Noname manuscript No.
(will be inserted by the editor)

Graph-based slice-to-volume deformable registration

Enzo Ferrante[1,2] · **Nikos Paragios**[1,3]

Received: date / Accepted: date

Abstract Deformable image registration is a fundamental problem in computer vision and medical image computing. In this paper we investigate the use of graphical models in the context of a particular type of image registration problem, known as slice-to-volume registration. We introduce a scalable, modular and flexible formulation that can accommodate low-rank and high order terms, that simultaneously selects the plane and estimates the in-plane deformation through a single shot optimization approach. The proposed framework is instantiated into different variants seeking either a compromise between computational efficiency (soft plane selection constraints and approximate definition of the data similarity terms through pair-wise components) or exact definition of the data terms and the constraints on the plane selection. Simulated and real-data in the context of ultrasound and magnetic resonance registration (where both framework instantiations as well as different optimization strategies are considered) demonstrate the potentials of our method.

Keywords Slice-to-volume registration · graphical models · deformable registration · discrete optimization

[1] Center for Visual Computing, CentraleSupelec, INRIA, Universite Paris-Saclay, (Paris, France)
[2] Biomedical Image Analysis (BioMedIA) Group, Department of Computing, Imperial College London, (London, UK)
[3] TheraPanacea (Paris, France)
E-mail: ferrante.enzo@gmail.com

1 Introduction

Slice-to-volume deformable registration is an important problem in the communities of computer vision and medical image computing, which has received considerable attention during the last decade. In general terms, it seeks to determine the slice (corresponding to an arbitrary plane) from a given target volume that corresponds to the deformed version of a source 2D image. This slice is generally specified by a rigid transformation $\hat{T}$. The source 2D image is deformed by a deformation field $\hat{D}$ towards improving the matching consistency between the deformed source image and the target slice.

Slice-to-volume registration is sometimes referred as 2D/3D registration, primarily due to dimension of the images involved in the registration process. Note that this term describes two different problems depending on the technology used to capture the 2D image: it might be a projective (e.g. x-ray) or sliced (e.g. ultrasound (US)) image. In this work we only focus on the latter case. Projective images have to be treated in a different way (basically a pixel in the 2D image does not correspond only to a voxel from the target volume, but to a projection of a set of them in certain perspective) and they are out of the scope of this paper. This is principally due to the fact that conventional image similarity terms cannot be used in the projective case. However, it should be noted that the proposed formulation with an appropriate definition of the matching and regularization cost could also accommodate a solution to this problem. We refer

Images in this work are better viewed in color.

the reader to the comprehensive survey by [29] for further information about this topic.

1.1 Motivation

A broad number of medical image computing applications benefit from slice-to-volume registration. One can cite, for example, image guided surgeries and therapies [7], biopsies [46], tracking of particular organs [12] and minimally-invasive procedures [28,17]. In such a context, slice-to-volume registration is a key element for bringing high resolution annotated data into the operating room. Generally, pre-operative 3D images such as computed tomography (CT) or magnetic resonance images (MRI) are acquired for diagnosis and manually annotated by expert physicians prior to the operation. During the procedure, 2D real time images are generated using different technologies (e.g. fluoroCT, US or interventional MRI slices). These intra-operative images refer to challenging acquisition constraints and inherit lower resolution and quality than the pre-operative ones. Moreover, tissue shift collapse as well as breathing and heart motion during the procedure, causes elastic deformation in the images. Non-rigid image registration is suitable to address this issue. The alignment of intra-operative images with pre-operative volumes augments the information that doctors have access to, and allows them to navigate the volumetric annotation while performing the operation.

Another interesting application is motion correction for image reconstruction. Here, the goal is to correct for misaligned slices when reconstructing a volume of a certain modality. A typical approach to solve it consists of mapping individual slices within a volume onto another reference volume in order to correct the inter-slice misalignment. The popular map-slice-to-volume (MSV) method that introduced this idea in the context of functional MRI (fMRI) was presented by [20]. More recently, applications of slice-to-volume registration to the same problem in different contexts like cardiac magnetic resonance (CMR) [5], fetal images [41] and diffusion tensor imaging (DTI) [18] have shown promising results.

Although the goals of the motivating problems we have described are different, all of them require to perform (to some extent) slice-to-volume registration. In this work, we focus on the applications where we need to navigate a pre-operative volume using intra-operative images. However, the method we present is modular enough to be adapted to different image modalities and settings, and therefore can be applied to any of these problems.

1.2 Previous work

Several methods have been proposed during the recent years to deal with slice-to-volume registration. Some of them deal only with rigid registration, and therefore they cannot manage deformations due to tissue shift, breathing or heart motion. [40], for example, proposed a method to register endoscopic and laparoscopic ultrasound images with pre-operative computed tomography volumes that potentially could work in real time. It is based on a new phase correlation technique called LEPART and it handles rigid registration. [12] tracks intra-operative MRI slices of prostate images with a pre-operative MRI volume. This monomodal registration is designed to provide patient tracking information for prostate biopsy performed under MR guidance, but is also constrained to rigid transformations. More recently, [6] proposed a method that uses smart phone as a navigation tool for initial slice alignment followed by an overlap invariant mutual information-based refinement that estimates the rigid transformation.

Other methods tackle the challenging problem of non-rigid slice-to-volume registration using non-linear models. Among these, there is a sub-category of approaches that uses several slices instead of a single one, in order to improve the quality of the results. Some examples are [33] which uses a variational approach and [46] who designed a two-step algorithm where initial rigid registration is followed by B-spline based deformable registration. Using several slices restricts the potential applications to the ones where more than one slice is available from the beginning. It also simplifies the problem by increasing the amount of available information. Our method performs slice-to-volume registration using a single input slice. Consequently, it can be adapted to a broader range of applications where just one slice is available at a time. We refer the reader to [11] for a complete survey about alternative slice-to-volume registration methods proposed in the literature of medical image registration.

Most of the aforementioned slice-to-volume registration approaches, rely on continuous methods to model and perform parameter estimation. In this paper we extend our previous work presented in [10,9,8] through the introduction of a single,

mathematically rigorous and theoretically sound framework derived as a discrete labeling problem on a graphical model. Graphical models and discrete optimization are powerful formalisms that have been successfully used during the past years in the field of computer vision [45]. In particular, rigid as well as non-rigid image registration have been formulated as a minimal cost graph problem where the nodes of the graph correspond to the deformation grid and the graph connectivity encodes regularization constraints. However, this technique has been applied mainly to mono-dimensional cases (2D-2D or 3D-3D). To the best of our knowledge, the only work that focuses on multi-dimensional image registration (apart of our previous articles that have been referenced at the beginning of this paragraph) using this type of techniques is [47]. However, it estimates only rigid transformations and works with projective images.

Discrete methods have several advantages when compared with continuous approaches for slice-to-volume registration. First, discrete algorithms are inherently gradient-free, while most part of continuous methods require the objective function to be differentiable. Gradient-free methods do not require computation of the energy derivative. Therefore, it may be applied to any complex energy function (allowing the user to define its own similarity measures in case of registration problems). The only requirement is that this function must be evaluable in a variety of possible discrete labelings. Second, most part of the continuous methods are prone to be stuck in local minima when the functions are not convex. In case of discrete methods, even complicated functions could potentially be optimized using large neighbor search methods. The main limitation is the discretization of the continuous space; however, as suggested by [13], '*the optimality is bounded by the discretization, but with intelligent refinement strategy the accuracy of continuous methods can be achieved*'. Third, parallel architectures can be used to perform non-sequential tasks required by several discrete algorithms leading to more efficient implementations. Fourth, by using a discrete label space we can explicitly control its range and resolution (it can be useful to introduce prior information, as it will be shown in this work), while in continuous models it is not clear how this type of information can be used to constraint the solution. Last but not least, discrete frameworks such as discrete MRF provide a modular and principled way to combine prior knowledge with data likelihood (through the energy formulation), what makes it applicable to a wide range of vision tasks [45], particularly, to the challenging slice-to-volume registration problem.

1.3 Contribution

This article contributes to enrich the standard graph-based deformable registration theory by extending it to the case of slice-to-volume registration. We present three different models to solve this challenging problem which vary in terms of graph topology, label space definition and energy construction. Our aim is to demonstrate how flexible and powerful the graph theory is in terms of expressive potential of the modeling process, while solving a new problem using graphical models. We analyze the strong and weak points of every model and we perform comparative experiments. Validation is done using a monomodal MRI cardiac dataset and a multimodal brain dataset [30] including different inference methods.

2 Graph-based slice-to-volume deformable registration

An enormous variety of tasks in computer vision and medical image analysis can be expressed as discrete labeling problems [35]. Low, mid and high-level vision tasks can be addressed within this framework. To this end, a visual perception task is addressed by specifying a task-specific parametric model, associating it to the available observations (images) through an objective function and optimizing the model parameters given both, the objective and the observations [36].

In the context of graph-based discrete labeling problems, the model is composed by a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ where vertices in $\mathcal{V}$ correspond to the variables while $\mathcal{E}$ is a neighborhood system (pair-wise & higher order cliques) that encodes the relationships among these variables. We also consider a discretized version of the search space that is represented by a discrete set of labels $l \in L$. The aim is to assign to every variable $v \in \mathcal{V}$ a label $l_v \in L$. Each time we choose to assign a label, say, l_{v_1} to a variable v_1, we are forced to pay a price according to the so-called energy function. This objective function is domain-specific and associates the observations to the model. It is formulated as the sum of singleton terms $g_v(l_v)$ (which depend only on one label l_v), pairwise terms $f_{v_1 v_2}(l_{v_1}, l_{v_2})$ (which depend on two variables l_{v_1}, l_{v_2}) and high-order

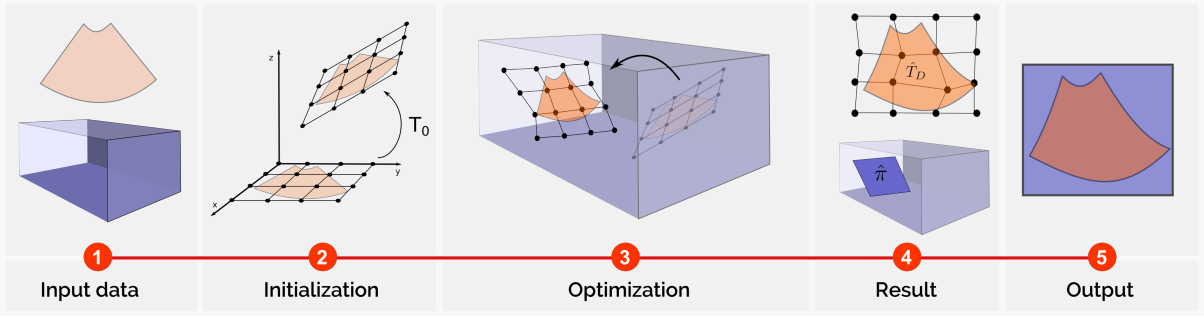


Figure 1 Basic workflow to perform slice-to-volume registration based on graphical models. (1) A 2D input image I and a 3D target volume J are given as input data. (2) A grid is superimposed to image I. The process is initialized using a 6-DOF rigid transformation T_0 that specifies the initial position of the grid within the volume J. (3) The grid is deformed by optimizing an energy function. (4) The plane $\hat{\pi}$ and the deformation field $\hat{T}_D$ are reconstructed from the final state of the optimized grid. (5) $\hat{T}_D$ is used to deform image I, and it is provided as output together with the corresponding slice $\hat{\pi}[J]$

terms $f_{v_1 \dots v_n}(l_{v_{i1}}, \dots, l_{v_i^{|C_i|}})$ (which are associated to high-order cliques C_i that depend on more than two variables). Our goal is then to choose a labeling which will allow us to recover the solution corresponding to the minimal value of the objective function. In other words, we want to choose a labeling that minimizes the sum of all the energy potentials, or equivalently the energy $\mathcal{P}(g, f)$. This amounts to solving the following optimization problem:

$$\underset{l_p}{\operatorname{argmin}}\, \mathcal{P}(g,f) = \sum_{v \in \mathcal{V}} g_p(l_v) + \sum_{(v_1,v_2)\in\mathcal{E}} f_{v_1 v_2}(l_{v_1}, l_{v_2}) + \sum_{C_i \in \mathcal{E}} f_{v_1 \dots v_n}(l_{v_{i1}}, \dots, l_{v_i^{|C_i|}}), \tag{1}$$

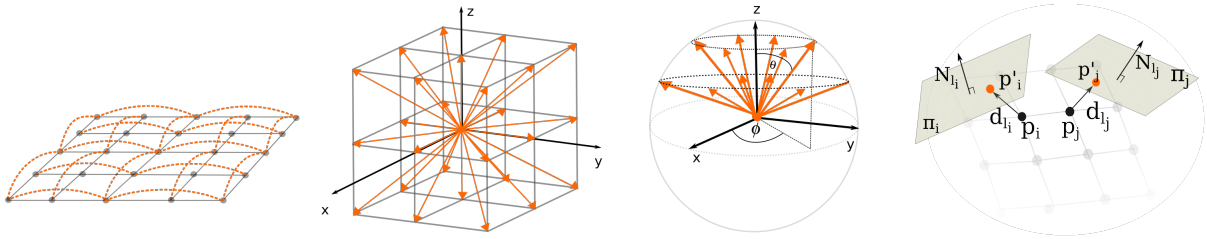
Performing parameter inference on this graphical model, could be an effective solution to a big variety of problems in computational medicine. Note that we make a distinction between singleton, pairwise and high-order terms, depending on the number of variables jointly interacting. It should be noted that most part of the graph-based vision models have explored mainly pairwise constraints (pairwise Conditional and Markov Random Field (CRF/MRF) models), because in these cases exact or approximate efficient inference of Maximum a Posteriori (MAP) solutions can be done. However, during the last few years, more and more high-order models and inference algorithms have been developed which offer higher modeling power and can lead to more accurate solutions of the problems [21,22]. Given such a general setting, let us now try to explore the expressive power of such models in the context of slice-to-volume deformable registration.

