



Non-Rigid Image Registration using a Hierarchical Partition of Unity Finite Element Method

Sherif Makram-Ebeid, Oudom Somphone

► To cite this version:

Sherif Makram-Ebeid, Oudom Somphone. Non-Rigid Image Registration using a Hierarchical Partition of Unity Finite Element Method. IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007., Oct 2007, Rio de Janeiro, Brazil. pp.1-8, 10.1109/ICCV.2007.4409078 . hal-00439374

HAL Id: hal-00439374

<https://hal.science/hal-00439374v1>

Submitted on 7 Dec 2009

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

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



HAL Authorization

Non-Rigid Image Registration using a Hierarchical Partition of Unity Finite Element Method

Sherif Makram-Ebeid and Oudom Somphone
Philips Medical Systems Research Paris
51, rue Carnot, F92156 Suresnes Cedex, France
sherif.makram-ebeid@philips.com

Abstract

We use a Hierarchical Partition of Unity Finite Element Method (H-PUFEM) to represent and analyse the non-rigid deformation fields involved in multidimensional image registration. We make use of the Ritz-Galerkin direct variational method to solve non-rigid image registration problems with various deformation constraints. In this method, we directly seek a set of parameters that minimizes the objective function. We thereby avoid the loss of information that may occur when an Euler-Lagrange formulation is used. Experiments are conducted to demonstrate the advantages of our approach when registering synthetic images having little of or no localizing features. As a special case, conformal mapping problems can be accurately solved in this manner. We also illustrate our approach with an application to Cardiac Magnetic Resonance temporal sequences.

1. Introduction

There are various motivations for making use of non-rigid image registration depending on the application field. Often, one just wishes to find a ‘reasonably smooth’ warping law which is such that a measure of dissimilarity between a reference image R and the warped template image T is made as small as possible [1, 2]. However, it is sometimes useful to obtain an estimate of the warping law itself. This is particularly hard if one cannot unambiguously match corresponding localizing features in the images T and R . For example, one may wish to register a pair of binary images containing objects with smooth contours and few or no corners. The corresponding edges are curved lines in 2D or smooth surfaces in 3D that should be matched together. Given a particular edge point in R , one does not know off-hand to which edge point in T it should be associated. The related difficulty – known as the *aperture problem* [3] –

does not have a satisfactory solution but it can at least be alleviated by injecting prior knowledge when estimating the warping field. One way for doing this is to seek the warping law that minimizes an objective function made up of a sum of terms. One term measures the dissimilarity between the reference image and the warped template and the others introduce soft regularization constraints which should bias the warping law towards a behaviour compatible with domain knowledge. A reliable strategy has to be defined in order to minimize this objective function without getting trapped in local minima.

In Medical Imaging, one often wishes to get a reliable warping field estimate from an image sequence (2D or 3D) in order to deduce the strain to which an organ has been subjected. For example, it is desirable to capture the myocard strain including cardiac wall thickening from a temporal sequence of heart images (in 2D or 3D). In such cases, the regularization constraint should provide robustness against noise and other imaging artefacts without introducing an artificial bias in the estimated warping fields.

A large range of variational approaches exist today to deal with *non-rigid intensity-based image registration*. Many authors reduce the problem to solving the corresponding Euler-Lagrange Partial Differential Equation (EL-PDE) using a finite difference scheme (see for example [1] and the references therein). Some authors make use of a mesh-based Finite Element Method (FEM) to solve the EL-PDE (see for example [5]). Others represent the warping field with B-splines or a block wise affine model and express the objective function in terms of their representation parameters. They then directly seek the parameters minimizing their objective function (see for example [6, 7] and the references therein). In the present paper, we propose a new formulation of Non-Rigid Image Registration Variational problems inspired from the work of Melenk and Babuška on the Partition of Unity Finite Element Method [8, 9] which provides a generic framework for solving a

large class of PDEs. As in [6, 7], we look for the parameters of our model that directly optimize the objective function. Such a direct method is referred to as the Ritz-Galerkin approach in the FEM terminology.

