [40]	0.98	0.83	1.08	1.45	1.55	1.52	1.29	1.75	1.22	1.47	1.31											
	aTV [22]	0.76	0.78	0.82	1.31	1.25	1.11	0.97	1.28	1.04	0.99	1.03											
	pTV	0.76	0.77	0.90	1.24	1.12	**0.85**	**0.80**	1.34	0.92	0.82	0.95	0.77	2.22	0.82	0.85	0.77	0.86	0.74	0.81	0.83	0.92	**0.96**

Fig. 5: Motion magnitude and interpolated TRE maps for lung 4D CT (a-c) and liver MR (d-f) images. The top row depicts the results of smoothness regularized (E_R^{sm}) registration, bottom row shows the results of the proposed pTV method. (a) axial slice of inferior-superior component of the displacement field; (b) axial and (c) coronal CT image slices with overlaid TRE map; (d) sagittal slice of the displacement magnitude map; (e) sagittal and (f) axial MR liver slices with overlaid TRE map. TRE maps are computed by using the Nadaraya-Watson model with a Gaussian kernel (σ=12 mm).

previous experiments, landmark registration error maps in Figs. 5 (e,f) indicate registration quality improvement around expected sliding interface (depicted in Fig. 6 (c)) computed using Eq. (20). Fig. 7 shows an example displacement field.

Comparing with the 4D CT data, this MR database has higher image voxel anisotropy, more image artifacts, and relatively large motion magnitudes; making the registration challenging. On the other hand, liver images have smaller image sizes and hence grid sizes, which decreases dimensionality of the optimization problem. This improves optimization behavior and allows for a more exhaustive search.

E. Parameter sensitivity

As was mentioned earlier, a single set of parameters was used for all the experiments. To show the influence of the LCC weighting kernel w, we tested our method with 10 evenly
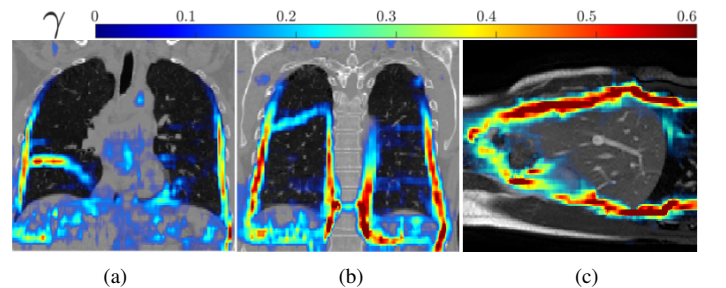

Fig. 6: Sliding amount $\gamma[\mathbf{x}]$ estimated for the displacement fields produced by pTV. Result for (a) dorsal and (b) ventral coronal slices of the 4D CT case # 4, (c) sagittal slice of the 4D liver MRI.

spaced w values in [0.5, 4.5] and 10 evenly spaced λ values in [0, 0.5] for all cases of the 4D MR liver database. The resulting
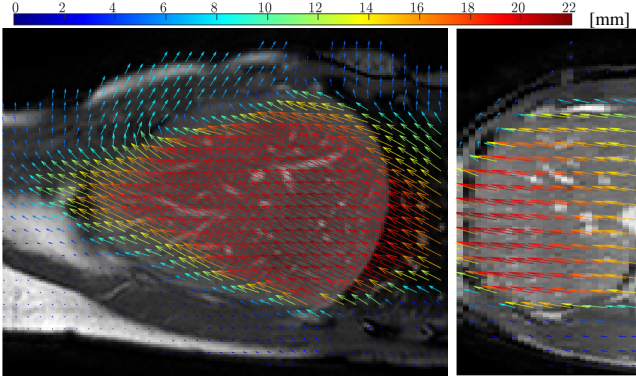
Fig. 7: Sagittal and axial slice of the displacement field estimated by pTV for the 4D liver MRI, as color-coded by displacement magnitude.
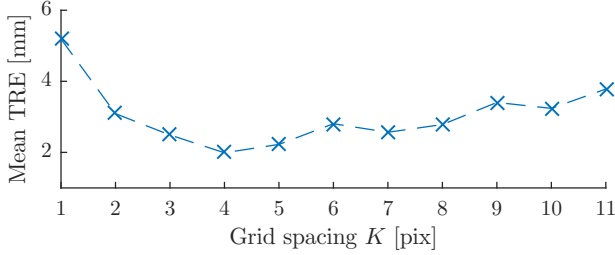

Fig. 8: Sensitivity of TRE to grid spacing of pTV method on the MRI liver database.

mean TRE versus w plots did not have any prominent minima. The TRE's standard deviation averaged over all cases and λ values was 0.13 mm. Hence, we fixed $w = 2.5$ pixels.