The task of slice-to-volume deformable registration can be expressed mathematically as follows. Given a 2D source image I and a 3D target volume J, we seek the 2D-2D in-plane local deformation field $\hat{T}_D$ and the plane $\hat{\pi}[J]$ (i.e. a bi-dimensional slice from the volume J) which in the most general case minimize the following objective function:

$$\hat{T}_D, \hat{\pi} = \underset{T_D, \pi}{\operatorname{argmin}}\, \mathcal{M}(I \circ T_D(\boldsymbol{x}), \pi[J](\boldsymbol{x})) + \mathcal{R}(T_D, \pi), \tag{2}$$

where $\mathcal{M}$ represents the data similarity term and $\mathcal{R}$ the regularization term. The data term $\mathcal{M}$ measures the matching quality between the deformed 2D source image and the corresponding 3D slice. The regularization term $\mathcal{R}$ imposes certain constraints on the solution that can be used to render the problem well posed. It also imposes certain expected geometric properties on the extended (plane selection and plane deformation) deformation field. The plane $\hat{\pi}$, that minimizes the equation, indicates the location of the 3D volume slice that best matches the deformed source image. The deformation field $\hat{T}_D$ represents the in-plane deformations that must be applied to the source image in order to minimize the energy function.

The fundamental idea behind our approaches is quite intuitive: we aim at deforming a planar 2D grid in the 3D space, which encodes both the deformation field $\hat{T}_D$ and the plane $\hat{\pi}$ at the same time. This grid is super-imposed to the 2D source image and consists of control points that jointly represent the in-plane deformation and the current position of the 2D image into the 3D volume.



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

Figure 2 (a) Connectivity structure of the graph for a grid of size 5x5. The gray edges are standard 4-neighbor connections while the orange ones correspond to the extra cliques introduced to improve the geometrical constraints propagation. (b) Displacement vectors corresponding to the first three elements of a label from the overparameterized approach $\boldsymbol{d_i} = (d_x, d_y, d_z)$. (c) Unit vectors in spherical coordinates corresponding to the last two coordinates of a label from the overparameterized approach $\boldsymbol{N_i} = (\phi, \theta)$. (d) Displacement of the control points $\boldsymbol{p_i}$ and $\boldsymbol{p_j}$ when the corresponding labels $\boldsymbol{l_i} = (\boldsymbol{d_i}, \boldsymbol{N_i})$ and $\boldsymbol{l_j} = (\boldsymbol{d_j}, \boldsymbol{N_j})$ are applied. The planes π_i and π_j are those that contain the control points $\boldsymbol{p_i} + \boldsymbol{d_i}, \boldsymbol{p_j} + \boldsymbol{d_j}$ and whose normals are $\boldsymbol{N_i}, \boldsymbol{N_j}$ respectively.

The source image is positioned within the volume by applying different displacement vectors with respect to the control points of the superimposed grid. These displacements are chosen such that a given energy (see Eq. 2) is minimized to best fit the matching criterion $\mathcal{M}$. Since they can be moved without any restriction, geometric constraints are imposed through the regularization term $\mathcal{R}$ in order to keep a smooth deformation field and a planar grid. Given that we impose a soft planar constraint, the resulting grid is approximately planar. Therefore, we reconstruct the final solution by projecting all the points into a regression plane which is estimated out of the current position of the points. The rigid transformation that indicates the position of the regression plane is considered as $\hat{\pi}$. Finally, the projected grid is interpreted as a 2D Free Form Deformation model (FFD) [39] where each control point has local influence on the deformation and is used to approximate the dense deformation field $\hat{T}_D$ (other control point interpolation models could be used as well). Alternatively, depending on the application, one may prefer to deform the sliced image $\pi[J]$ instead of the source image I. Note that this can be done by simply using the inverse of the deformation field T_D. To guarantee the existence of the inverse, we can restrict the generated deformation fields to be diffeomorphic. This can be easily guaranteed in our framework by restricting the displacements size to 0.4 times the size of the current grid, as indicated in [16]. Figure 1 illustrates the complete workflow described in this paragraph.

In this work, we restrict the geometry of the final solution to in-plane deformations only (i.e. 2D deformations acting only in the plane π). As explained in the previous paragraph, we do that by projecting the final position of the control points into a regression plane estimated out of the current position of those points. We follow this strategy since we found that it improves the stability of the method by restricting the solution space to only 2D deformation fields. However, considering out-of-plane deformations in the proposed framework would only require sidestepping the control points projection step. In our current formulation, since we allow the control points to move freely within the 3D space, the grid is actually deformed in 3D. Actually, the regularization terms imposing plane consistency are soft constraints which can be violated if the data term indicates large matching values. Indeed, they are commonly violated; otherwise we would not require a projection step. Consequently, avoiding the step where we project every control point to the regression plane and interpreting the deformed 2D grid as a 3D deformation field, would be enough to incorporate out-of-plane deformations if required.

This general formulation can be expressed through different discrete labeling problems on a graph by changing its topology, the label space definition and the energy terms. As we mentioned, in this work we propose three different approaches to derive slice-to-volume registration as a discrete graph labeling problem. First, we propose the so-called *overparameterized* method, which combines linear and deformable parameters within a coupled formulation on a 5-dimensional label space [10]. The main advantage of such a model is the simplicity provided by its pairwise structure, while the main disadvantage is the dimensionality of the label space which makes inference computationally inefficient and approximate (limited sampling of search space). Motivated by the work of [42], we present a *de-*

coupled model where linear and deformable parameters are separated into two interconnected subgraphs which refer to lower dimensional label spaces [9]. It allows us to reduce the dimensionality of the label space by increasing the number of edges and vertices, while keeping a pairwise graph. Finally, in the *high-order* approach [8], we achieve this dimensionality reduction by augmenting the order of the graphical model, using third-order cliques which exploits the expression power of this type of variable interactions. Such a model provides better satisfaction of the global deformation constraints at the expense of quite challenging inference.

2.1 Overparameterized approach

Let us consider an undirected pair-wise graph $G_O = \langle V, E \rangle$ super-imposed to the 2D image domain with a set of nodes V and a set of cliques E. The nodes V (a regular lattice) are interpreted as control points of the bi-dimensional quasi-planar grid that we defined in the previous section. The set of edges E is formed by regular 4-neighbors grid connections and some extra edges introduced to improve the propagation of the geometrical constraints (see Figure 2.a). The vertices $v_i \in V$ are moved by assigning them different labels $u_i \in L$ (where L corresponds to the label space) until an optimal position is found.

In order to deform the graph, we need to define a label space able to describe the inplane deformations and the plane selection variables. To this end, we consider a label space L that consists of 5-tuples $\boldsymbol{l} = (d_x, d_y, d_z, \phi, \theta)$, where the first three parameters (d_x, d_y, d_z) define a displacement vector $\boldsymbol{d_i}$ in the cartesian coordinate system (see Figure 2.b), and the angles (ϕ, θ) define a vector $\boldsymbol{N_i}$ on a unit sphere, expressed using spherical coordinates (see Figure 2.c). Let us say we have a control point $\boldsymbol{p_i} = (p_{xi}, p_{yi}, p_{zi})$ and we assign the label $\boldsymbol{l_i} = (d_{xi}, d_{yi}, d_{zi}, \phi_i, \theta_i)$ to this point. So, the new point position $\boldsymbol{p'_i}$ after assigning the label is calculated using the displacement vector as given by the following equation:

$$\boldsymbol{p'_i} = (\boldsymbol{p_{xi}} + \boldsymbol{d_{xi}}, \boldsymbol{p_{yi}} + \boldsymbol{d_{yi}}, \boldsymbol{p_{zi}} + \boldsymbol{d_{zi}}). \quad (3)$$

Additionally, we define a plane π_i containing the displaced control point $\boldsymbol{p'_i}$ and whose unit normal vector (expressed in spherical coordinates and with constant radius $r = 1$) is $\boldsymbol{N_i} = (\phi_i, \theta_i)$. One of the most important constraints to be considered is that our transformed graph should have a quasi-planar structure, i.e. it should be similar to a plane; the plane π_i associated with every control point $\boldsymbol{p_i}$ is used by the energy term to take into account this constraint. Figure 2.d shows how to interpret the labels for two given points $\boldsymbol{p_i}$ and $\boldsymbol{p_j}$.

The energy to be optimized is formed by data terms $G = \{g_i(\cdot)\}$ (or unary potentials) associated with each graph vertex and regularization terms $F = \{f_{ij}(\cdot, \cdot)\}$ (or pairwise potentials) associated with the edges. As we described in section 2, the first ones are typically used for encoding some sort of data likelihood, whereas the later ones act as regularizers and thus play an important role in obtaining high-quality results [16]. The minimization energy problem for the overparameterized formulaton is thus defined as:

$$\mathcal{P}_{\mathcal{O}}(G, F) = \min \sum_{i \in V} g_i(\boldsymbol{l_i}) + \gamma \sum_{(i,j) \in E} f_{ij}(\boldsymbol{l_i}, \boldsymbol{l_j}), \quad (4)$$

where $l_i, l_j \in L$ are the labels assigned to the vertices $v_i, v_j \in V$ respectively.

The formulation of the unary potentials that we propose is independent of the similarity measure. It is calculated for each control point given any intensity based metric δ capable of measuring the similarity between two bi-dimensional images (e.g sum of absolute differences, mutual information, normalized cross correlation). This calculation is done for each control point $\boldsymbol{p_i}$, using its associated plane π_i in the target image J and the source 2D image I. An oriented patch Ω_i over the plane π_i (centered at $\boldsymbol{p_i}$) is extracted from the volume J, so that the metric δ can be calculated between that patch and the corresponding area from the source 2D image (see Figure 3). Please note that this patch will be sampled from the 3D image, given the current position of the control point $\mathbf{p_i}$. Since a single point is not enough to define a unique patch, we refer to the "patch Ω_i over the plane π_k" to stress the fact that this patch will be sampled from the area surrounding the point $\mathbf{p_i}$, only considering those points living in the plane π_i defined by the normal vector $\boldsymbol{N_i}$. The unary potential is then defined as:

$$g_i(\boldsymbol{l_i}) = \int_{\Omega_i} \delta(I \circ T_D(\boldsymbol{x}), \pi_i[J](\boldsymbol{x})) d\boldsymbol{x}. \quad (5)$$

One of the simplest and commonly used similarity measures is the Sum of Absolute Differences (SAD) of the pixel intensity values. It is useful in the monomodal scenario, where two images of the same modality are compared and, therefore, the grey intensity level itself is discriminant enough to

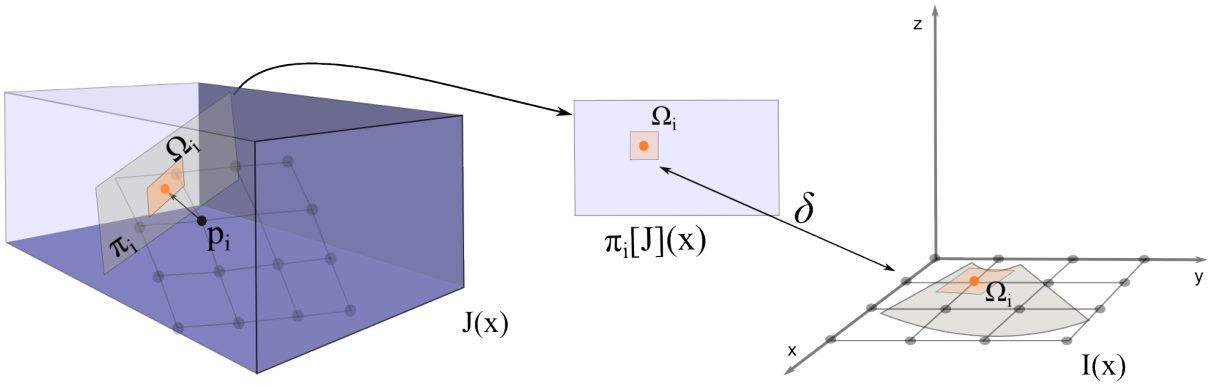


Figure 3 Data term formulation for the overparameterized approach. The points $\boldsymbol{x} \in \Omega_i$ are used to calculate the unary potential. $\pi[J](\boldsymbol{x})$ returns the intensity of the point in the 2D slice corresponding to the plane π_i in the 3D image, whereas $I(\boldsymbol{x})$ returns the 2D image intensity. δ represents the similarity measure.

determine how related are the two images. Its formulation in our framework is:

$$g_{SAD,i}(\boldsymbol{l_i}) = \int_{\Omega_i} \mid I \circ T_D(\boldsymbol{x}) - \pi_i[J](\boldsymbol{x}) \mid d\boldsymbol{x}. \quad (6)$$

In multimodal scenarios, where different modalities are compared (e.g. CT with Ultrasound images), statistical similarity measures such as Mutual Information (MI) are generally used since we can not assume that corresponding objects have the same intensities in the two images. MI is defined using the joint intensity distribution $p(i,j)$ and the marginal intensity distribution $p(i)$ and $p(j)$ of the images as:

$$g_{MI,i}(\boldsymbol{l_i}) = -\int_{\Omega_i} \log \frac{p(I \circ T_D(\boldsymbol{x}), \pi_i[J](\boldsymbol{x}))}{p(I \circ T_D(\boldsymbol{x}))p(\pi_i[J](\boldsymbol{x})))} d\boldsymbol{x}. \quad (7)$$

As we can see in the previous examples, our framework can encode any local similarity measure defined over two two-dimensional images. Please note that by *local* similarity measure we stress the fact that the metric is computed locally around the control point, as opposed to global similarity measures which are computed using the complete image.

Let us now proceed with the definition of the regularization term. Generally, these terms are used to impose smoothness on the displacement field. In our formulation, the pairwise potentials are defined using a linear combination of two terms: the first (F_1) controls the grid deformation assuming that it is a plane, whereas the second (F_2) maintains the plane structure of the mesh. They are weighted by a coefficient α as indicates the following equation:

$$f_{ij}(\boldsymbol{l_i}, \boldsymbol{l_j}) = \alpha F_{1i,j}(\boldsymbol{l_i}, \boldsymbol{l_j}) + (1-\alpha) F_{2i,j}(\boldsymbol{l_i}, \boldsymbol{l_j}). \quad (8)$$

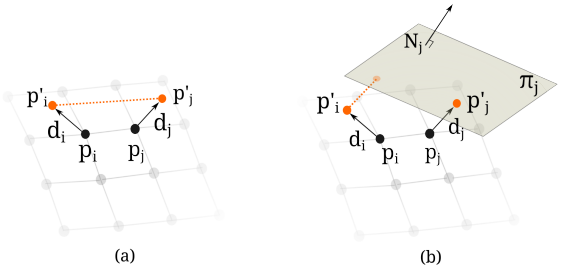


Figure 4 (a) In plane regularization term: the dotted line represents the distance used in F_1, i.e. the distance between the points assuming they are coplanar. (b) Plane structure regularization term: the dotted line represents the distance between one of the control points and the plane corresponding to the other one. This information is used to compute the term F_2.