Unlike methods using finite differences for solving PDEs, our problem formulation does not require imposing unnatural conditions on the boundaries of the image domain (of Dirichlet or Neumann type). Among other advantages, this allows us to get reliable warping field estimations even in the neighbourhood of the domain boundaries.

2. PUFEM Scalable Representation of the Warping Field

Let $\Omega \subset \mathbb{R}^d$ be a rectangular image domain and let us consider a set of points, called *nodes*, distributed over Ω . To each node n , we associate a compact subdomain $\Omega_n \subset \mathbb{R}^d$, and a non-negative window function $\varphi^{(n)}(\underline{x})$ which vanishes outside Ω_n . The *Partition of Unity* condition imposes:

$$\Omega \subset \bigcup_n \Omega_n \quad \text{and} \quad \forall \underline{x} \in \Omega, \quad \sum_n \varphi^{(n)}(\underline{x}) = 1 \quad (1)$$

The warping field is defined as a mapping ($\underline{x} \mapsto \underline{x} + \underline{U}(\underline{x})$) associating a point $\underline{x} + \underline{U}(\underline{x})$ in the domain of the template image T to any given point $\underline{x}$ in Ω , the domain of the reference image R . Each component $U_i(\underline{x})$ of the displacement vector field $\underline{U}(\underline{x}) \equiv (U_1, \dots, U_d)^T$ is approximated by a blending of local polynomials in the form:

$$\tilde{U}_i(\underline{x}) = \sum_n \varphi^{(n)}(\underline{x}) \tilde{U}_i^{(n)}(\underline{x}) \quad (2)$$

where each local polynomial $\tilde{U}_i^{(n)}(\underline{x})$ is expanded in a local basis $\{v_r^{(n)}\}$ as:

$$\tilde{U}_i^{(n)}(\underline{x}) = \sum_r A_{i,r}^{(n)} v_r^{(n)}(\underline{x}) \quad (3)$$

where the $A_{i,r}^{(n)}$ are the scalar coefficients of the expansion.

For any point $\underline{x} \in \Omega$, there are at most M window functions such that $\varphi^{(n)}(\underline{x}) \neq 0$. For the sake of computational efficiency, we choose our nodes over a regular rectangular array with inter-node spacing h_i along the i^{th} coordinate axis ($i = 1, \dots, d$). In our implementation, we have $M = 2^d$ as discussed below. Within a window n , the coordinates x_i of a point $\underline{x}$ are replaced by normalized local coordinates $\zeta_i^{(n)} = (x_i - \xi_i^{(n)})/h_i$ for which the window centre coordinates $\xi_i^{(n)}$ are taken as origin.

Each window function $\varphi^{(n)}(\underline{x})$ is taken as the product of single-variable piecewise polynomial functions:

$$\varphi^{(n)}(\underline{x}) = \prod_{i=1}^{i=d} P(\zeta_i^{(n)}) \quad (4)$$

The function $P(\zeta)$ is designed in such a manner that $P(\zeta) + P(1-\zeta) = 1$ and $P(-\zeta) = P(\zeta)$ for $0 \leq \zeta \leq 1$, $P(\zeta) = 0$ for $\zeta < -1$ or $\zeta > 1$. A piecewise polynomial expression for $P(\zeta)$ which has a prescribed C^k continuity ($k \geq 0$) is readily found. For C^0 continuity: $P(\zeta) = 1 - |\zeta|$, for C^1 continuity: $P(\zeta) = 1 - 3|\zeta|^2 + 2|\zeta|^3$. Defining the window functions from Eq.(4) is advantageous because they are simple to compute and they automatically satisfy the normalization condition $\sum_n \varphi^{(n)}(\underline{x}) = 1$ with $M = 2^d$ non-vanishing windows for any $\underline{x}$. When selecting for $P(\zeta)$ the C^0 continuous expression, we just obtain the familiar multi-linear interpolation window function (bilinear in 2D and trilinear in 3D).

The different basis functions $v_r^{(n)}(\underline{x})$ associated with node n are monomials of all orders up to a maximum p , for example, if $p = 2$ and $d = 2$, there would be six such monomials $v_r^{(n)}(\underline{x})$ ($r = 1$ to 6) defined by $\{1, \zeta_1^{(n)}, \zeta_2^{(n)}, (\zeta_1^{(n)})^2, \zeta_1^{(n)}\zeta_2^{(n)}, (\zeta_2^{(n)})^2\}$.