Fig. 9 shows the TRE dependence on λ for the different databases. Results show that for breathing motion all tested databases of CT or MR images have optimal regularization values $\lambda^* \approx 0.1$. Particularly interesting is the TRE behavior of the different 4D MR liver sequences, which consists of 4 image pairs with relatively small motion (5–8 mm) and 4 image pairs with larger motion (14–17 mm). As can be seen from Fig. 9, pTV is relatively insensitive to λ for the cases with small motion. However, for the cases with prominent motion, setting λ to large values leads to over-regularization, which results in poorer registration accuracy. We did not investigate adjusting λ to the image noise level as described in Sec. II-F.

Considering these results, we propose the following practical tuning scheme: (i) fix $w = 2.5$ pixels; (ii) select from the database one pair of images with large motion and annotate it; (iii) perform grid search for λ^*, which minimizes mean TRE.

To tune control point grid spacing K, one can evaluate registration accuracy with reasonable grid spacings. Usually at very fine grid spacing ($K \in \{1, 2, 3\}$) the optimization search space becomes too large, yielding poor minimization result. The trade-off between model flexibility (grid spacing) and optimization robustness is common in model selection. Registration accuracies for different grid spacings on the 4D MRI liver database are given in Fig. 8. The best mean accuracy was achieved for 4-voxel grid spacing. It is superior (62% TRE reduction) to K=1 spacing, which is approximating the non-parametric registration scenario.
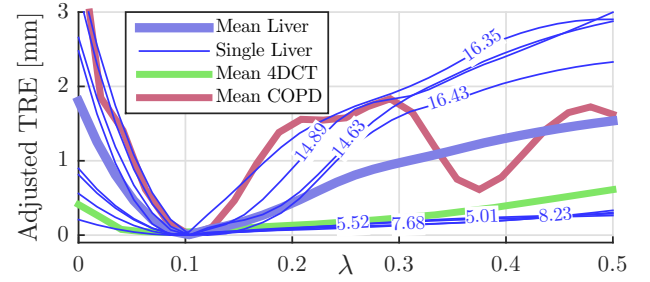

Fig. 9: Sensitivity of TRE to regularization parameter λ of pTV method. For better illustration, each line was shifted vertically (hence adjusted TRE) to align its minimal value with zero. Mean values for different databases are shown with bold lines. For the 4D liver database the adjusted TRE line for each image pair is shown with its mean motion value.

TABLE III: Average mean (max) TRE [mm] for liver MRIs.

Method	Inside the liver	Outside the liver	Overall
No reg.	14.18 (18.58)	5.71 (16.75)	10.98 (19.16)
Demons	3.57 (12.87)	4.25 (14.38)	3.83 (15.77)
Aniso. Demons [10]	3.00 (11.86)	4.05 (13.15)	3.40 (14.03)
aTV [22]	2.42 (11.00)	2.64 (7.52)	2.56 (9.78)
pTV	**1.86 (9.87)**	**2.21** (10.00)	**1.99 (11.87)**

IV. CONCLUSIONS

In this paper we have further developed and validated a parametric approach for image registration with total variation regularization by incorporation of an isotropic TV and efficient LCC image metric. The method was successfully evaluated on the imaging data of different modalities (CT, MR) and anatomies. An advantage of the proposed parametric registration based on TV regularization (pTV) is that it does not rely on a specific anatomy or image modality, and can be used generally for the quantitative analysis of organ sliding motion. The fact that the same algorithm parameters could be used for all four clinical datasets indicates that the method might not need any further tuning in similar scenarios, i.e. breathing motion estimation from CT and MR images. Our method does not use organ masks and provides motion estimates on both sides of the sliding interface, which were similar or more accurate than the results of the methods with masks. On the breath-hold lung CT COPDgene dataset, pTV outperformed all other methods in terms of mean TRE by a minimum margin of 11%, which we see as an important practical result.