The in-plane deformation is controlled using a distance preserving approach: it tries to preserve the original distance between the control points of the grid. Since this metric is based on the euclidean distance between the points, it assumes that they are coplanar. We use a distance based on the ratio between the current position of the control points $\boldsymbol{p_i}, \boldsymbol{p_j}$ and their original position $\boldsymbol{p_{o,i}}, \boldsymbol{p_{o,j}}$:

$$\psi_{i,j}(\boldsymbol{d_i}, \boldsymbol{d_j}) = \frac{\mid\mid (\boldsymbol{p_i} + \boldsymbol{d_i}) - (\boldsymbol{p_j} + \boldsymbol{d_j}) \mid\mid}{\mid\mid (\boldsymbol{p_{o,i}}) - (\boldsymbol{p_{o,j}}) \mid\mid}. \quad (9)$$

Once we have defined ψ_{ij}, the regularizer should fulfill two conditions: (i) it has to be symmetric with respect to the displacement of the points, i.e. it must penalize equally whenever the control points are closer or more distant; (ii) the energy has to be zero when the points are preserving distances and monotonically increasing with respect to the violation of the constraint. The following regularization term fulfills both conditions for a

couple of nodes $i, j \in V$ labeled with labels $\boldsymbol{l_i}, \boldsymbol{l_j}$:

$$F_{1i,j}(\boldsymbol{l_i}, \boldsymbol{l_j}) = (1-\psi_{i,j}(\boldsymbol{d_i}, \boldsymbol{d_j}))^2 + (1-\psi_{i,j}(\boldsymbol{d_i}, \boldsymbol{d_j})^{-1})^2, \quad (10)$$

The plane preservation term is based on the average distance between a given control point and the plane defined from the neighboring ones (see Figure 4.b). The aim is to maintain the quasi-planar structure of the grid. Given that the distance between a point and a plane is zero when the point lies on the plane, this term will be minimum when the control points for which we are calculating the pairwise potential are on the same plane.

The distance between a point $\boldsymbol{p} = (p_x, p_y, p_z)$ and a plane π defined by the normal vector $\boldsymbol{N} = (n_x, n_y, n_z)$ and the point $\boldsymbol{q} = (q_x, q_y, q_z)$ is calculated as:

$$D_\pi(\boldsymbol{p}) = \frac{\mid n_x(p_x - q_x) + n_y(p_y - q_y) + n_z(p_z - q_z) \mid}{\sqrt{n_x^2 + n_y^2 + n_z^2}}. \quad (11)$$

F_2 is defined using this distance (Equation 11) and corresponds to the average of $D_{\pi_j}(\boldsymbol{p_i}+\boldsymbol{d_i})$ and $D_{\pi_i}(\boldsymbol{p_j} + \boldsymbol{d_j})$:

$$F_{2i,j}(\boldsymbol{l_i}, \boldsymbol{l_j}) = \frac{1}{2}(D_{\pi_j}(\boldsymbol{p_i}+\boldsymbol{d_i}) + D_{\pi_i}(\boldsymbol{p_j}+\boldsymbol{d_j})). \quad (12)$$

Recall that normal vectors in our label space are expressed using spherical coordinates with a fixed radius $r = 1$ (unit sphere). However, the formulation that we presented uses cartesian coordinates. Therefore, the mapping from one space to another is done as follows:

$$x = r\sin(\theta)\cos(\phi), y = s\sin(\theta)\sin(\phi), z = r\cos(\theta). \quad (13)$$

Note that such pairwise terms are non submodular since we include the current position of the points (which can be arbitrary) in their formulation and therefore the submodularity constraint is not fulfilled. In this context, even if there is no energy bounding that guarantees certain quality for the solution of the optimization problem, good empirical solutions are feasible since we are in a pairwise scenario. Still, two issues do arise: (i) high dimensionality of the label space and consequently high computational cost, (ii) insufficient sampling of the search space and therefore suboptimal solutions. In order to address these issues while maintaining the pairwise nature of the methods, we propose the *decoupled* method inspired by [42]. We consider decoupling the label space into two different ones and redefining the topology of the graph, so that we can still capture rigid plane displacements and in-plane deformation.

2.2 Decoupled approach

We propose to overcome the limitations of the over-parameterized method by decoupling every node of the previous approach into two different ones: one modeling the in-plane deformation and another the position of the plane. This is somewhat analogous to creating two separated graphs of the same size and topology corresponding to different random variables and label spaces. Once spaces have been decoupled, different sampling strategies can be used for them. Another advantage of this approach is that we can define distinct regularization terms for edges connecting deformation nodes or plane position nodes. It allows to regularize in a different way the deformation and the plane position, imposing alternative geometrical constraints for every case.

Since data term computation requires the exact location of the node, both position and deformation labels are necessary. Both graphs can thus be connected through a pairwise edge between every pair of corresponding nodes. Therefore, new pairwise potentials are associated with these edges in order to encode the matching measure.

Formally, the decoupled formulation consists of an undirected pair-wise graph $G_D = \langle V, E \rangle$ with a set of nodes $V = V_I \cup V_P$ and a set of cliques $E = E_I \cup E_P \cup E_D$. V_I and V_P have the same cardinality and 4-neighbor grid structure. Nodes in V_I are labeled with labels that model in-plane deformation, while labels used in V_P model the plane position. Edges from E_I and E_P correspond to classical grid connections for nodes in V_I and V_P respectively; they are associated with regularization terms. Edges in E_D link every node from V_I with its corresponding node from V_P, creating a graph with a three dimensional structure; those terms encode the matching similarity measure. Note that E_I and E_P can be extended with the same type of extra edges defined in Section 2.1 (see Figure 2.a) to improve the satisfaction of the desired geometrical constraints.

We define two different label spaces, one associated with V_I and one associated with V_P. The first label space, L_I, is a bidimensional space that mod-

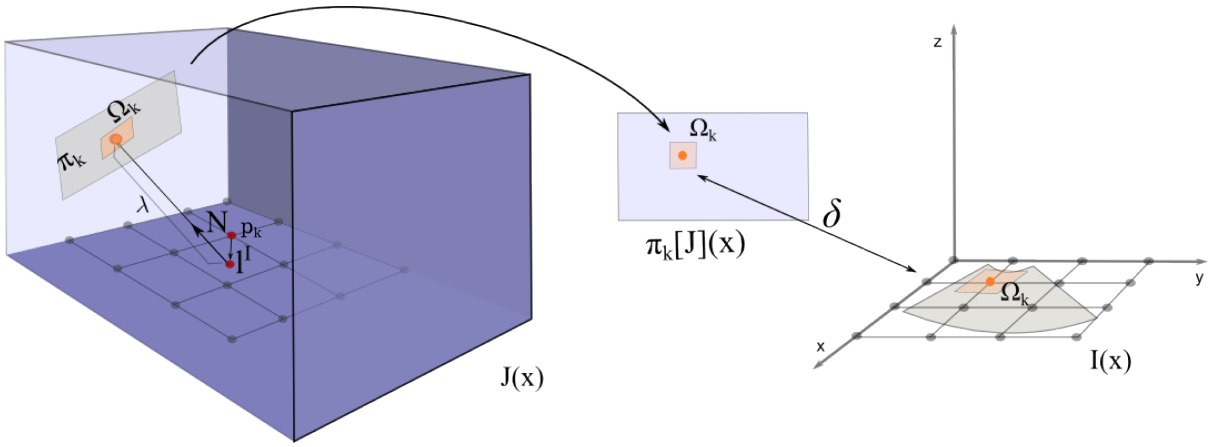


Figure 5 Data term formulation for the decoupled approach. It is similar to the formulation shown in Figure 3, but it combines labels from different label spaces. The points $\boldsymbol{x} \in \Omega_k$ are used to calculate the unary potential. $\pi_k[J](\boldsymbol{x})$ returns the intensity of the point in the 2D slice corresponding to the plane π_k in the 3D image, whereas $I(\boldsymbol{x})$ returns the 2D image intensity. δ represents the similarity measure. In order to compute the final position of the sampled patch in the volume, the in-plane deformation label $l^I = (d_x, d_y)$ is applied to the corresponding imaginary grid point $\boldsymbol{p_k}$. Then, label $l^P = (N, \lambda)$ is used: the point is translated in the direction given by vector $\mathbf{N}$ and scaled by a factor λ. In other words, we simply add the vector $\mathbf{N} * \lambda$. Finally, the patch Ω_k is sampled from plane π_k with normal N, centered at the displaced point $\boldsymbol{p_k}$ (in orange).

els in-plane deformation using displacement vectors $\boldsymbol{l^I} = (d_x, d_y)$. The second label space, L_P, indicates the plane in which the corresponding control point is located and consists of labels $\boldsymbol{l^P}$ representing different planes. In order to specify the plane and the orientation of the grid on it, we consider an orthonormal basis acting on a reference point in this plane. Using this information, we can reconstruct the position of the control points of the grid. The planes parametrization is given by $\boldsymbol{l^P} = (\phi, \theta, \lambda)$, where angles ϕ and θ define a vector $\boldsymbol{N}$ over a unit sphere, expressed through its spherical coordinates (see Figure 2.c). This value, together with parameter λ, defines the position of the plane associated with the given control point. This is an important advantage of our method: we could use prior knowledge to improve the way we explore the plane space, just by changing the plane space sampling method.

As it concerns the considered plane sampling method, the final position of every control point $\boldsymbol{p_k}$ of the grid is determined using the pairwise term between two graph nodes ($v_k^I \in V_I$ and $v_k^P \in V_P$) and their respective labels ($l_k^I \in L_I$ and $l_k^P \in L_P$). Imagine we have a plane π_k with normal vector $\boldsymbol{N}$ that contains the displaced control point $\boldsymbol{p_k} + \boldsymbol{l_k^I}$. Parameter λ indicates the magnitude of the translation we apply to π_k in the direction given by $\boldsymbol{N}$ in order to determine the plane's final position (see Figure 5 for a complete explanation). Given that we can associate different planes to different control points (by assigning them different labels $\boldsymbol{l^P}$), we need to impose constraints that will force the final solution to refer to a unique plane.

The energy that guides the optimization process involves three different pairwise terms, which encode data consistency between the source and the target, smoothness of the deformation and unique plane selection:

$$\mathcal{P}_{\mathcal{D}}(I, P, D) = \min \alpha \sum_{(i,j)\in E_I} e_{i,j}^I(\boldsymbol{l_i^I}, \boldsymbol{l_j^I}) + \beta \sum_{(i,j)\in E_P} e_{i,j}^P(\boldsymbol{l_i^P}, \boldsymbol{l_j^P}) + \sum_{(i,j)\in E_D} e_{i,j}^D(\boldsymbol{l_i^I}, \boldsymbol{l_j^P}), \tag{14}$$

where α, β are scaling factors, $e_{i,j}^I \in I$ are *in-plane* deformation regularizers (associated to edges in E^I), $e_{i,j}^P \in P$ are *plane* consistency constraints (associated with edges in E^P) and $e_{i,j}^D \in D$ are *data* terms (associated with edges in E^D). l_i^I, l_i^P are labels from label spaces L_I and L_P respectively.

The data term is defined for every control point of the imaginary grid $\boldsymbol{p_k}$ using the information provided by two associated graph nodes. It is encoded in the pairwise term $e^D \in E^D$. To this end, we extract an oriented patch Ω_k over the plane π_k (centered at $\boldsymbol{p_k}$) from the volume J, so that the similarity measure δ can be calculated between that patch and the corresponding area over the source 2D image (see Figure 5):

$$e_{i,j}^D(\boldsymbol{l_i^I}, \boldsymbol{l_j^P}) = \int_{\Omega_k} \delta(I \circ T_D(\boldsymbol{x}), \pi_k[J](\boldsymbol{x}))d\boldsymbol{x}. \tag{15}$$

We define two different regularization terms. The first controls the in-plane deformation; it is defined on V_I and corresponds to a symmetric distance preserving penalty:

$$e^I_{i,j}(\boldsymbol{l^I_i}, \boldsymbol{l^I_j}) = (1-\psi_{i,j}(\boldsymbol{l^I_i}, \boldsymbol{l^I_j}))^2 + (1-\psi_{i,j}(\boldsymbol{l^I_i}, \boldsymbol{l^I_j})^{-1})^2, \tag{16}$$

where $\psi_{i,j}$ is the distance defined in Equation 9.

The second term penalizes inconsistencies in terms of plane selection, and is defined on V_P. We use the earlier defined (at is concerns the overparameterized model, in Equation 12) point-to-plane distance:

$$e^P_{i,j}(\boldsymbol{l^P_i}, \boldsymbol{l^P_j}) = \frac{1}{2}(D_{\pi_j}(\boldsymbol{p_i}') + D_{\pi_i}(\boldsymbol{p_j}')). \tag{17}$$

where $\boldsymbol{p_i}'$ and $\boldsymbol{p_j}'$ are the positions after applying label $\boldsymbol{l^P_i}$, $\boldsymbol{l^P_j}$ to $\boldsymbol{p_i}$, $\boldsymbol{p_j}$ respectively.

Note that these terms are similar to the ones of the former approach. However, there is an important difference regarding the parameters they use. In case of the overparameterized approach, parameters are always 5-dimensional labels. In the current approach, parameters are at most 3-dimensional, thus reducing the complexity of the optimization process while also allowing a denser sampling of the solution space. Conventional pairwise inference algorithms could be used to optimize the objective function corresponding to the previously defined decoupled model. Such a model offers a good compromise between expression power and computational efficiency. However, the pairwise nature of such an approach introduces limited expression power in terms of energy potentials. The smoothness (regularization) terms with second order cliques are not invariant to linear transformations such as rotation and scaling [14], while being approximate in the sense that plane consistency is imposed in a rather soft manner. These concerns could be partially addressed through a higher order formulation acting directly on the displacements of the 2D grid with 3D deformation labels. Furthermore, the data term is just a local approximation of the real matching score between the deformed source 2D image and the corresponding target plane; by introducing high-order terms we could define it more accurately.