Global and Local Approximation Errors. We now turn to the problem of approximating a given warping field $\underline{U}(\underline{x})$ with the above representation Eqs.(2,3). We do this by looking for the coefficients $A_{i,r}^{(n)}$ that, for each component U_i of the warping field, minimize the L^2 error $Q_i = \|\underline{U}_i - \tilde{U}_i\|_{L^2}^2$ incurred when replacing the given $U_i(\underline{x})$ by this representation. Using the partition of unity property $\sum_n \varphi^{(n)}(\underline{x}) = 1$, this can be written as:

$$Q_i = \int_{\Omega} \left(\sum_n \varphi^{(n)}(\underline{x}) (U_i(\underline{x}) - \tilde{U}_i^{(n)}(\underline{x})) \right)^2 d\underline{x} \quad (5)$$

From Cauchy-Schwarz inequality, it is straightforward to prove that $(\sum_n g_n z_n)^2 \leq (\sum_n g_n)(\sum_n g_n z_n^2)$ provided $g_n \geq 0$ for all n . Applying this inequality to the integrand, we get an upper bound for the global quadratic error Q_i in the approximation of $U_i(\underline{x})$ in the form:

$$Q_i \leq \sum_n Q_i^{(n)} \quad (6)$$

where

$$Q_i^{(n)} = \int_{\Omega} \varphi^{(n)}(\underline{x}) (U_i(\underline{x}) - \tilde{U}_i^{(n)}(\underline{x}))^2 d\underline{x} \quad (7)$$

In other words, in order to obtain a small global quadratic approximation error Q_i , it is sufficient to minimize the local weighted quadratic errors $Q_i^{(n)}$ within each of the windows n . We should recall that $Q_i^{(n)}$ according to (3) depends only on the local coefficients $A_{i,r}^{(n)}$ in window n . Minimizing $Q_i^{(n)}$ therefore boils down to a standard LMS problem solvable using classical linear algebra techniques.

PUFEM Hierarchy. In order to obtain fast convergence rates and to reduce the probability of getting trapped in local minima, we have to deal with representations having different levels of detail (coarse to fine). To do this, we define a dyadic pyramid of PUFEM arrays. In our implementation, the array of a level has either identical node separation distances h_i as those h_i^{fine} of the next finer level or has exactly twice those separations ($h_i = 2h_i^{fine}$ for $i = 1, \dots, d$). All levels of the pyramid cover the same rectangular domain Ω . The order p of the polynomial approximation is also allowed to vary from one level to the other. We have to design prolongation operators allowing to convert a coarse level representation into the next finer level. We also need restriction operators in order to realize the reverse fine to coarse conversions. Both operators are realized using the least square approximation procedure defined above. In both types of prolongation and restriction operators, assume that an input displacement field $U_i(\underline{x})$ is defined by a node array with coefficients $B_{i,s}^{(m)}$, window functions $\psi^{(m)}(\underline{x})$ and monomials $w_s^{(m)}(\underline{x})$. The coefficient $A_{i,r}^{(n)}$ of the output representation are obtained by looking for minimum quadratic errors $Q_i^{(n)}$, leading to:

$$\frac{1}{2} \frac{\partial Q_i^{(n)}}{\partial A_{i,r}^{(n)}} = \sum_s K_{r,s}^{(n)} A_{i,s}^{(n)} - Z_{i,r}^{(n)} = 0 \quad (8)$$

$$\text{where } K_{r,s}^{(n)} = \int \varphi^{(n)} v_r^{(n)} v_s^{(n)} d\underline{x} \quad (9)$$

$$\text{and } Z_{i,r}^{(n)} = \sum_{s,m} \left(\int \psi^{(m)} \varphi^{(n)} w_s^{(m)} v_r^{(n)} d\underline{x} \right) B_{i,s}^{(m)} \quad (10)$$

