Simultaneous Linear and Deformable Registration Through a Higher Order MRF Model

Vivien Fécamp\textsuperscript{1,2}, Aristeidis Sotiras\textsuperscript{3}, and Nikos Paragios\textsuperscript{1,2}

\textsuperscript{1} CVC, Ecole CentraleSupélec, France
\textsuperscript{2} GALEN Team, INRIA Saclay \texttt{vivien.fecamp,nikos.paragios@inria.fr}
\textsuperscript{3} Section of Biomedical Image Analysis University of Pennsylvania \texttt{aristeidis.sotiras@uphs.upenn.edu}

Abstract. In this paper, we present a new approach to tackle simultaneously linear and deformable registration between a pair of images. Our combined formulation avoids the bias created when linear registration is performed independently before a deformable registration. Our registration problem is formulated as a discrete Markov Random Field employing a higher order objective function. To cope with both linear and deformable registration, we introduce two graphical models, one for each subproblem. The two graphical models consist of identical node systems. The nodes of the first graph encode the local translations of the global linear registration component, while the nodes of the second graph encode the local translations of the deformable registration component. A higher-order edge system (third and fourth order interactions) that imposes the linearity of the transformation is introduced for the first graph, while a simple pairwise edge system that promotes the smoothness of the deformation field is employed for the second graph. The two graphs are coupled by additional edges that connect homologous nodes and encode the data term, while unary potentials are used only for the deformable part and penalize large deformations. The resulting formulation is modular with respect to the image metric used to evaluate the correctness of mapping as well as with respect to the nature of the linear transformation (rigid, similarity, affine). Inference on this graph is performed efficiently through Alternating Direction Method of Multipliers. Promising results on medical 3D images demonstrate the potentials of our approach.

Keywords: Registration, Markov Random Fields, Higher Order Potentials, Alternating Direction Method of Multipliers

1 Introduction

Linear [9] and deformable registration [7, 19] are among the computational pillars of medical image analysis. Linear methods aim to establish approximate
correspondences using global models (rigid, similarity, affine, etc.), whereas deformable methods seek a one-to-one mapping between the images.

Linear registration is either geometric or iconic. On the one hand geometric methods seek a global transformation that corresponds to the lowest potential of the Euclidean distance between corresponding points. These methods are not within the scope of the paper and therefore will not be reviewed. On the other hand, iconic methods aim to determine the set of parameters of a given linear transformation that minimizes an image-based objective function. Global methods adopt often standard similarity metrics like Sum of Absolute Differences (SAD), Normalized Cross Correlation (NCC), Sum of Squared Differences (SSD), Mutual information (MI) and determine the optimal transformation through a gradient-driven optimization method. In order to cope with local minima and reduce the dependency on the initial conditions, optimization approaches like simplex methods [17], cutting planes methods [11] or more recently discrete optimization ones [22], have been investigated. Jenkinson et al. [9] uses a hybrid scheme that combines both local and global search.

Once linear mapping has been addressed, deformable registration is used to provide dense correspondences. These methods are in most of the cases image-based and one can refer to an important number of successful developments in the recent years [19]. Without attempting to give a full overview of this rich field, let us note that most methods can be classified based on the optimization technique they employ. The first class of methods regroups registration approaches that solve the problem by means of continuous optimization techniques. Typical examples of this class include the Demons algorithm [20] that iterates between estimating correspondences and smoothing, Free-Form Deformations (FFD) registration [16] using gradient descent, and Advanced Normalization Tools (ANTS) [2]. The algorithm used in ELASTIX [10] is based on a B-spline deformation model and is a good compromise between complexity and performance. The second class of registration methods consists of approaches that quantize the solution space and use discrete optimization techniques. DROP [6] is a modular, metric free, computationally efficient approach to deformable registration. DRAMMS [15] use MRF solvers while [3] uses random walker approaches.