2.3 High-order approach

The new formulation consists of an undirected graph $G_H = \langle V, E \rangle$ with a set of nodes V and a set of third-order potentials $E = E_D \cup E_R$. The nodes are control points of our two-dimensional quasi-planar grid and they are displaced using 3D vectors $l_i \in L_H$. We define two types of cliques in E. Cliques in E_D are triplets of vertices with a triangular shape and they are associated with data terms. Cliques in E_R are collinear triplets of vertices (aligned in horizontal and vertical directions) forming third-order cliques associated with regularization terms.

Unlike the previous methods, which require extra labels to explicitly model the plane selection, high-order potentials explicitly encode them. Furthermore, third-order triangular cliques can also explicitly encode data terms, since the corresponding plane can be precisely determined using the position of these 3 vertices. We use triplets of collinear points for regularization terms. According to [26], this allows us to encode a smoothness prior based on the discrete approximation of the second-order derivatives using only the vertices' position. Therefore, we define a simple three dimensional label space of displacement vectors which is sampled as shown in Figure 2.b.

The energy to be minimized consists of data terms D_{ijk} associated with triangular triplets of graph vertices $(i,j,k) \in E_D$ and regularization terms R_{ijk} associated with collinear horizontal and vertical triplets $(i,j,k) \in E_R$. The minimization energy problem becomes:

$$\begin{aligned} \mathcal{P}_{\mathcal{H}}(D,R) = \min \sum_{(i,j,k)\in E_D} D_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}) + \\ \gamma \sum_{(i,j,k)\in E_R} R_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}), \end{aligned} \tag{18}$$

where γ is a scaling factor and $\boldsymbol{l_i}$ is a label associated with a displacement vector (d_x, d_y, d_z) and assigned to the node i.

The data term is defined over a disjoint set of triangular cliques, covering the entire 2D domain, as shown in Figure 6.a. Its formulation is independent of the similarity measure δ and it is calculated for each clique $\boldsymbol{c} = (i,j,k) \in E_D$ using the source 2D image I and the corresponding plane $\pi_d[J]$ extracted from the target volume J, defined by the three control points of the clique. For a given similarity measure δ, the data term associated with

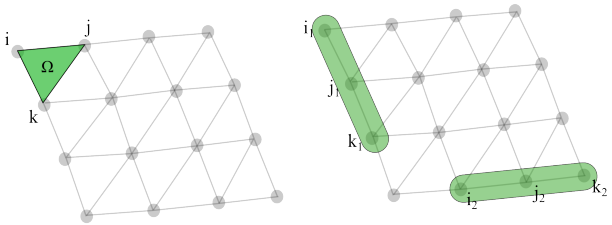


Figure 6 Different types of cliques used in the formulation. (a) Example of a triangular clique used for data term computation. The green patch Ω corresponds to the clique (i, j, k) and it is used to calculate the data term. (b) Examples of vertical (i_1, j_1, k_1) and horizontal (i_2, j_2, k_2) collinear third-order cliques used to regularize the grid structure.

the clique $\boldsymbol{c}$ is thus defined as:

$$D_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}) = \int_{\Omega_{(l_i, l_j, l_k)}} \delta(I \circ T_D(\boldsymbol{x}), \pi_d[J](\boldsymbol{x}))d\boldsymbol{x}, \tag{19}$$

where $\boldsymbol{x} \in \Omega_{(l_i,l_j,l_k)}$, and $\Omega_{(l_i,l_j,l_k)}$ corresponds to the triangular area defined by the control points of clique $\boldsymbol{c} = (i, j, k)$ over the plane $\pi_d[J]$, after applying the corresponding labels $\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}$ to the vertices.

Smoothness and plane consistency are also imposed using higher order cliques. We define a clique for every set of three collinear and contiguous grid nodes (in horizontal and vertical directions as depicts Figure 6.b). We also introduce extra cliques formed by nodes that are collinear but not contiguous. The aim is to propagate the regularization so that the planar structure is conserved. The regularization term, as noted previously, seeks to satisfy the plane structure of the grid and the smoothness nature of the in-plane deformations.

Planar consistency can be easily enforced by propagating a null second-derivative constraint among collinear triplets of points. In fact, a null second-derivative for these cliques does not impose just a planarity constraint but it also aims at regularizing the grid structure. Thanks to the third-order cliques, we can accurately approximate a discrete version of the second-order derivative [26]. Given three contiguous control points $(\boldsymbol{p_i}, \boldsymbol{p_j}, \boldsymbol{p_k})$ and their corresponding displacement labels $(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k})$, it can be approximated as follows: $||\ (\boldsymbol{p_i} + \boldsymbol{l_i}) + (\boldsymbol{p_k} + \boldsymbol{l_k}) - 2 \cdot (\boldsymbol{p_j} + \boldsymbol{l_j})\ ||$.

Based on this idea, we define the following energy term that is proportional to the second derivative, and normalized with the original distance between the control points, d:

$$R^A_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}) = \frac{||(\boldsymbol{p_i}+\boldsymbol{l_i})+(\boldsymbol{p_k}+\boldsymbol{l_k})-2\cdot(\boldsymbol{p_j}+\boldsymbol{l_j})||}{d^2}^2, \tag{20}$$

In-plane deformation smoothness is reinforced in the same manner as the previous models - through a symmetric distance preserving approach. For the sake of clarity, we redefine Equation 10 as $\Psi_{ij}(\boldsymbol{l_i}, \boldsymbol{l_j}) = (1 - \psi_{i,j}(\boldsymbol{l_i}, \boldsymbol{l_j}))^2 + (1 - \psi_{i,j}(\boldsymbol{l_i}, \boldsymbol{l_j})^{-1})^2$, and we apply it to both pairs of contiguous points that form the clique (i, j, k):

$$R^B_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}) = \frac{\Psi_{ij}(\boldsymbol{l_i}, \boldsymbol{l_j}) + \Psi_{jk}(\boldsymbol{l_j}, \boldsymbol{l_k})}{2}. \tag{21}$$

The equation that regularizes the grid is a weighted combination of both terms R^A_{ijk} and R^B_{ijk}:

$$R_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}) = (1 - \alpha)R^A_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}) + \alpha R^B_{ijk}(\boldsymbol{l_i}, \boldsymbol{l_j}, \boldsymbol{l_k}), \tag{22}$$

where α represents a weighting factor used to calibrate the regularization term.

3 Results and discussion

Let us now proceed with a systematic evaluation of the proposed methods. One of the main aspects shared across methods is the inference algorithms used to produce the desired solution.

3.1 Inference methods

Depending on their cardinality and regularity, objective functions can be optimized using a variety of discrete optimization algorithms which offer different guaranties. It must be noted that the regularization terms presented in our three models are non submodular, since we include the current position of the points (which can be arbitrary) in their formulation. Therefore, submodularity constraint is fulfilled neither in the pairwise nor in the high-order terms (for a clear definition of submodularity in pairwise and high-order energies, we refer the reader to the work of [38]).

In [10], the overparameterized approach was optimized using the FastPD algorithm [23] while for the decoupled [9] and the higher order models [8], we consider loopy belief propagation networks. For the sake of fairness, in order to improve the confidence of the comparison among the three

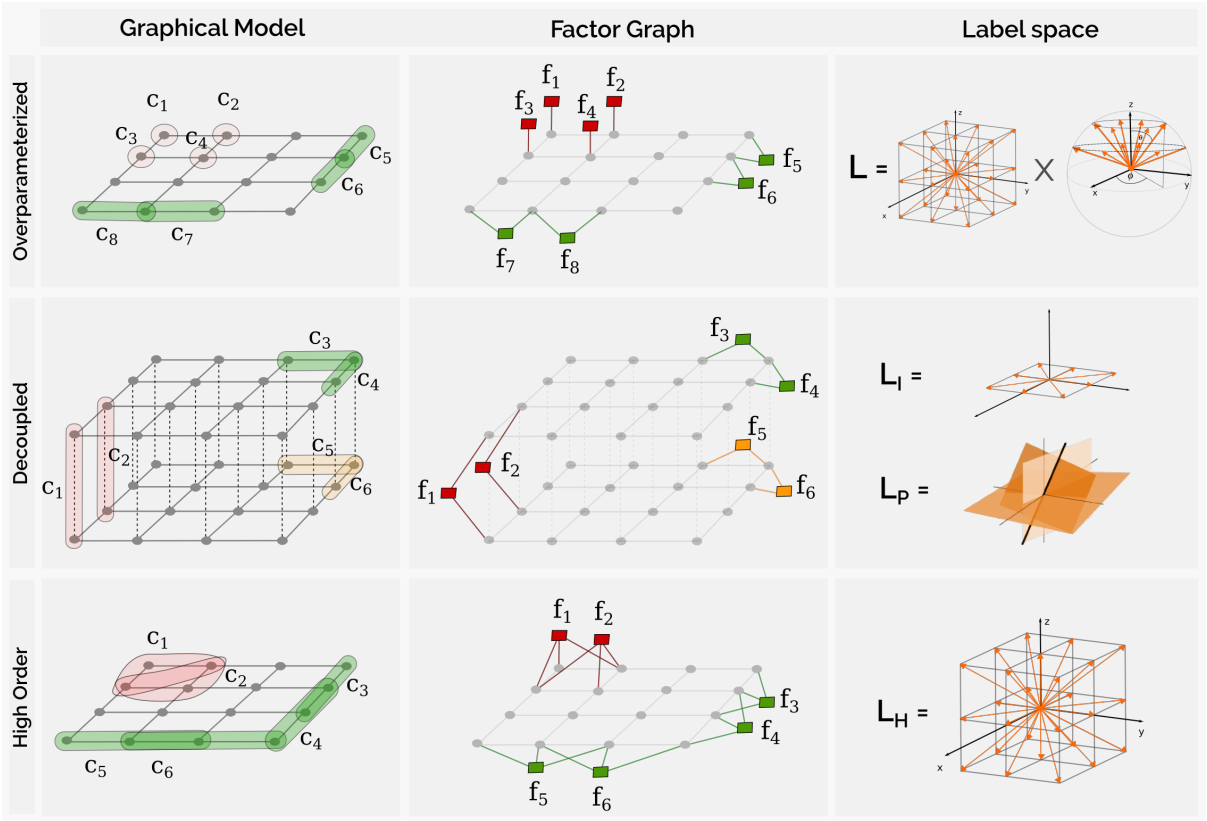


Figure 7 Factor graph derivation and labels spaces corresponding to the overparameterized, decoupled and high-order approaches. It shows the equivalence between cliques c_i in the first column (unary, pairwise and high-order, depending on the model) and the corresponding factors f_i in the second column. In red, we observe the cliques and factors associated with data terms, while in green and orange we represent those associated with regularization terms. In the third column, we include a figure representing the label space associated to every model (orange vectors and planes are associated to different labels). Note that the overparameterized approach is defined as the Cartesian product between the displacements and plane selection labels, while in the decoupled approach these label spaces are independent.

methods, in this work we adapted it to be optimized with the same algorithms. Therefore, results in this work can not be directly compared with our previous works.

Given the variety of models presented in this work, we chose two different inference methods that can deal with arbitrary graph topologies and clique orders, coming from two standard inference algorithm classes: (i) Loopy Belief Propagation (LBP), a well known message passing algorithm that has been extensively used in the literature; and (ii) the Lazy Flipper (LF) by [1], a move-making algorithm which is a generalization of the classical Iterated Conditional Modes (ICM) [3] and has provided good approximations for several non-submodular models in different benchmarks. Both are approximate inference methods that can accommodate arbitrary energy functions, graph topologies and label spaces, and allow us to show how the three proposed approaches perform under different optimization strategies.

3.1.1 Loopy Belief Propagation

LBP estimates a solution by iteratively passing local messages around the variables of the random field. These messages m_{ij} (sent from a node i to a node j) are actually vectors of size $\mid L \mid$ (cardinality of the label space), where every scalar entry represents what node i *thinks* about assigning label l to the node j. Once a node i receives all the messages from its neighbors, it compute its beliefs (also vectors of size $\mid L \mid$) in a label l_i. The messages are iteratively passed from one node to its neighbors until no change occurs from one iteration to the next one. When convergence is achieved, the MAP labeling is obtained for every node i as the label l_i that minimizes the corresponding belief. Note that both, messages and beliefs computed for a given

node, depend on the messages received from its neighbors. Therefore, if the graph that underlies the MRF is a tree, this process is initialized in the roots since messages for these nodes can be calculated considering just their potentials. In this case, at convergence, the solution is guaranteed to be optimal for arbitrary energies. If the structure is not a tree, messages are passed in any arbitrary order, but the algorithm is not guaranteed to converge in a finite number of iterations. Nonetheless, LBP has shown good performance in empirical studies [31].

3.1.2 Lazy Flipper

LF is a move-making algorithm proposed by [1]. It is a generalization of the well-known ICM which offers a systematic way to explore (exhaustively or not) the search space. The idea is to start from an arbitrary initial assignment and perform successive flips of variables that reduce the energy to be minimized. A greedy strategy is adopted to explore the space of solutions: as soon as a flip reducing the energy is found, the current configuration is updated accordingly. In a first stage, only one variable is flipped at a time (as in ICM). However, once a configuration is found whose energy can no longer be reduced by flips of one variable, a new stage starts where all subsets of two connected variables (i.e. variables that are linked by an edge in the graph) are considered. This strategy is applied, considering sets of maximum size k. This parameter controls the search depth. For $k = 1$, it specializes to ICM. For bigger values of k a trade-off between approximation quality and runtime is established, which in the limit converges to an exhaustive search over only the connected subgraphs (intractable in most of the cases).

3.1.3 Factor graphs

We have adopted the OpenGM2 library [19] which implements both inference methods, and makes it possible to perform fair comparisons. It requires construction of a factor graph for every scheme (see Figure 7).