For each output node n , the summation extends over all labels s of input monomials $w_s^{(m)}(\underline{x})$ and over all input nodes m having overlapping windows with n (i.e. such that $\psi^{(m)} \varphi^{(n)}$ is not identically zero). The resulting system in Eqs.(8) is solvable by matrix inversion because the matrices $K_{r,s}^{(n)}$ are positive definite. In Eqs.(9,10), we have dropped the explicit dependence on $(\underline{x})$ of the functions $\psi^{(m)}$, $\varphi^{(n)}$, $w_s^{(m)}$ and $v_r^{(n)}$. All matrix entries $K_{r,s}^{(n)}$ and all coefficients of $B_{i,s}^{(m)}$ in the expression for $Z_{i,r}^{(n)}$ are easily computed since they are dimensionally separable integrals of piecewise polynomials. The coefficients of the output representation $A_{i,r}^{(n)}$ can now be expressed as linear combinations of the input coefficients $B_{i,s}^{(m)}$. All matrix inversions are precomputed in order to avoid costly computations within iterative algorithms.

PUFEM Global and Local Warping Field Derivatives. Let $F(\underline{x})$ stand for any component of the warping displacement vector field $\tilde{U}_i(\underline{x})$ and let $F^{(n)}(\underline{x}) = \tilde{U}_i^{(n)}(\underline{x})$ stand for the corresponding polynomial representation in

node n . We wish to evaluate the partial derivative $\partial_j F$ of F with respect to j^{th} component of the positional vector $\underline{x}$. To do this, we need to take into account the variation of the window functions from Eq.(2). Dropping the explicit mention of the independent variable $(\underline{x})$ we get:

$$\partial_j F = \sum_n \varphi^{(n)} \partial_j F^{(n)} + \chi_j \quad (11)$$

$$\text{where } \chi_j = \sum_n \left(F^{(n)} - \sum_m \left(\varphi^{(m)} F^{(m)} \right) \right) \partial_j \varphi^{(n)}$$

as obtained by making use of the partition of unity properties $\sum_m \varphi^{(m)} = 1$ and $\sum_m \partial_j \varphi^{(m)} = 0$. The first term in the right hand side of the above equation is just the blending of local derivatives $\partial_j F^{(n)}$ using the window function weighting. The second term χ_j involves $|\partial_j \varphi^{(n)}|$ the upper bound of which can always be written as C/h where C is a positive constant and $h = \min_{i \in [1,d]} \{h_i\}$. This leads to the inequality:

$$\begin{aligned} |\chi_j(\underline{x})| &\leq \frac{C}{h} \left| \sum_n \left(F^{(n)}(\underline{x}) - \sum_m \left(\varphi^{(m)} F^{(m)}(\underline{x}) \right) \right) \right| \\ &\leq \frac{MC}{h} \max_{(m,n)} \left| F^{(m)}(\underline{x}) - F^{(n)}(\underline{x}) \right| \end{aligned} \quad (12)$$

where M is the number of non-zero window functions $\varphi^{(m)}$ at point $\underline{x}$ while (m,n) stands for any pair of element numbers for which $\varphi^{(m)}(\underline{x}) \varphi^{(n)}(\underline{x}) > 0$. Therefore $\partial_j F$ can be expressed as a weighted combination of local derivatives $\partial_j F^{(n)}$ provided that the different local approximations $F^{(m)}$ agree in regions where the element windows overlap. As noted by Melenk and Babuška [9], this is possible if we impose inter-element continuity (conformity) and absorb those extra constraints by making use of a sufficiently large approximation space (h small and/or p large).