From an application point of view, the major limiting factor of pTV is the linear parametrization of the displacement field. Reducing the degrees of freedom helps to better condition the optimization process, although linear interpolation might not always accurately represent sliding motion discontinuities. Alternative solutions for this problem are not straightforward and also have drawbacks: For example, displacement fields parametrized with cubic function basis still cannot model sharp transitions, and the interpolated values are furthermore not bounded by control grid displacements. Wavelet transforms have the drawback that their displacement parametrization properties heavily depend on the choice of the generating functions, and they generally parametrize smooth transitions poorly. Adaptive grids require a local refinement strategy, which implicitly increases the degrees of freedom and slows

down computation due to nonuniform grid spacing. While non-parametric approaches have the flexibility to capture sliding interfaces anywhere, their registration accuracy is often worse since the regularization required to enable sliding appears to be insufficient to make the problem well-posed. A way to overcome the limitations of 1st order B-spline model, could be an overcomplete parametrization of displacement fields with different families (e.g. linear and cubic) of B-splines while enforcing sparsity constraints on the coefficients of those basis functions. A similar approach was introduced in [28] in the context of estimating smooth cardiac motion using cubic B-splines with different grid spacing.

Although no tuning was needed for breathing motion estimation, further research should investigate the sensitivity of the regularization parameter for different motion types, which might make it possible to automatically tune this parameter based on motion patterns.

REFERENCES

[1] M. Burger, J. Modersitzki, and L. Ruthotto, "A hyperelastic regularization energy for image registration," *SIAM Journal on Scientific Computing*, vol. 35, no. 1, pp. B132–B148, 2013.

[2] S. Hu, E. Hoffman, and J. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. on Medical Imaging*, vol. 20, no. 6, pp. 490–8, 2001.

[3] M. Heinrich, H. Handels, and I. Simpson, "Estimating large lung motion in COPD patients by symmetric regularised correspondence fields," in *Proc. Int. Conf. MICCAI*, 2015, vol. 9350, pp. 338–45.

[4] L. Konig and J. Rühaak, "A fast and accurate parallel algorithm for non-linear image registration using normalized gradient fields," in *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, 2014, pp. 580–583.

[5] V. Delmon, S. Rit, R. Pinho, and D. Sarrut, "Registration of sliding objects using direction dependent B-splines decomposition," *Phys. Med. Biol.*, vol. 58, no. 5, pp. 1303–14, 2013.

[6] R. Amelon, K. Cao, J. Reinhardt, G. Christensen, and M. Raghavan, "A measure for characterizing sliding on lung boundaries," *Annals of Biomedical Engineering*, vol. 42, no. 3, pp. 642–650, 2014.

[7] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling," *Phys. Med. Biol.*, vol. 55, no. 1, pp. 305–27, 2010.

[8] L. Risser, F.-X. Vialard, H. Y. Baluwala, and J. A. Schnabel, "Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions," *Medical Image Analysis*, vol. 17, no. 2, pp. 182–193, 2013.

[9] B. Papież, M. Heinrich, J. Fehrenbach, L. Risser, and J. Schnabel, "An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration," *Medical image analysis*, vol. 18, no. 8, pp. 1299–1311, 2014.

[10] C. Tanner, G. Samei, and G. Székely, "Investigating anisotropic diffusion for the registration of abdominal MR images," in *IEEE Int. Symp Biomedical Imaging (ISBI)*, 2013, pp. 484–7.

[11] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.

[12] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2432–9.

[13] S. Kiriyanthan, K. Fundana, and P. C. Cattin, "Discontinuity preserving registration of abdominal MR images with apparent sliding organ motion," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist.Intervent.*, 2012, pp. 231–239.

[14] T. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM Journal on Applied Mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.

[15] T. Pock, M. Urschler, C. Zach, R. Beichel, and H. Bischof, "A duality based algorithm for TV-L1-optical-flow image registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist.Intervent.*, 2007, pp. 511–8.

[16] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through MRFs and efficient linear programming," *Medical Image Analysis*, vol. 12, no. 6, pp. 731–741, 2008.

[17] H. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel, "MRF-based deformable registration and ventilation estimation of lung CT," *IEEE Trans. on Medical Imaging*, vol. 32, no. 7, pp. 1239–1248, 2013.

[18] M. Rajchl, J. S. Baxter, W. Qiu, A. R. Khan, A. Fenster, T. M. Peters, and J. Yuan, "RANCOR: Non-linear image registration with total variation regularization," *arXiv preprint arXiv:1404.2571*, 2014.