A factor graph G' is a bipartite graph that factorizes a given global energy function, expressing which variables are arguments of which local functions [25]. Given a graphical model of any order $G = \langle V, E\rangle$ (like the ones described in this work), we can derive a factor graph $G' = \langle V', F', E'\rangle$. Here, V' is the set of variable nodes formed by the nodes of G, F' is a the set of all the factors $f \in F'$ (where every f is associated to one clique G), and the set $E' \subset V' \times F'$ defines the relation between the nodes and the factors. Every factor f has a function $\varphi_f : V'^n \rightarrow \mathbb{R}$ associated with it, that might correspond to one of the data or regularization terms defined in previous sections. The energy function of our discrete labeling problem in the context of factor graphs is then given by:

$$\mathcal{E}(x) = \sum_{f \in F'} \varphi_f(l_1^f, ..., l_n^f), \tag{23}$$

where x corresponds to a given labeling for the complete graph and $l_1^f ... l_n^f$ are labels given to the variables in the neighborhood (or scope) of the factor f. Figure 7 shows a comparison between the three models and the derivation of the corresponding factor graph in each case.

3.1.4 Incremental approach

In order to improve the quality of the label space sampling (and therefore the accuracy of the results) while keeping a low computational cost, we adopted a greedy incremental approach where the label space is refined for every time we run the inference algorithm. In that way we explore a wider range of parameters which result in more accurate sampling when composed after several iterations. A similar approach has been successfully used in previous graph-based registration papers ([10, 8, 9, 15, 16]).

3.2 Experimental validation

We compute results on two different datasets for the three methods, using the two inference algorithms (LBP and LF) in order to validate both the resulting 2D-2D deformation field and the final plane estimation. The first one is a monomodal MRI heart dataset while the second one consists of 6 sequences of multimodal US-MRI brain images.

For every registration case, we run the inference algorithm several times (more precisely, the inference method is executed a number of times equal to the product between grid refinement levels and label refinement levels). For a single execution of both inference methods, we use the same compound stopping criterion based on the energy gap between iterations and maximum running time. The algorithms run until the energy improvement between two iterations is smaller than a fraction of

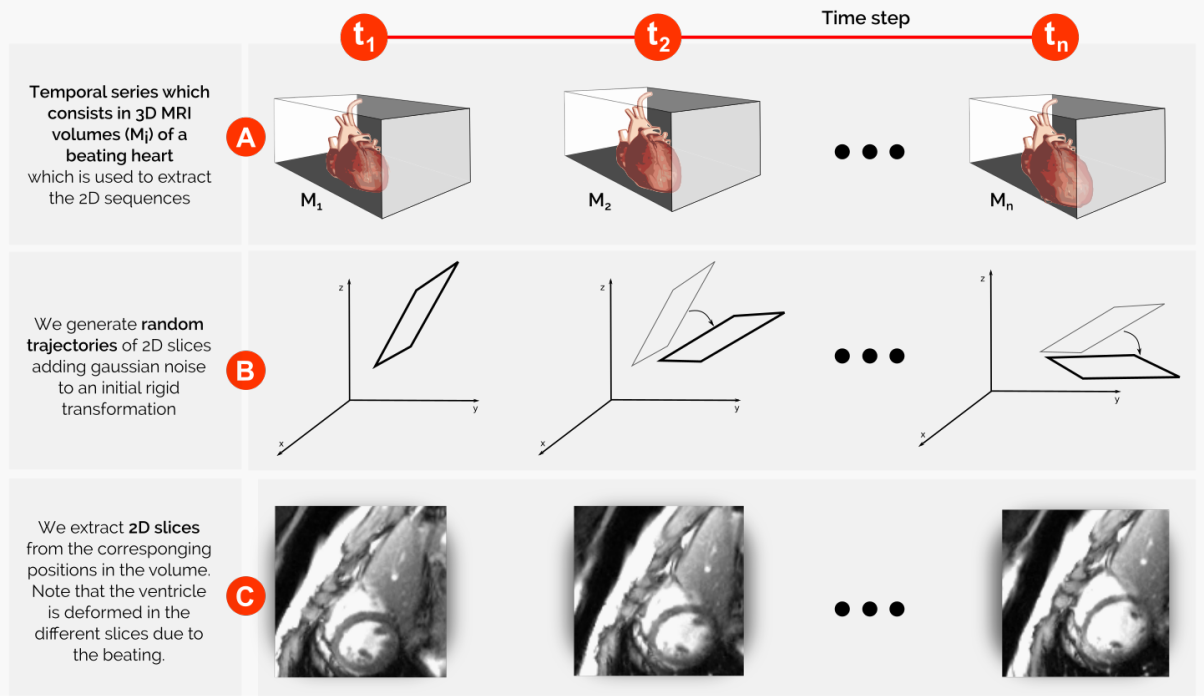


Figure 8 Heart dataset construction. Given a series of 3D MRI volumes of a beating heart (A), we extract ten different random trajectories (B). Every trajectory is composed of twenty different positions from which we extract the 2D slices (C).

the current energy (we use $\epsilon = 0.01\%$). If convergence is not achieved before a timeout is reached, the algorithm stops and returns the best explored solution. A timeout of 60 seconds is used since we observe that it is enough to achieve convergence in most of the registration cases. When it is not achieved within this time, it can take too long. For LF we used a maximum depth of $k = 2$ (for details about LF, we refer the reader to section 3.1 or to the work of [19]).

We run the same experiments using a continuous approach to estimate rigid and deformable parameters, which serve as baseline for comparison. We adopted the best deformable approach (namely *Cont Def - Two Steps*) from a brief comparative analysis included in Appendix 4, where we discussed alternative continuous models for slice-to-volume registration. Please refer to Appendix 4 for a detailed discussion about this model. The continuous optimization is performed using the simplex algorithm proposed by [32]. Also known as Nelder-Mead, downhill simplex or amoeba, the simplex method is one of the most popular continuous derivative-free methods. It relies on the notion of simplex (a $n + 1$ vertices polytope living in a n-dimensional space) to explore the space of solutions in a systematic way. At every iteration, the method constructs a simplex over the search surface, and the objective function is evaluated on its vertices. In the simplest version, the algorithm moves across the surface by replacing, at every iteration, the worst vertex of the current set by a point reflected through the centroid of the remaining n points. The method can find a local optimum when the objective function varies smoothly and is unimodal. It has also shown to be more robust when dealing with complicated parameter space than standard gradient-based methods, providing a good compromise between robustness and convergence time [27]. It has been widely used in a variety of slice-to-volume applications, to estimate all kinds of transformation models optimizing a variety of similarity measures [7,4,12,34]. We optimized a global energy where the similarity measure was computed for the complete image, since no local deformation model is considered.

In the following subsections we describe the datasets and present quantitative and qualitative results.

3.2.1 Monomodal dataset experiment

The monomodal dataset was derived from a temporal series of 3D heart MRI volumes. It consists of 10 sequences of 19 MRI slices which have to be registered with an initial volume. The slices are

extracted from random positions in the volumes while satisfying spatio-temporal consistency. The ground truth associated with this dataset refers to the rigid transformation used to extract every 2D slice of every sequence (it is used to validate the plane estimation or rigid registration) and a segmentation mask of the left endocardium, that can be used to validate the quality of the estimated deformation field.

The dataset was generated from a temporal series of 3D heart MRI volumes M_i as shown in Figure 8. For a given sequence in the heart dataset, every 2D slice I_i was extracted from the corresponding volume M_i at a position which is calculated as follows. Starting from a random initial translation $T_0 = (T_{x_0}, T_{y_0}, T_{z_0})$ and rotation $R_0 = (R_{x_0}, R_{y_0}, R_{z_0})$, we extract the first 2D slice I_0 from the initial volume M_0. Then, gaussian noise is added to every parameter of the transformation in order to generate the position of the next slice at the next volume. We used $\sigma_r = 3°$ as rotation and $\sigma_t = 5$mm as translation parameters. Those parameters generate maximum distances of about 25mm between the current and the succeeding plane. In this way, we generated 2D sequences that correspond to trajectories inside the volumes. Since the initial 3D series consists of temporally spaced volumes of the heart, there are local deformations between them due to the heartbeat; therefore, extracted slices are also deformed.

The resolution of the MRI volume is $192 \times 192 \times 11$ voxels and the voxel size is $1.25\text{mm} \times 1.25\text{mm} \times 8\text{mm}$. The slices of the 2D sequences are 120×120 pixels with a voxel size of $1.25\text{mm} \times 1.25\text{mm}$.

Experiments for the 3 methods were performed using equivalent configurations. In all of them we used 3 grid refinement levels, 4 steps of label refinement per grid level, initial grid size of 40mm and minimum patch size (for similarity measure calculation) of 20mm. In case of the overparameterized approach we used $\alpha = 0.8$, $\gamma = 1$ and 342 labels; for the decoupled approach we used $\alpha = 0.8$, $\beta = 0.2$, 25 labels in the 2D deformation space and 91 in the plane selection space; and finally, for the high-order approach we used $\alpha = 0.5$, $\gamma = 1.10$ and 19 labels. Parameters α, β, γ were chosen using cross-validation. The number of labels in every label space was chosen to make the search spaces as similar as possible. Recall that alternative label spaces were adopted in every approach: the overparameterized model uses 5-dimensional labels describing in-plane deformation and plane selection variables; the decoupled model divides this unified label space into two separate ones, the in-plane deformations label space and the plane selection label space; finally, the high-order model uses a unique and simpler label space composed of 3-dimensional displacement vectors.

Results are reported (for every approach and every inference method) for 10 sequences of 19 images, giving a total of 190 registration cases. We also included the results corresponding to the rigid approach optimized using simplex method. We used SAD as similarity measure given that we are dealing with monomodal registration. The idea is to register every 2D slice I_i (which plays the role of an intra-operative image) to the same initial volume M_0 (which acts as the pre-operative image). The resulting position of the slice I_i was used to initialize the registration of slice I_{i+1}.

Figure 10 shows results in terms of rigid transformation estimation. We measured the distance between the transformation parameters, and reported the average of the 190 registration cases. It resulted in less than 0.02rad (1.14°) for rotation and less than 1.5mm for translation parameters in all the discrete approaches and optimization methods. The discrete methods outperform the results obtained using the continuous baselines. The decoupled method dominates the other two by orders of magnitude in terms of reduction of the standard deviation and the mean error. However, in terms of performance, both decoupled and high-order methods are equally good when compared to the overparameterized approach whose computational time is higher (as expected, given the high dimensionality of the label space). This can be observed in figures 13, 14 and 15.

To measure the influence of the deformation in the final results, we used the segmentations being associated with the dataset. We computed statistics for the segmentation overlapping at three different stages: before registration (i.e. between the source image and the target volume slice corresponding to the initial transformation), after rigid registration (i.e. between the source image and the target volume slice corresponding to the estimated transformation) and after deformable registration (i.e. between the deformed source image and the target volume slice corresponding to the estimated transformation). We evaluated accuracy computing DICE coefficient, Hausdorff distance and Contour Mean Distance (CMD). We also provided sensitivity (which measures how many pixels from the reference image are correctly segmented in test image) and specificity (which measures how many

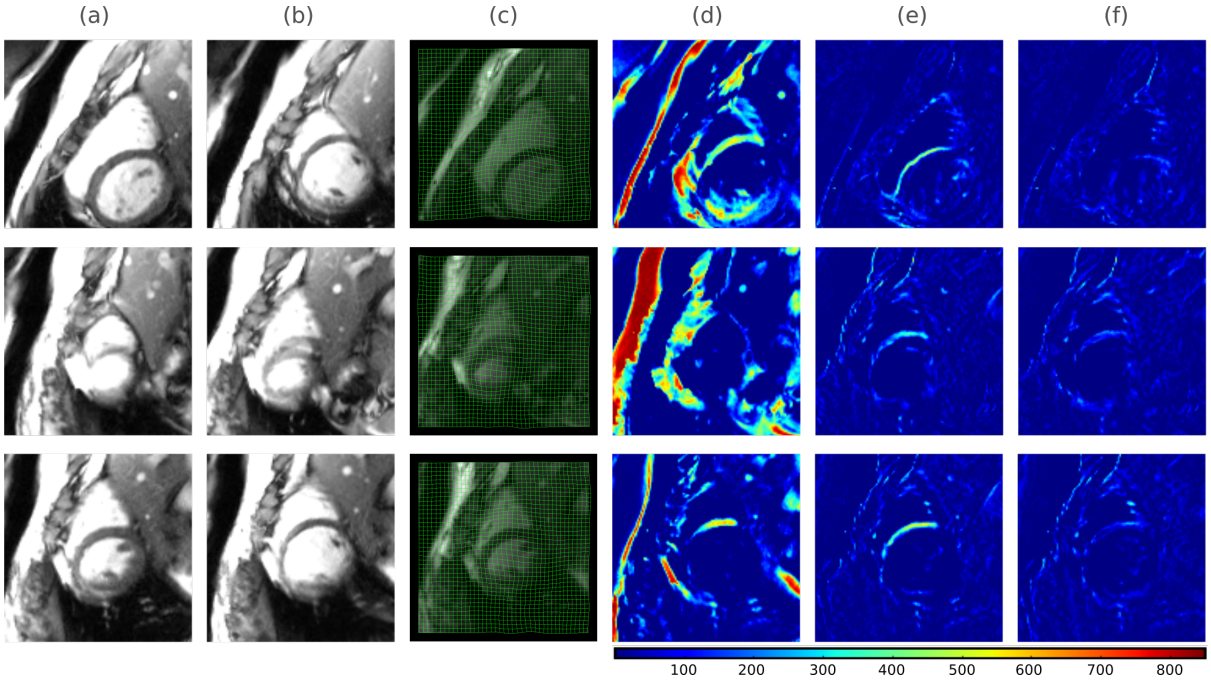


Figure 9 Slices extracted from three different sequences of the heart dataset before and after registration. Input slice (a) is initialized in a position specified by a rigid transformation within the volume, whose slice corresponds to (b). After deformable registration, a deformation field (c) is estimated. (d) shows the difference between initial images (a) and (b), while (e) shows the difference between (a) and the corresponding slice extracted after rigid registration. Finally, (f) corresponds to the results after deformable registration (i.e., the difference between the deformed version of slice (a) and the slice corresponding to the estimated transformation). Red indicates bigger differences between the images. Note how these values are changing before (d), after rigid (e) and after deformable (f) registration.

pixels outside the reference image are correctly excluded from the test image) coefficients to complete the analysis. Results presented in Figure 11 show the mean and standard deviation of the indicators at the three stages, for the three approaches and the two inference methods. The discrete methods outperform the continuous approaches (rigid and deformable) in all the cases. It can be seen that results improve at each stage, achieving DICE coefficient of around 0.9 after deformation. Hausdorff distance and CMD decreased at each stage until a total reduction of around 66%. Decoupled method still outperforms the others after deformation in all the indicators, and presents a substantial improvement in terms of standard deviation reduction with respect to them (it is consistent with the results we showed in Figure 10 for the rigid parameters). Figure 12 complements these results by showing DICE values per sequence, while Figure 9 shows some qualitative results before, after rigid and after deformable registration.