Sobolev Non-Conformity Measures. From the above discussion, the estimation of first derivatives from the local polynomial representations is allowed provided their values differ as little as possible within overlapping window regions. More generally, one may need to estimate derivatives of higher orders $D^\alpha F = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d} F$ where $\alpha = (\alpha_1, \dots, \alpha_d)$ is the derivation order multi-index. The above analysis can be easily generalized but then, inter-element continuity of higher order derivatives of $F^{(m)}$ has to be enforced. For the representation of F , we define the non-conformity measure for any two adjacent nodes m and n as a Sobolev Norm weighted by the product $\varphi^{(m)} \varphi^{(n)}$ of window functions:

$$S_k^{(m,n)}(F) = \sum_{|\alpha| \leq k} \int \varphi^{(m)} \varphi^{(n)} \left(D^\alpha F^{(m)} - D^\alpha F^{(n)} \right)^2 d\underline{x} \quad (13)$$

where $|\underline{\alpha}| = \sum_{i=1}^d \alpha_i$ is the total derivation order while k is the highest Sobolev differential smoothness order required. For $k = 0$, we just get a weighted L^2 non-conformity measure.

3. Ritz-Galerkin Variational Formulation for PUFEM-based Image Registration

In our variational formulation of the image registration problem, the objective function $\mathcal{E}$ to minimize can be written as the sum of three terms:

$$\mathcal{E} = \mathcal{M} + \beta \mathcal{S}_k + \kappa \mathcal{D} \quad (14)$$

in this expression, $\mathcal{M}$ penalizes the mismatch between the reference image $R(\underline{x})$ and the warped template image $T(\underline{x} + \underline{U})$, $\mathcal{S}_k$ penalizes the non-conformity of the PUFEM representation of the warping field $\underline{U}(\underline{x})$. $\mathcal{D}$ is a deformation energy term which is meant to inject prior knowledge about the expected spatial-behaviour of the deformation field. The constants β and κ allow to control the relative influence of the constraints $\mathcal{S}_k$ and $\mathcal{D}$ on the final result. The Ritz-Galerkin method that we adopt consists in expressing the objective function $\mathcal{E}$ in terms of the coefficients $A_{i,r}^{(n)}$ representing the warping field $\underline{U}(\underline{x})$ in the PUFEM framework. We then directly seek the combinations of these parameters that minimize $\mathcal{E}$. For a given node n and a given displacement component, the parameters are represented in the following by the column vector $\mathbf{Y}_i^{(n)} = [A_{i,1}^{(n)}, \dots, A_{i,\rho}^{(n)}]^T$ where ρ is the number of monomials $v_r^{(n)}$ used in the representation. The different components for $i = 1$ to d are piled up to form the column vector $\mathbf{Y}^{(n)} = [\mathbf{Y}_1^{(n)}, \dots, \mathbf{Y}_d^{(n)}]^T$ which has ρd elements.

The Non Conformity Penalty $\mathcal{S}_k$ is the sum of pairwise non-conformity measures $S_k^{(m,n)}(U_i)$, from Eq.(13), over all adjacent node-pairs (m, n) and all components of U_i of the warping field. We recall that the suffix k stands for the Sobolev continuity order. Each of the terms $S_k^{(m,n)}(U_i)$ can be expressed as a non-negative definite quadratic function of $\mathbf{Y}^{(m)}$ and $\mathbf{Y}^{(n)}$ by replacing $U_i^{(m)}$ and $U_i^{(n)}$ by their local expressions from Eq.(3). Writing this in terms of the local coefficient vectors $\mathbf{Y}_i^{(m)}$ and $\mathbf{Y}_i^{(n)}$, we obtain an expression of the form:

$$\begin{aligned} S_k^{(m,n)}(U_i) &= \left[\mathbf{Y}_i^{(m)} \right]^T \mathbf{S}^{(m,n)} \left[\mathbf{Y}_i^{(m)} \right] \\ &+ \left[\mathbf{Y}_i^{(n)} \right]^T \mathbf{S}^{(n,m)} \left[\mathbf{Y}_i^{(n)} \right] \\ &- 2 \left[\mathbf{Y}_i^{(m)} \right]^T \mathbf{C}^{(n,m)} \left[\mathbf{Y}_i^{(n)} \right] \end{aligned} \quad (15)$$

where $\mathbf{S}^{(m,n)}$, $\mathbf{S}^{(n,m)}$ and $\mathbf{C}^{(n,m)}$ are $(\rho d \times \rho d)$ square matrices. It should be noted that this is a non-negative

quadratic expression of the two vectors $\mathbf{Y}_i^{(m)}$ and $\mathbf{Y}_i^{(n)}$ because $S_k^{(m,n)}(U_i)$ is never negative. In particular, both $\mathbf{S}^{(m,n)}$ and $\mathbf{S}^{(n,m)}$ are symmetric non-negative definite matrices.