Linear and deformable registration have been considered separately up to now. Deformable registration is usually built upon the result of the linear one. Therefore, one can expect that a bias is introduced in the deformable registration by the linear transformation. Moreover, the performance of deformable registration depends heavily on the input of the linear one and might fail to provide appropriate correspondences. Coupling the two problems would remove the bias and by allowing them to share information better results may be achieved. To the best of our knowledge, only three previous works exist. Haber et al. [8] use continuous optimization which causes problems with the similarity criterion. Kwon et al. [14] introduce a curvature regularisation as a soft constraint and solve the problem on the factor graph with the tree-reweighted algorithm but there is no possible reconstruction of the rigid part of the transform. Their regu-
larisation is seen as a penalty to get a soft deformation field, and not to have an exact mathematical decomposition. Finally Ferrante et al. [5] presented a similar decomposition to perform 2D-3D registration, but with a soft higher order constraint again. In this paper, we solve the problem in the discrete case which allows us to be modular with the metric space and the regularization constraint.

In this paper, we introduce a novel graphical model that consists of two interconnected components and estimates simultaneously the two registration components by considering a local deformation approach. The first graph consists of a regular grid where higher order constraints between nodes impose the expected nature of linearity of the transformation as defined in our previous work [4]. The second graph - inspired by the one proposed in [6] - adopts an identical grid endowed with singleton terms that penalize the magnitude of displacements and pairwise constraints imposing deformation smoothness. The two graphs are interconnected and in these edges the exact data term is modeled through the composition of the two transformations. The resulting formulation can deal with arbitrary types of linear mapping, arbitrary similarity criteria and various regularization terms. The optimization of this graphical model is performed through a master-slave framework that is based on the dual decomposition with the Alternating Direction Method of Multipliers (DD-ADMM) approach [1].

The remainder of the paper is organized as follows: Sect. 2 presents the graphical model formulation for the linear mapping, that is endowed with the deformable component. The optimization of the complete framework and the associated implementation details are presented in Sect. 2.3. Sec 3 presents implementation details and experimental validation. The last section concludes the paper and provides future directions.

2 Method

The registration problem consists in finding a transformation $T$ that aligns an image $J$ (typically referred to as source) to a reference image $I$ (typically referred to as target). A common approach for modeling this problem is by energy minimization:

$$\hat{T} = \arg\min_T \xi (I, J \circ T),$$  \hspace{1cm} (1)

where $\hat{T}$ is the optimal transformation and $\xi$ is a similarity measure. Our purpose is to model the image transformation through the displacements of a small set of control points. More specifically, we consider two grid-based deformation systems that model the affine and the deformable components of the registration, respectively. The two grids of nodes are isomorphic. A control point is a point in the image whose displacement we are looking for. A control point is not a node, but the control points also form a grid superimposed on the image. The displacement of a control point is encoded into two nodes, one for the linear part, the other for the deformable part. The nodes, edges, and hyperedges form
We aim to find the optimal displacements of the control points, by making use of Markov Random Fields (MRF) theory on the graph.

Therefore, we will formulate (1) as a discrete label assignment problem with the use of Markov Random Fields theory.

We model the simultaneous linear and deformable registration problem through a hypergraph \( G = (V, E, C) \), where \( V \) denotes the set of nodes, \( E \) the set of edges, and \( C \) the set of higher order cliques. Let \( L = \{l_1, \ldots, l_n\} \) be the set of labels that corresponds to a quantized version of the solution space, which means a label is associated to a displacement vector, and \( l_p \) denotes the label assigned to node \( p \). The goal of the proposed model is to assign a label \( l_p \) to each node \( p \) such that the two images get aligned. The energy of the MRF can be written as:

\[
E_{MRF} = \sum_{p \in V} U_p(l_p) + \sum_{(p,q) \in E} V_{p,q}(l_p, l_q) + \sum_{c \in C} H_c(l_c),
\]

(2)

where \( U_p(l_p) \) denotes the unary potentials, \( V_{p,q}(l_p, l_q) \) denotes the binary potentials, \( l_c = \{l_p, p \in c\} \), is the set of labels assigned to the nodes in the clique \( c \) and \( H_c(l_c) \) denotes the higher order potentials.