Finally, in terms of running time, Figure 13 presents the average value for the three approaches and the two inference methods, together with the distribution with respect to data cost computation and optimization time. As we can see, the decoupled method again outperforms the other two when inference is performed using LBP. We run all the experiments (brain and heart datasets) on an Intel Xeon W3670 with 6 Cores, 64bits and 16GB of RAM.

3.2.2 Multimodal experiment

Another dataset was used to test our approaches on multimodal image registration. The dataset consists of a preoperative brain MRI volume (voxel size of 0.5mm $\times$ 0.5mm $\times$ 0.5mm and resolution of 394 $\times$ 466 $\times$ 378 voxels) and 6 series of 9 US images extracted from the patient 01 of the database MNI BITE presented in [30]. The intra-operative US images were acquired using the prototype neuronavigation system IBIS NeuroNav. We generated 6 different sequences of 9 2D US images of the brain ventricles, with resolution around 161 $\times$ 126 pixels and pixel size of 0.3mm $\times$ 0.3mm. The brain ventricles were manually segmented in both modalities. The estimated position of the slice n was used to initialize the registration process of slice $n+1$. Slice 0 was initialized in a position near

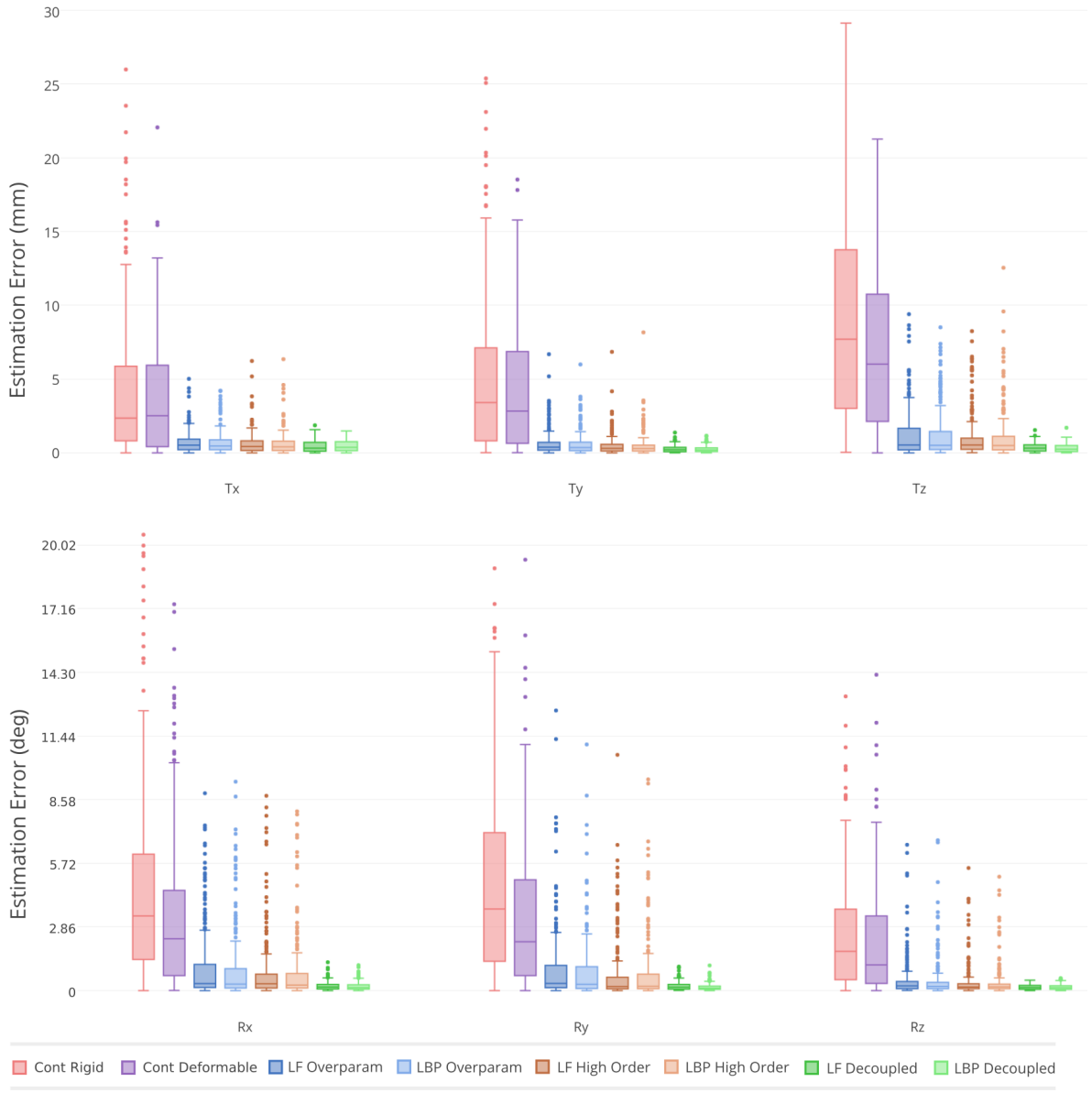


Figure 10 Rigid transformation estimation error for the heart dataset. We measured the distance for every one of the 6 rigid parameters, for the three approaches using LF and LBP as inference methods and for the continuous rigid and deformable approaches. The discrete methods outperform the results obtained using the continuous baselines. Independently of the inference method, the decoupled approach outperforms the other two in terms of average and standard deviation of the estimated error, for all the 6 parameters.

the ground truth using the rigid transformation provided together with the dataset. We computed statistics as we did in the previous experiment, but in this case based on the overlap between ventricle segmentations. Since we registered input images of different modalities, we used Mutual Information as similarity measure instead of SAD.

Figure 11 summarizes the average DICE, specificity, sensibility, Hausdorff distance and Contour Mean Distance coefficients for all the series, while Figure 13 reports the running times. Figure 12 complements these results by showing DICE values disaggregated per sequence. Note that the decoupled method does better in terms of computational time (independently of the inference method). However, the high-order method achieves better results in terms of segmentation statistics (in the order of 5% in terms of DICE, 2mm for Hausdorff distance and 0.5mm for contour mean distance) while keeping low running times, specially when using LF as optimization strategy (see figures 14 and 15 for a comparison between running time and

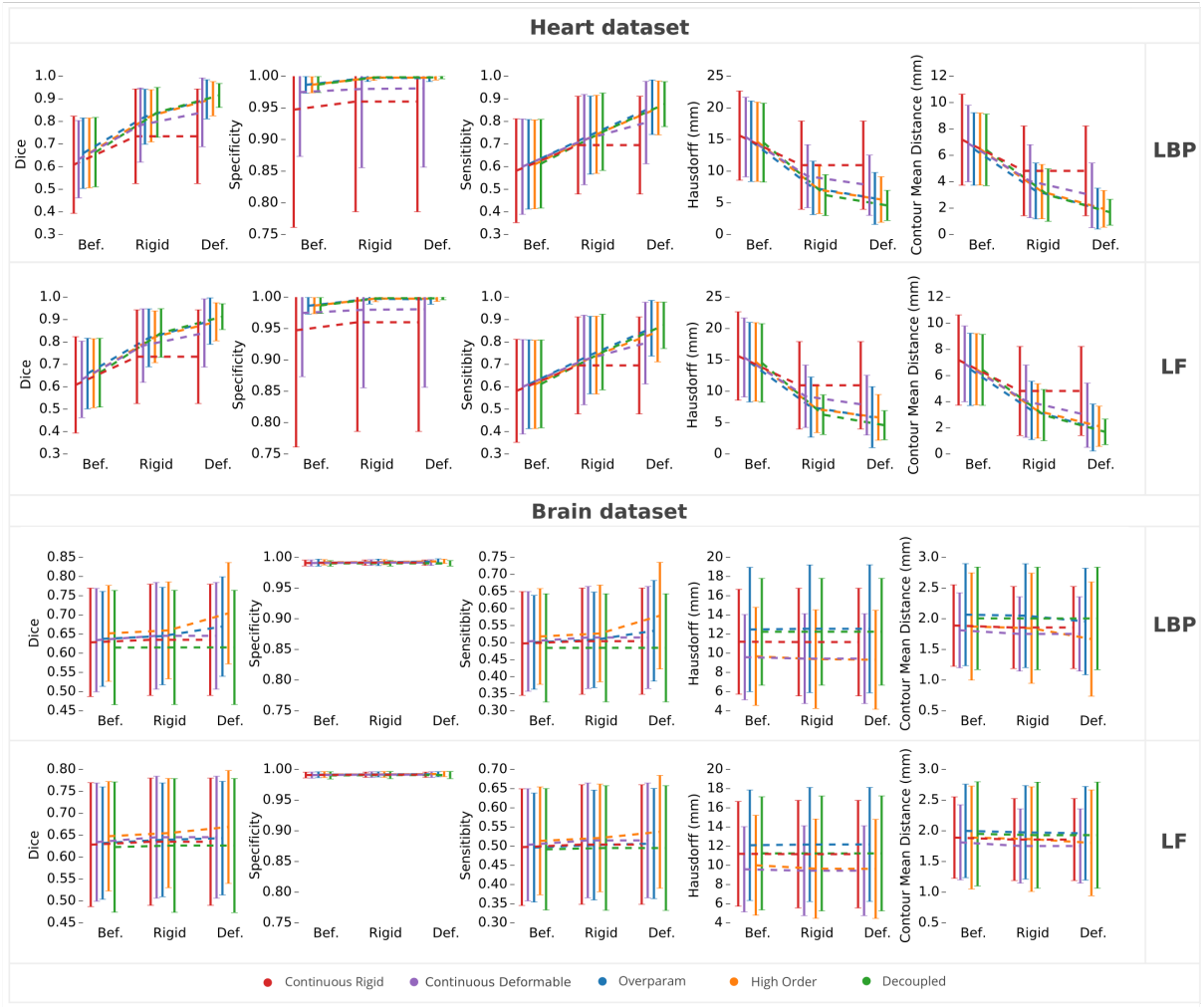


Figure 11 Segmentation overlapping statistics computed before, after rigid and after deformable registration for both datasets (190 registration cases for the heart dataset and 54 for the brain dataset). In the case of deformable registration, the source segmentation mask was deformed using the estimated deformation field. Results are reported for the continuous approach that only estimates rigid parameters using the simplex method, the continuous approach that estimates both rigid and deformable parameters, and for three discrete methods (overparameterized, high-order and decoupled) using both inference strategies (LBP and LF). Note that for every sequence of several contiguous 2D images, the resulting transformation from slice I_i was used to initialize the registration of slice I_{i+1}. Therefore, the accumulated error during the registration of successive 2D images for every method lead to different scores 'before registration'.

energy or accuracy, respectively). It must be noted that, in this case, we are dealing with a more complex problem than in the case of monomodal registration; consequently, the increment obtained in terms of accuracy for both, rigid and deformable registration, is smaller. Given that we are dealing with highly challenging images of low resolution being heavily corrupted from speckle, those results are extremely promising. It is known to the medical imaging community that explaining correspondences between different modalities is an extremely difficult task.

In all brain experiments, we used initial grid size of 8mm, a minimum patch size of $13mm$, histograms of 16 bins to measure mutual information similarity, a grid level of 3 and 4 steps of label refinement per grid level. In case of the overparameterized approach, we used $\alpha = 0.9$, $\gamma = 0.1$ and 342 labels; for the decoupled approach $\alpha = 0.015$, $\beta = 0.135$, 25 labels in the 2D deformation space and 91 in the plane selection space; finally, for the high-order approach $\alpha = 0.7$, $\gamma = 0.05$ and 19 labels. Parameters were chosen similarly as in the heart experiments.

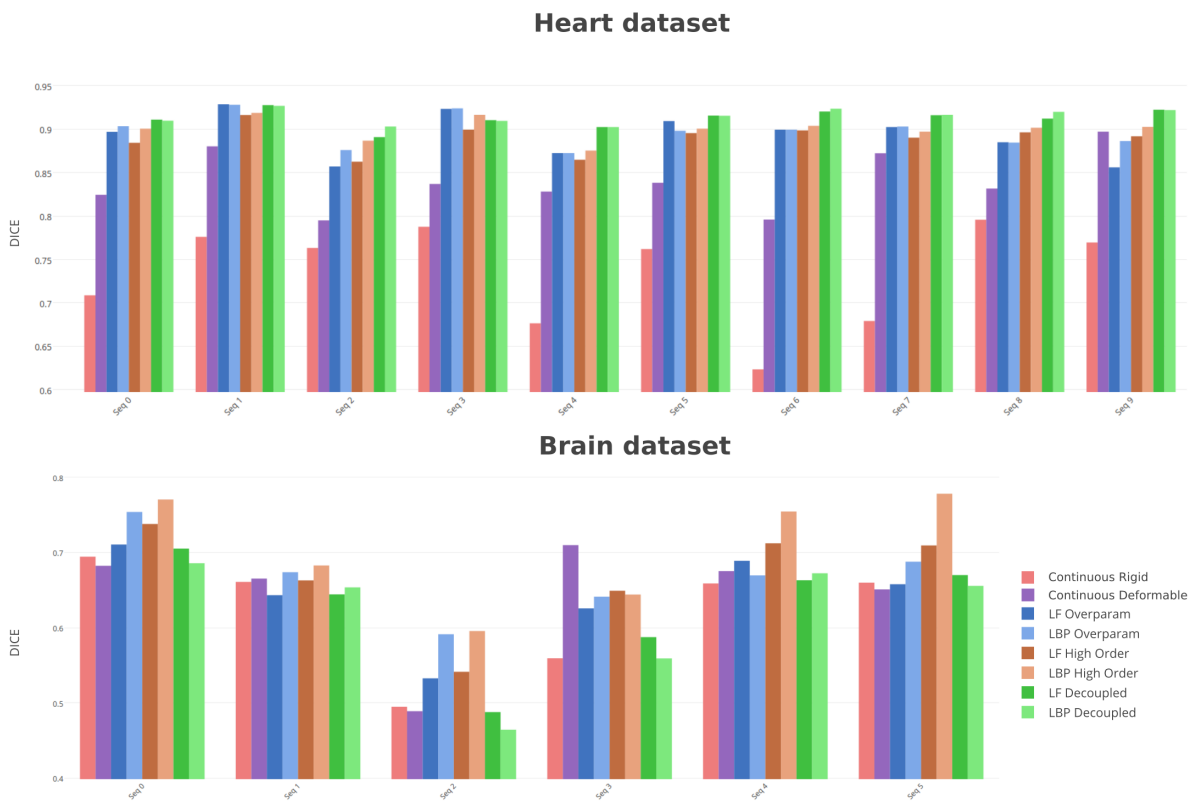


Figure 12 Final DICE (after deformation) comparison for every sequence (10 sequences in the heart dataset and 6 sequences in the brain dataset). Results are shown for the rigid approach optimized using simplex method, as well as the three discrete approaches where inference is performed using two different methods. The discrete methods outperform the continuous approaches (both rigid and deformable models). In case of the heart dataset, decoupled method outperforms the other two discrete in most part of the sequences. In the brain dataset, the high-order approach shows best performance in most cases. This is coherent with the aggregated results shown in Figure 11.