The Deformation Penalty $\mathcal{D}$ that can be dealt within our Ritz-Galerkin formulation is any non-negative quadratic functional of the warping field derivatives. It can, in general be expressed as a sum of simple quadratic terms:

$$\mathcal{D} = \sum_{\theta} \int_{\Omega} \left(\sum_{i, \underline{\alpha}, |\underline{\alpha}| > 0} \gamma_{i, \underline{\alpha}}^{(\theta)} D^{\underline{\alpha}} U_i \right)^2 d\underline{x} \quad (16)$$

In our implementation, we deal with the case where the coefficients $\gamma_{i, \underline{\alpha}}^{(\theta)}$ are considered constant within each of the window function supports. If inter-node conformity is enforced, we may replace $D^{\underline{\alpha}} U_i$ by a blending of local derivatives $D^{\underline{\alpha}} \tilde{U}_i^{(n)}$. The reasoning leading to inequality (6) can then be repeated and $\mathcal{D}$ is easily shown to have as upper bound a sum of weighted local contributions $\sum_n \mathcal{D}^{(n)}$, each of which can be expressed as:

$$\mathcal{D}^{(n)} = \left[\mathbf{Y}^{(n)} \right]^T \mathbf{D}^{(n)} \left[\mathbf{Y}^{(n)} \right] \quad (17)$$

where $\mathbf{D}^{(n)}$ is a symmetric non-negative definite matrix of size $(\rho d \times \rho d)$.

The Image Mismatch Penalty $\mathcal{M}$ is defined by:

$$\mathcal{M} = \int_{\Omega} (T(\underline{x} + \underline{U}) - R(\underline{x}))^2 d\underline{x} \quad (18)$$

where we select the sum of square grey-level difference mismatch metric.

3.1. Optical Flow Step

In the form of Eq.(18), $\mathcal{M}$ cannot be directly expressed as a sum of local quadratic functions of the coefficient vectors $\mathbf{Y}^{(n)}$. In order to allow us to use a standard quadratic minimization procedure, we need to break down the objective function minimization into a succession of steps of Optical Flow [3], within each of which $\underline{U}$ undergoes sufficiently small changes as to allow linear approximations of the grey level differences $T(\underline{x} + \underline{U}) - R(\underline{x})$. Starting from a previously obtained value or from an initial guess for the warping field $\underline{U}(\underline{x})$, we look for the small incremental warping field $\underline{u}(\underline{x})$ needed to reduce $\mathcal{M} + \beta \mathcal{S}_k + \kappa \mathcal{D}$. At the end of the step, we update $\underline{U}(\underline{x})$ by adding the increment found and repeat up to convergence. We use the optical flow approximation [3] which consists in replacing $(T(\underline{x} + \underline{U} + \underline{u}) - R(\underline{x}))$ by its first order Taylor series ap-

proximation resulting in:

$$\mathcal{M} \cong \int_{\Omega} \left(T(\underline{x} + \underline{U}) - R(\underline{x}) + \sum_i u_i \partial_i T|_{\underline{x} + \underline{U}} \right)^2 d\underline{x} \quad (19)$$

Assuming that the changes u_i in U_i are small enough, the reasoning leading to inequality (6), allows us to express the change in $\mathcal{M}$ as a sum $\sum_n (\Delta \mathcal{M}^{(n)})$ of windowed contributions. By expressing the increments u_i in terms of the difference $\mathbf{Y}^{(n)} - \mathbf{Y}_o^{(n)}$ between the coefficient vector $\mathbf{Y}^{(n)}$ at the current step and that of the previous step $\mathbf{Y}_o^{(n)}$, we have:

$$\begin{aligned} \Delta \mathcal{M}^{(n)} &\cong \left[\mathbf{Y}^{(n)} - \mathbf{Y}_o^{(n)} \right]^T \mathbf{M}^{(n)} \left[\mathbf{Y}^{(n)} - \mathbf{Y}_o^{(n)} \right] \\ &\quad - 2 \left[\mathbf{Y}^{(n)} - \mathbf{Y}_o^{(n)} \right]^T \mathbf{V}^{(n)} \end{aligned} \quad (20)$$