The alignment is measured by a cost function. The choice of the cost function depends on the modalities of the two images as well as on our assumption regarding the intensity relationship that relates them. Our framework is modular with respect to the choice of the similarity function. The only mild requirement regarding this choice is that it can be evaluated over a patch. We superimpose a grid of control points over the image domain (see Fig. 1a). Patches are naturally defined in our framework by taking into account the grid-based deformation model. The deformation grid gives rise to patches that are centered around the control points of the grid. The size of these patches depends on the deformation grid resolution and changes along the iterations of the algorithm.

We want to separate and determine simultaneously the linear and deformable transformations. As mentioned previously, we decompose the displacement of each control point into a linear and a deformable part. Thus our graph contains twice as many nodes as control points in the image. Our graph contains two parts: the first part will encode the linear displacements, the second part will encode the deformable displacements (see Fig. 1b). Therefore, each control point is represented by two corresponding nodes of the graph. Each node is assigned a label which corresponds to a displacement vector. The resulting displacement of a control point is then the sum of the displacements of its linear and its deformable parts.

To get a decomposition of the linear and deformable displacements, we needed two displacement vectors. We consider \( L_d \) potential displacement vectors. If we had put them in the same nodes, our number of labels per node would be \((L_d)^2\). Our decomposition allows to keep a reasonable number of nodes (twice as many) and labels \((L_d)\), and therefore to greatly decrease the computational cost of performing inference in the graph. A similar decomposition was used in [18] for the x- and y- axes. Here we use the same idea but for the linear and non-linear parts. Let us note \( V1 \) the nodes in the first part of the duplicated graph (linear
part), and $V_2$ those in the second part (deformable part):

$$V = V_1 \cup V_2. \quad (3)$$

The displacement of each control point is the sum of the two displacements vectors, or labels, associated to the two nodes, one in $V_1$ and one in $V_2$ associated to the control point. The labels of all the nodes in $V_1$ give the affine transformation, those in $V_2$ the deformable field and the sum corresponds to the whole registration.

### 2.1 Graph Construction

Let us now define $C$, the set of cliques. In our framework, the cliques have very different goals. Edges connecting nodes in $V_2$ ensure the smoothness of the deformable displacements. So there is an edge between each pair of neighbouring nodes, which form a grid that is similar to the one in [16], that was used to compute deformable deformations. The cliques in $V_1$ are higher order edges, and they ensure that the linear displacements of all the points form a coherent linear transformation of the image. Finally, the data term should capture the interactions between pairs of linear and deformable displacements so each pair of duplicated nodes (one in $V_1$, one in $V_2$) will be linked by an edge.

**Unary Potentials.** To ensure the algorithm prefers large linear displacements instead of large deformable ones, we employ a unary potential penalizing the
norm of the vector of the displacement vector for nodes in the deformable part $V_2$. 

$$U_p(l_p) = \|l_p\| \quad \forall p \in V_2. \quad (4)$$

This potential is defined for every node in $V_2$, where $l_p$ corresponds to a deformable displacement. There are no unary potentials for the nodes in $V_1$.

**Binary Potentials.**

*Smoothing Term.* A regularization term operating between nodes in $V_2$ is necessary in order to ensure the deformable registration is smooth. This can be achieved by penalizing the vector differences between neighbouring nodes:

$$V_{p,q}(l_p, l_q) = \frac{\| (q + l_q) - (p + l_p) \|}{\| q - p \|} \quad \forall (p, q) \in V_2^2, \quad (5)$$

where $p$ and $q$ represent two neighbouring nodes, both in $V_2$.

*Data Term.* In order to quantify the alignment of the two images, we employ a patch-based similarity criterion, i.e. we compare a patch from the source image $B_{p,q}$ with a patch in the target domain $B_{l_p+l_q}$. This patch is centered at one control point but its displacement is defined by two nodes, $p \in V_1, q \in V_2$. These represent the concatenation of the affine and deformable part of the deformation. The data term is defined as:

$$V_{p,q}(l_p, l_q) = \rho(B_{p,q}B_{l_p+l_q}) \quad \forall (p, q) \in V_1 \times V_2. \quad (6)$$

There is no constraint imposed on the choice of the matching criterion $\rho$. The proposed model can encompass a wide choice of intensity-based similarity measures, from the sum of absolute difference (SAD) to statistical measures for multimodal registration like mutual information [21]. There are no pairwise edges between nodes in $V_1$. Instead, nodes in $V_1$ are connected through hyperedges (see next paragraph).