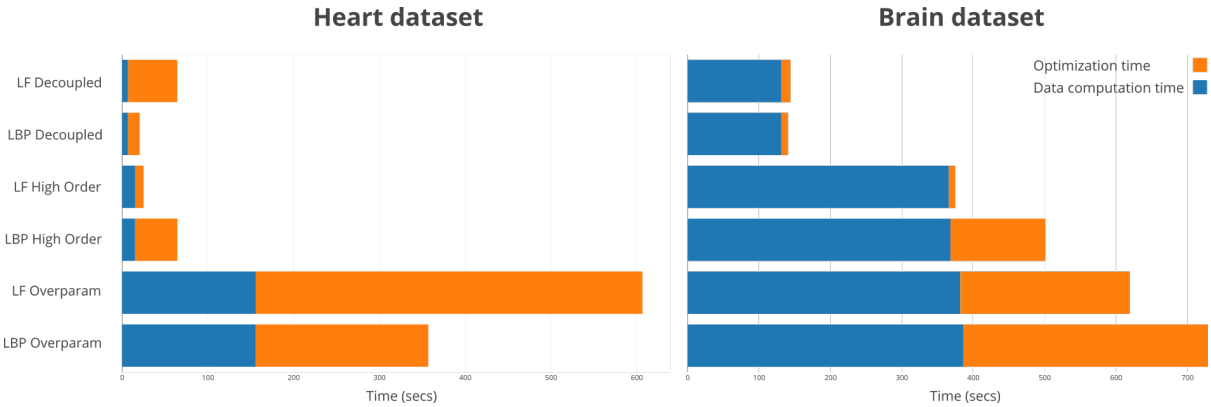


Figure 13 Average running time expressed in seconds, for one registration case, for the three approaches running on the heart dataset (a) and the brain dataset (b), using LF and LBP. Blue part corresponds to data cost computation while orange part corresponds to the optimization time. As we can observe, data cost computation represents a bigger portion of the total time in the brain dataset than in the heart dataset. This is due to the similarity measure: while in the monomodal case (heart) we use a simple SAD, in the multimodal case (brain) we need a more complex measure like mutual information. Note that data cost computation time remains constant when we vary the inference method (with small fluctuations due to operating system routines which ran during the experiment) but not across different models.

3.3 Comparative analysis

In this section, we aim at comparing different aspects of the three approaches we have presented in this paper, namely label spaces, graph topology and computational time. Without loss of gener-

ality, some assumptions are made regarding the models. First, we consider only square grids where N is the number of control points and consequently $\sqrt{N}$ is the number of nodes per side. Second, for the sake of simplicity we do not consider the extra cliques introduced to improve the geometrical constraints propagation, since they are contemplated as an alternative strategy which may or may not be adopted.

Figure 14 shows a comparative analysis between the three approaches, using the two proposed inference methods, in terms of optimization time and final energy, while figure 15 includes a similar graph representing time vs accuracy (measured using DICE coefficient). Note that in figure 14, both methods are equivalent with respect to the final energy in general (without considering the outliers). However, there are more important differences in terms of computational time. In the high-order approach, where the label space is small, LF outperforms LBP since convergence is achieved in a few seconds, independently of the dataset. For bigger label spaces (like decoupled and overparameterized approaches), LBP converges faster in case of the heart dataset, where SAD is used as similarity measure and therefore the energy is smooth. The last case is when we use MI as similarity measure (brain dataset) and we have big label spaces: there is no clear pattern in this case. Note that these results are consistent with those shown in figure 15. Indeed, one can observe that graphs in figure 15 are essentially a flipped version (over the X axis) of graphs included in 14. This evidences a high correlation between low energy values and high accuracy of the results, proving that the energy is appropriately modeled.

Table 3.3 presents a compendium of the most critical parameters related to the proposed methods. Let us start with the label spaces. We divide them into two types: displacement space (L_D) and plane selection space (L_P). The first one contains the displacement vectors (2D or 3D, depending on the model) applied to the control points, while the second one contains the set of planes that can be chosen. In terms of cardinality of the label spaces, the overparameterized approach has the highest complexity, given by the cartesian product between the displacements and all the possible planes, $|L_D \times L_P|$. The decoupled model is dominated by the maximum of the cardinality of both label spaces, $\max(|L_D|, |L_P|)$. Finally, for the high-order model it depends only on $|L_D|$ since it is not necessary anymore to explicitly model which planes can be chosen - the triangles defined by the triplets of points describe a plane (and even more, a patch on this plane) by themselves. It clearly illustrates how we can reduce the complexity of a given label space by making smart decisions in terms of energy definition and graph topology.

Table 1 Memory footprint comparison among the three methods, using two different optimizers. The reported value corresponds to the maximum amount of memory that the process used while running in every case.

	Overparameterized	**Decoupled**	**High-order**
LBP	362 MB	37.9 MB	65.4 MB
LF	342 MB	34.9 MB	44.7 MB

However, there is always a trade-off. This strong reduction in the size of the label space, has an effect on other parameters like number of cliques and number of variables. In case of the decoupled model, the main advantage is related to the fact that while the number of variables and edges augment linearly (it goes from N to $2N$ in case of variables, and from $2N - 2\sqrt{N}$ to $5N - 4\sqrt{N}$ in case of pairwise edges), the number of labels decreases quadratically (from $|L_D \times L_P|$ to $\max(|L_D|, |L_P|)$). It results in better performance for the decoupled method as can be observed in Figure 13. A consequence of the third-order cliques in the high-order method is higher computation costs. Even then, judging from the running times reported in Figure 13, we achieve good experimental computation time because of the smaller label space.

Finally, we include a comparison in terms of memory footprints (see table 1) among the three methods, using two different optimizers. We reported the maximum amount of memory that a process consumed while running one registration case for the heart dataset. As expected, the overparemterized model requires more memory than the other two approaches. Results also suggest that LF is more efficient in terms of memory consumption than LBP, since given the same graphical model, LF always outperforms LBP in terms of memory consumption.

4 Conclusion

We derived three new models from the standard graph-based deformable registration theory for slice-to-volume registration. We have shown promising results in a monomodal and a multimodal case, using different inference methods, and we compare them with baseline rigid and non-rigid approaches were inference is performed using con-



Figure 14 Comparison between total optimization time and final energy using two different optimizers (LBP corresponds to circles and LF to crosses). Results are shown for the overparameterized approach (in blue), the high-order approach (in orange) and the decoupled approach (in green). The gray lines connect data points corresponding to the same registration case.

Table 2 Comparison among the three methods in terms of label space and graph topology.

Parameter	Overparameterized	Decoupled	High-order
Label space	$\|L_D \times L_P\|$	$\max(\|L_D\|, \|L_P\|)$	$\|L_D\|$
# variables	N	$2N$	N
# 1st order cliques	N	-	-
# 2nd order cliques	$2N - 2\sqrt{N}$	$5N - 4\sqrt{N}$	-
# 3rd order cliques	-	-	$4N - 6\sqrt{N} + 2$

tinuous optimization. The proposed framework inherits the advantages of graph-based registration theory: modularity with respect to the similarity measure, flexibility to incorporate new types of prior knowledge within the registration process (through new energy terms) and scalability given by its parallelization potential.

The three methods we have presented aim at optimizing different types of energy functions in order to get both, rigid and deformable transformations that can be applied independently, according to the problem we are trying to solve. An extensive evaluation in terms of different statistical indicators has been presented, together with a comparative analysis of the algorithmic and computational complexity of each model. This work constitutes a clear example of the modeling power of graphical models, and it pushes the limits of the state-of-the-art by showing how a new problem can be solved not just in one, but in three different ways.

Numerous future developments built upon the proposed framework can be imagined. In this work, we proposed a joint model which encodes rigid and deformable parameters through a 2D grid of control points living in 3D space. An alternative approach, standard in the literature of slice-to-volume registration using continuous methods (i.e.

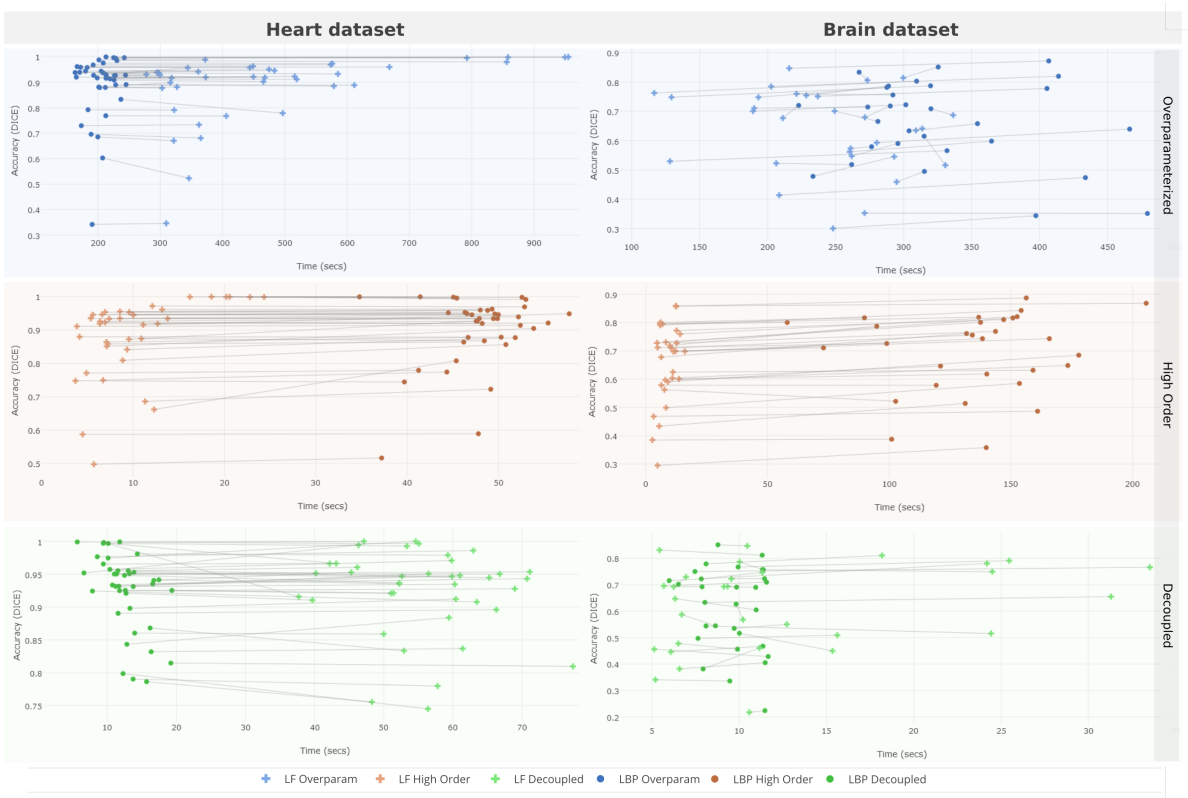


Figure 15 Comparison between total optimization time and results accuracy (measured using DICE coefficient) using two different optimizers (LBP corresponds to circles and LF to crosses). Results are shown for the overparameterized approach (in blue), the high-order approach (in orange) and the decoupled approach (in green). The gray lines connect data points corresponding to the same registration case.

[34]), consists in decoupling the parameters into a unique global rigid transformation (6 DOF) for plane selection, and a 2D deformation model, which can be optimized in two-steps or simultaneously, as we discussed in Appendix 4. Adopting a similar model in the discrete case would help to reduce the number of parameters in the label space, by increasing the complexity of the graphical model itself. In that sense, the recent work presented by [37] suggests a strategy to optimize global transformations through discrete graphical models in the context of slice-to-volume registration, which could be combined with a simplified version of the proposed models encoding the deformable parameters.

Alternative optimization methods and in particular second order methods in the context of higher order inference could improve the quality of the obtained solution while decreasing the computational complexity. The integration of geometric information (landmark correspondences) combined with iconic similarity measures [44] could also be an interesting additional component of the registration criterion. Last but not least, domain/problem specific parameter learning [2,24] towards improving the proposed models could have a positive influence on the obtained results.

Appendix 1: Continuous slice-to-volume registration

In this appendix, we include a brief comparative study among alternative continuous models for slice-to-volume registration. Comparison is performed using the monomodal heart dataset (see Section 3.2.1 for a complete description). The aim of this experiment was to choose the most accurate method for deformable registration which (together with the standard rigid model) was then used as baseline for comparison with the discrete approaches proposed in this work (see Section 3.2).