[19] S. Hermann and R. Werner, "TV-L1-based 3D medical image registration with the census cost function," in *Img. and Vid. Tech.*, 2014, pp. 149–61.

[20] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Tran. Med. Imag.*, vol. 18, pp. 712–21, 1999.

[21] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. on Image Processing*, vol. 12, no. 11, pp. 1427–1442, 2003.

[22] V. Vishnevskiy, T. Gass, G. Székely, and O. Goksel, "Total variation regularization of displacements in parametric image registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist.Intervent.*, 2014, vol. 8676, pp. 211–220.

[23] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, "Iconic feature based nonrigid registration: the PASHA algorithm," *Computer Vision and Image Understanding*, vol. 89, no. 2–3, pp. 272–298, 2003.

[24] P. Cachier and X. Pennec, "3D non-rigid registration by gradient descent on a Gaussian-windowed similarity measure using convolutions," in *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2000, pp. 182–189.

[25] P. Blomgren and T. Chan, "Color TV: total variation methods for restoration of vector-valued images," *IEEE Trans. on Image Processing*, vol. 7, no. 3, pp. 304–309, 1998.

[26] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[27] M. A. Figueiredo and J. M. Bioucas-Dias, "Algorithms for imaging inverse problems under sparsity regularization," in *IEEE Int. Workshop on Cognitive Information Processing (CIP)*, 2012, pp. 1–6.

[28] W. Shi, M. Jantsch, P. Aljabar, L. Pizarro, W. Bai, H. Wang, D. O'Regan, X. Zhuang, and D. Rueckert, "Temporal sparse free-form deformations," *Medical Image Analysis*, vol. 17, no. 7, pp. 779–789, 2013.

[29] L. A. Schwarz, "Non-rigid registration using free-form deformations," Ph.D. dissertation, Technische Universität München, Germany, 2007.

[30] O. Goksel, V. Vishnevskiy, G. A. Carrillo, and C. Tanner, "Imaging of sliding visceral interfaces during breathing," in *IEEE Int. Symposium on Biomedical Imaging*, 2016.

[31] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable," *IEEE Trans. on Medical Imaging*, vol. 31, no. 2, pp. 153–163, Feb 2012.

[32] J. Vandemeulebroucke, D. Sarrut, P. Clarysse *et al.*, "The POPI-model, a point-validated pixel-based breathing thorax model," in *Int. Conf. Computers in Radiation Therapy (ICCR)*, 2007, pp. 195–9.

[33] S. Hermann and R. Werner, "High accuracy optical flow for 3D medical image registration using the census cost function," in *Image and Video Technology*, 2014, vol. 8333, pp. 23–35.

[34] J. Rühaak, S. Heldmann, T. Kipshagen, and B. Fischer, "Highly accurate fast lung CT registration," vol. 8669, 2013, pp. 86 690Y–86 690Y–9.

[35] T. Polzin, J. Rühaak, R. Werner, J. Strehlow, S. Heldmann, H. Handels, and J. Modersitzki, "Combining automatic landmark detection and variational methods for lung CT registration," in *Fifth Int. Workshop on Pulmonary Image Analysis*, 2013.

[36] S. Hermann, "Evaluation of scan-line optimization for 3D medical image registration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3073–3080.

[37] E. Castillo, R. Castillo, B. White, J. Rojo, and T. Guerrero, "Least median of squares filtering of locally optimal point matches for compressible flow image registration," *Physics in Medicine and Biology*, vol. 57, no. 15, p. 4827, 2012.

[38] E. Castillo, R. Castillo, D. Fuentes, and T. Guerrero, "Computing global minimizers to a constrained B-spline image registration problem from optimal ℓ_1 perturbations to block match data," *Med. Phys.*, vol. 41, 2014.

[39] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling," *Phys. in Med. and Biol.*, vol. 55, no. 1, p. 305, 2010.

[40] C. Hoog, Antink, T. Singh, P. Singla, and M. Podgorsak, "Evaluation of advanced LukasKanade optical flow on thoracic 4D-CT," *Journal of Clinical Monitoring and Computing*, vol. 27, no. 4, pp. 433–441, 2013.

[41] R. Castillo, E. Castillo, D. Fuentes, M. Ahmad, A. Wood, M. Ludwig, and T. Guerrero, "A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive," *Phys. in Med. and Biol.*, vol. 58, no. 9, p. 2861, 2013.