**Higher Order Terms.** The higher order potentials are defined as in [4]. Triplets and a special clique called $\lambda$-clique ensure the linearity of the transformation. $V_1$ is a grid of points with the shape of a cube in 3D (or a square in 2D). A triplet regroups any set of three neighbouring nodes in $V_1$ which form a line parallel to one of the axes of the cube ($x$, $y$, $z$). A $\lambda$-clique is an hyperedge containing four nodes in the shape of a $\lambda$. There is one $\lambda$-clique on each face of the cube in 3D, so six in total (and only one in 2D). An example of those cliques for a 2D grid used for a 2D registration is shown in Fig. 2a. For a triplet $c = p, q, r$ and their respective labels $l_p, l_q, l_r$, the higher order potential is defined as:

$$H_c(l_c) = \begin{cases} 
0 & \text{if } (l_p + l_r - 2l_q) = 0 \\
\infty & \text{otherwise} 
\end{cases} \quad (7)$$
The λ-clique contains four points p, q, r, and s. The points p, q, r satisfy the same condition as the previously mentioned triplet. They are not part of the former set because they are aligned along a diagonal, not an axis of the cube. If we are searching for a rigid transformation, or for a similarity, an additional constraint is imposed in the λ-clique. It is detailed in [4]. If we are looking for an affine transformation, we do not have to add anything else. These higher order potentials alone ensure the global transformation estimated by the linear part of the graph V1 lies within the requested set of transformations (affine, rigid, similarity).

2.2 Optimization algorithm

To solve the MRF, we use DD-ADMM [1]. Dual Decomposition [12] consists in decomposing a global difficult problem into smaller solvable subproblems (referred to as slaves) and then extracting a solution by cleverly combining the solutions from these subproblems. The only requirement for the choice of the subproblems is that they cover (at least once) every node, edge, and hyperedge of the graph that models the problem. DD-ADMM improves Dual Decomposition by accelerating its convergence. In our experiments, Dual Decomposition needed too many iterations (more than 10000 iterations) and sometimes it never agreed to a consensus while DD-ADMM reached a consensus generally between 100 and 1000 iterations. Other optimisation algorithms performed badly in our experiments, mainly because of the hard constraint on the higher order terms. The time required for the optimisation was more than one day. In our case, the difficulty of the inference of the optimization displacements lies in the presence of the higher order cliques. Here, the graph is decomposed into a grid and into trees that constitute the set of subproblems. The slave problems can be
solved independently from one another, thus allowing for the parallelization of the computation.

In our case, in the affine part of the graph, a slave problem is defined for each line parallel to a coordinate axis in the first part of the grid, and a slave for each \( \lambda \)-clique. An example of the different slaves in 2D is illustrated in Fig. 2b. One of the slave contains all the nodes of the graph but only the edges corresponding to the smoothing term and the edges encoding the data term. In this slave, the nodes in \( V_1 \) only belong to one edge. Thus we can send a message from this node to the other end of the edge, like in Message-Passing algorithm, to put all needed information into the unary of the node in \( V_2 \). Then we have a simple slave and we optimize it using the Fast-PD algorithm [13].

3 Experimental Validation

3.1 Implementation Details

The algorithm uses an iterative coarse-to-fine refinement process. The resolution of the image is reduced at the first steps to accelerate the computation. The label space is successively refined to explore a large number of displacements while keeping a reasonable execution time. The label space corresponds to a discretization of potential displacement vectors that are densely sampled following a regular grid pattern around each node. The maximal length of the displacement vectors is 0.4 times the deformation grid spacing. The length is iteratively reduced along the iterations. We used up to 7 iterations in our experiments. The successive label space refinement allows to keep the number of labels quite small, \( 3^3 \) or \( 5^3 \), while reaching sub-millimeter registration accuracy. The grid contains \( 3^3 \) control points at the first iterations and is increased to \( 9^3 \).