Following the literature on slice-to-volume registration (for a complete survey on slice-to-volume registration see [11]), we adopted a decoupled model where the transformation consists in a global 6-DOF rigid transformation for plane selection, and a 2D FFD to represent the deformation field. To account for smooth deformations, the FFD is regu-

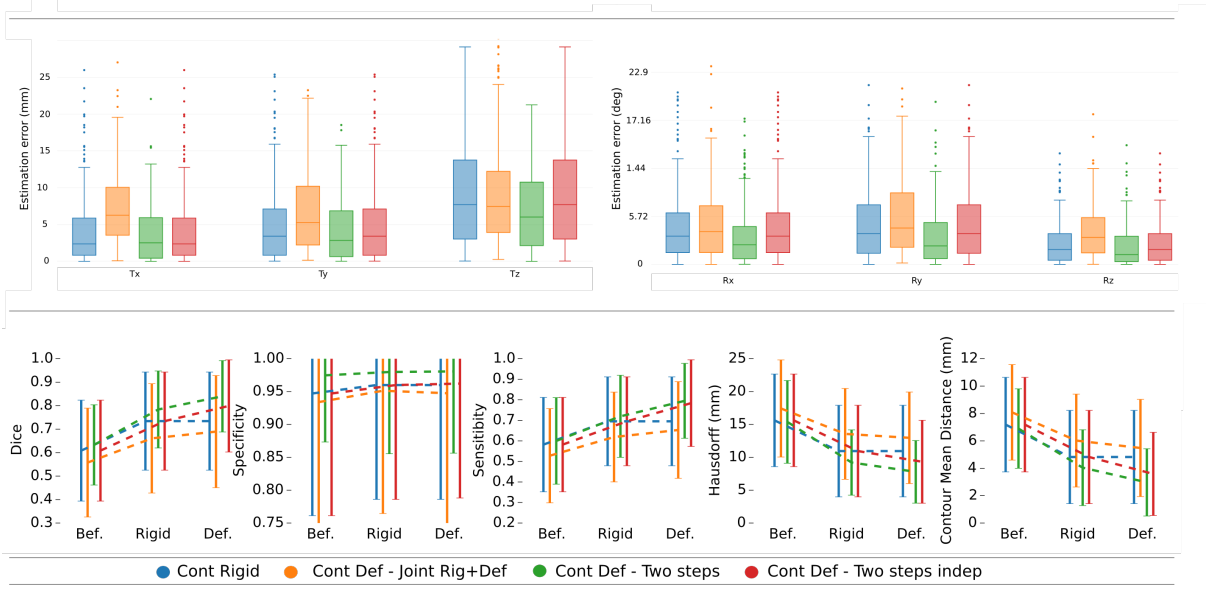


Figure 16 Comparison among four different slice-to-volume registration models optimized using continuous optimization (Nelder-Mean simplex algorithm). We compare a simple 6-DOF rigid transformation (*Cont Rigid*) and three variants of a decoupled model with a global 6-DOF rigid transformation and a 2D FFD. In the first variant (*Cont Def - Joint Rig+Def*), rigid and deformable parameters are optimized jointly. In the second case (*Cont Def - Two Steps*) a two steps strategy is adopted: first, only rigid parameters are optimized until convergence; then, both rigid and deformable parameters are optimized jointly. In the last case (*Cont Def - Two Steps Indep*), a two steps strategy is also adopted, the difference being that when optimizing the deformable parameters, the rigid ones are not modified. The *Cont Def - Two Steps* model outperforms the others according to all the metrics (distance between rigid transformations, Dice coefficient, specificity, sensitivity, Hausdorff and CMD.)

larized using the Jacobian of the deformation field, a common regularizer used in the deformable image registration community [43]. Optimization was performed through the continuous Nelder-Mead simplex algorithm described in Section 3.2, adopted in many slice-to-volume registration studies (see Section 4.1.2 of [11]). The grid resolution for the 2D FFDs was set to be equivalent to the resolutions used in the discrete experiments (see Sections 3.2.1 and 3.2.2). We run simplex optimization until convergence or until a maximum of 10000 simplex iterations were achieved. For the rigid model, convergence was always reached in a few seconds. In case of the deformable models, the algorithm did not converge in all the cases, achieving maximum running times of around 40 seconds for 10000 iterations. We experimented with more iterations (100000) but we did not reach significant improvements in the results.

Figure 16 summarizes the results for the comparative study including a simple 6-DOF rigid transformation and three variants of a decoupled model with a global 6-DOF rigid transformation and a 2D FFD. As it can be observed, the *Cont Def - Two Steps* outperforms the other models. That is why it was chosen as baseline for comparison with the discrete approaches proposed in this work.

5 Acknowledgements

This research was partially supported by European Research Council Starting Grant Diocles (ERC-STG-259112). We thank Mihir Sahasrabudhe for proof-reading the paper, and Puneet Kumar Dokania, Vivien Fecamp and Jorg Kappes for helpful discussions.

References

1. Andres, B., Kappes, J.H., Beier, T., Köthe, U., Hamprecht, F.A.: The lazy flipper: Efficient depth-limited exhaustive search in discrete graphical models. In: Computer Vision–ECCV 2012, pp. 154–166. Springer (2012)
2. Baudin, P.Y., Goodman, D., Kumar, P., Azzabou, N., Carlier, P.G., Paragios, N., Kumar, M.P.: Discriminative parameter estimation for random walks segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 219–226. Springer (2013)
3. Besag, J.: On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B (Methodological) pp. 259–302 (1986)

4. Birkfellner, W., Figl, M., Kettenbach, J., Hummel, J., Homolka, P., Schernthaner, R., Nau, T., Bergmann, H.: Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic ct images. Medical Physics **34**(1), 246 (2007). DOI 10.1118/1.2401661
5. Chandler, A.G., Pinder, R.J., Netsch, T., Schnabel, J.A., Hawkes, D.J., Hill, D.L., Razavi, R.: Correction of misaligned slices in multi-slice mr cardiac examinations by using slice-to-volume registration. In: Journal of Cardiovascular Magnetic Resonance2008, 10:13 (2008)
6. Eresen, A., Li, P., Ji, J.X.: Correlating 2D histological slice with 3D MRI image volume using smart phone as an interactive tool for muscle study. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, pp. 6418–6421. IEEE (2014)
7. Fei, B., Duerk, J.L., Wilson, D.L.: Automatic 3D registration for interventional MRI-guided treatment of prostate cancer. Computer Aided Surgery **7**(5), 257–267 (2002)
8. Ferrante, E., Fecamp, V., Paragios, N.: Implicit planar and in-plane deformable mapping in medical images through high order graphs. In: IEEE International Symposium on BIOMEDICAL IMAGING: From Nano to Macro (ISBI) (2015)
9. Ferrante, E., Fecamp, V., Paragios, N.: Slice-to-volume deformable registration: Efficient one shot consensus between plane selection and in-plane deformation. International Journal of Computer Assisted Radiology and Surgery (IJCARS) p. 16 (2015)
10. Ferrante, E., Paragios, N.: Non-rigid 2D-3D medical image registration using markov random fields. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 163–170. Springer (2013)
11. Ferrante, E., Paragios, N.: Slice-to-volume medical image registration: a survey. Medical Image Analysis (2017)
12. Gill, S., Abolmaesumi, P., Vikal, S., Mousavi, P., Fichtinger, G.: Intraoperative prostate tracking with slice-to-volume registration in MRI pp. 154–158 (2008)
13. Glocker, B.: Random Fields for Image Registration. PhD Thesis (2010)
14. Glocker, B., Komodakis, N., Paragios, N., Navab, N.: Approximated curvature penalty in non-rigid registration using pairwise MRFs. In: Advances in Visual Computing, pp. 1101–1109. Springer (2009)
15. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical Image Analysis **12**(6), 731 – 741 (2008). DOI 10.1016/j.media.2008.03.006. Special issue on information processing in medical imaging 2007
16. Glocker, B., Sotiras, A., Komodakis, N., Paragios, N.: Deformable medical image registration: setting the state of the art with discrete methods. Annu Rev Biomed Eng **13**, 219–244 (2011). DOI 10.1146/annurev-bioeng-071910-124649
17. Huang, X., Moore, J., Guiraudon, G., Jones, D.L., Bainbridge, D., Ren, J., Peters, T.M.: Dynamic 2D ultrasound and 3D ct image registration of the beating heart. Medical Imaging, IEEE Transactions on **28**(8), 1179–1189 (2009)
18. Jiang, S., Xue, H., Counsell, S., Anjari, M., Allsop, J., Rutherford, M., Rueckert, D., Hajnal, J.V.: Diffusion tensor imaging (dti) of the brain in moving subjects: Application to in-utero fetal and ex-utero studies. Magnetic Resonance in Medicine **62**(3), 645–655 (2009)
19. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Lellmann, J., Komodakis, N., Rother, C.: A comparative study of modern inference techniques for discrete energy minimization problem (2013)
20. Kim, B., Boes, J.L., Bland, P.H., Chenevert, T.L., Meyer, C.R.: Motion correction in fMRI via registration of individual slices into an anatomical volume (1999)
21. Kohli, P., Rother, C.: Higher-order models in computer vision. Chapter 1, Image Processing and Analysis with Graphs (CRC Press) (2012)
22. Komodakis, N., Paragios, N., Tziritas, G.: MRF energy minimization and beyond via dual decomposition. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on (2011)
23. Komodakis, N., Tziritas, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic MRFs. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pp. 1–8. IEEE (2007)
24. Komodakis, N., Xiang, B., Paragios, N.: A framework for efficient structured max-margin learning of high-order MRF models. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) (2015)
25. Kschischang, F.R., Frey, B.J., Loeliger, H.A.: Factor graphs and the sum-product algorithm. Information Theory, IEEE Transactions on **47**(2), 498–519 (2001)
26. Kwon, D., Lee, K.J., Yun, I.D., Lee, S.U.: Nonrigid image registration using dynamic higher-order MRF model. In: Computer Vision–ECCV 2008, pp. 373–386. Springer (2008)
27. Leung, K.Y.E., van Stralen, M., Nemes, A., Voormolen, M.M., van Burken, G., Geleijnse, M.L., Ten, F.J., Cate, J., Reiber, H.C.: Sparse Registration for Three-Dimensional Stress Echocardiography (November 2015) (2010)
28. Liao, R., Zhang, L., Sun, Y., Miao, S., Chefd'Hotel, C.: A review of recent advances in registration techniques applied to minimally invasive therapy. Multimedia, IEEE Transactions on **15**(5), 983–1000 (2013)
29. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3D/2D registration methods for image-guided interventions. Medical image analysis (2012)
30. Mercier, L., Del Maestro, R.F., Petrecca, K., Araujo, D., Haegelen, C., Collins, D.L.: Online database of clinical mr and ultrasound images of brain tumors. Medical Physics **39**, 3253 (2012)
31. Murphy, K.P., Weiss, Y., Jordan, M.I.: Loopy Belief Propagation for Approximate Inference: An Empirical Study. In: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (UAI'99), (1999)
32. Nelder, J.A., Mead, R.: A Simplex Method for Function Minimization. The Computer Journal **7**(4), 308–313 (1965). DOI 10.1093/comjnl/7.4.308
33. Olesch, J., Beuthien, B., Heldmann, S., Papenberg, N., Fischer, B.: Fast intra-operative non-

linear registration of 3D-ct to tracked, selected 2D-ultrasound slices. In: SPIE Medical Imaging, pp. 79,642R–79,642R. International Society for Optics and Photonics (2011)
34. Osechinskiy, S., Kruggel, F.: Slice-to-Volume Nonrigid Registration of Histological Sections to MR Images of the Human Brain. Anatomy Research International **2011**, 1–17 (2011). DOI 10.1155/2011/287860
35. Paragios, N., Ferrante, E., Glocker, B., Komodakis, N., Parisot, S., Zacharaki, E.I.: (Hyper)-graphical models in biomedical image analysis. Medical Image Analysis (2016)
36. Paragios, N., Komodakis, N.: Discrete Visual Perception. In: Pattern Recognition (ICPR), 2014 22nd International Conference on, pp. 18–25. IEEE (2014)
37. Porchetto, R., Stramana, F., Paragios, N., Ferrante, E.: Rigid slice-to-volume medical image registration through markov random fields. BAMBI Workshop, MICCAI 2016 (2016)
38. Ramalingam, S., Kohli, P., Alahari, K., Torr, P.H.: Exact inference in multi-label crfs with higher order cliques. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp. 1–8. IEEE (2008)
39. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast mr images. Medical Imaging, IEEE Transactions on **18**(8), 712–721 (1999)
40. San José Estépar, R., Westin, C., Vosburgh, K.: Towards real time 2D to 3D registration for ultrasound-guided endoscopic and laparoscopic procedures. International journal of computer assisted radiology and surgery **4**(6), 549–560 (2009)
41. Seshamani, S., Fogtmann, M., Cheng, X., Thomason, M., Gatenby, C., Studholme, C.: Cascaded slice to volume registration for moving fetal fMRI. In: Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, pp. 796–799. IEEE (2013)
42. Shekhovtsov, A., Kovtun, I., Hlaváč, V.: Efficient MRF deformation model for non-rigid image matching. Comput. Vis. Image Underst. **112**(1), 91–99 (2008)
43. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. Medical Imaging, IEEE Transactions on **32**(7), 1153–1190 (2013)
44. Sotiras, A., Ou, Y., Glocker, B., Davatzikos, C., Paragios, N.: Simultaneous geometric-iconic registration. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, pp. 676–683. Springer (2010)
45. Wang, C., Komodakis, N., Paragios, N.: Markov random field modeling, inference & learning in computer vision & image understanding: A survey. Computer Vision and Image Understanding **117**(11), 1610–1627 (2013)
46. Xu, H., Lasso, A., Fedorov, A., Tuncali, K., Tempany, C., Fichtinger, G.: Multi-slice-to-volume registration for MRI-guided transperineal prostate biopsy. International journal of computer assisted radiology and surgery pp. 1–10 (2014)
47. Zikic, D., Glocker, B., Kutter, O., Groher, M., Komodakis, N., Kamen, A., Paragios, N., Navab, N.: Linear intensity-based image registration by markov random fields and discrete optimization. Medical Image Analysis **14**(4), 550–562 (2010)