The algorithm is implemented in C++. The tests were performed on a 64 bits machine with a Intel Xeon W3670 processor and 16 Go of RAM. The mean running time for 3D volumes was about 160 seconds when using SAD as the similarity criterion.

3.2 Synthetic Data

We use a database of abdomen 3D CT images, containing 6 images of the same patient at different time points. Two organs have been manually segmented by medical doctors, the sigmoid and the bladder. The image dimension is about \( 512 \times 512 \times 121 \) with a physical spacing of 0.92*0.92*4 mm, with small variations on the images. We apply 22 simulated affine transformations on one image. We then applied a small deformation field (see Fig. 4a) to the transformed image. This deformation field is small in the sense it should not contain any global linear transformation. We then try to register these deformed images (see Fig. 4b) to the original one. Simulated rotations lie between 0˚ and 5˚ and translations reach 20mm. Mean initial error was about 10mm. We used the Sum of Absolute Differences (SAD). We want to compare the affine transformation we find
with the one we initially applied. Therefore we fixed 6 points in the images at some extremities of the bodies, and compute the mean distance between the two transformations. Our results show a mean distance of 2.61 mm. Most of the error comes from rotations which are not captured by the data term. The results could be improved by using a rotation invariant measure. One example of registration is shown in Fig. 5.

3.3 Clinical Data

We then use intra-patient images from the same database to compare our method with a sequential linear and deformable registration. Images are thus initially aligned with a linear registration. Then we apply a deformable registration algorithm, DROP [6]. In parallel, we apply our algorithm. The same set of parameters was used in the two cases. We compare the DICE we get from the two methods. Our results show a small improvement (cf. 1) of the DICE. Apparently, there is no bias for those images, but those results demonstrate our algorithm is performing well while by construction it has no bias compared to the standard two-step registration approach.

4 Conclusion

In this paper, we have proposed a discrete MRF formulation to solve the problem of simultaneous linear and deformable registration. The proposed formulation is metric-free (can deal with arbitrary similarity criterion), modular with respect to the nature of the linear transformation (rigid, similarity, affine and could be extended to projective) and exhibits computational efficiency due to
Fig. 4: a) A slice of the deformation field applied to the image. b) One example of an image transformed by the composition of simulated affine and deformable deformations.

<table>
<thead>
<tr>
<th></th>
<th>Bladder</th>
<th>Sigmoid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before registration</td>
<td>45.61</td>
<td>39.383</td>
</tr>
<tr>
<td>Consecutive registration</td>
<td>78.15</td>
<td>68.55</td>
</tr>
<tr>
<td>Our registration</td>
<td>78.47</td>
<td>68.64</td>
</tr>
</tbody>
</table>

Table 1: DICE overlap ratio obtained for two organs. Results before and after registration using the proposed framework and a standard two-step registration pipeline are shown.

its relative local nature and the designed search space. The performance of the method on 3D multi-modal medical data demonstrate its potential for applications. Contrary to the commonly used sequential linear/deformable registration, our scheme is based on a mathematical framework without any decoupling assumption which prevents from a potential bias. Despite the fact that it does not show significant improvement in our experiments, it performs as well as usual methods and would be able to tackle a potential bias where it appears. Moreover this approach is fast compared to state of the art methods.

Furthermore, mapping from 2D to 3D is a great problem of important interest either in vision or in medical imaging towards image-based navigation/guidance. The same concept that was proposed here to decompose 2D-2D or 3D-3D deformations in linear and non-linear components can be also applied to 2D-3D. The clinical impact of such a component in computer assisted surgery is currently under investigation.
Fig. 5: A registration, the two images are superimposed in different colours: (a) Before registration. (b) After registration.

**Acknowledgments** This research was supported by a DGA-MRIS scholarship and partially by European Research Council Starting Grant Diocles (ERC-STG-259112).

**References**