where the entries of the $(\rho d \times \rho d)$ symmetric non-negative definite square matrix $\mathbf{M}^{(n)}$ and the size (ρd) vector $\mathbf{V}^{(n)}$ are given by:

$$\begin{aligned} \mathbf{M}_{(i,r),(j,s)}^{(n)} &= \int \varphi^{(n)} v_r^{(n)} v_s^{(n)} (\partial_i T \partial_j T)|_{\underline{x} + \underline{U}} d\underline{x} \\ \mathbf{V}_{(i,r)}^{(n)} &= \int \varphi^{(n)} v_r^{(n)} (R(\underline{x}) - T(\underline{x} + \underline{U})) \partial_i T|_{\underline{x} + \underline{U}} d\underline{x} \end{aligned} \quad (21)$$

From Eqs.(15, 17), $\beta \mathcal{S}_k + \kappa \mathcal{D}$ is a quadratic function of vectors $\mathbf{Y}^{(n)}$. Within an optical flow step, the objective function $\mathcal{E} = \mathcal{M} + \beta \mathcal{S}_k + \kappa \mathcal{D}$ can therefore be put in the form:

$$\mathcal{E} = C + \mathbf{Y}^T \mathbf{G} \mathbf{Y} - 2 \mathbf{W}^T \mathbf{Y} \quad (22)$$

where C is a constant, $\mathbf{Y}$ is a $(N \rho d)$ column vector obtained by piling up the coefficient vectors $\mathbf{Y}^{(n)}$ of all N elements of the node array. The matrix $\mathbf{G}$ is a non-negative definite symmetric matrix of size $(N \rho d \times N \rho d)$ whereas $\mathbf{W}$ is a column vector of size $(N \rho d)$. Minimizing $\mathcal{E}$ in Eq.(22) is equivalent to solving the linear system:

$$\mathbf{G} \mathbf{Y} = \mathbf{W} \quad (23)$$

which yields the new vector $\mathbf{Y}$ resulting from the optical flow step. In practice, matrix $\mathbf{G}$ is sparse and can be very large. We adopt the conjugate gradient algorithm which is suited to solve such systems. The non-conformity energy term $\beta \mathcal{S}_k$ is found to have a stabilizing influence on the system (23) because it prevents the condition number of matrix $\mathbf{G}$ from being too large. In almost all the practical cases we have studied, less than 5 optical flow steps with 4 conjugate gradient iterations within each step are sufficient for convergence. Slow convergence only occurs when the weight κ of the deformation energy is so large that $\kappa \mathcal{D}$ provides a contribution to the trace of matrix $\mathbf{G}$ more than 100 times larger than that of the other terms. In our experience, one only need such a large κ weight when dealing with the hard Cassini-Oval conformal mapping problem discussed below.

3.2. The Image Registration Algorithm Sequence

A PUFEM hierarchy is defined as explained in section 2. To each level of the hierarchy, a Gaussian smoothing kernel of size σ is associated. Both reference and template images are smoothed with this kernel. We take σ to be a fraction of the average inter-element separation distance h at each level. For each level, the non-conformity penalty $\mathcal{S}_k$ and the deformation penalty $\mathcal{D}$ are mapped to their coefficient representations using the matrices $\mathbf{S}^{(m,n)}$, $\mathbf{S}^{(n,m)}$, $\mathbf{C}^{(n,m)}$ and $\mathbf{D}^{(n)}$ as defined by Eqs.(15, 17). The entries of those matrices turn out to be integral of polynomial expressions over rectangular domains and are precomputed accurately. The registration algorithm itself starts from the coarsest level with associated coarse σ images. Once the objective function has been minimized within a level, we apply the prolongation operator defined in section 2 to get the initial state at the next finer level. The procedure is resumed until convergence at the finest level has been realized. Registration at each level is achieved using a number of optical flow steps as described in section 3.1.

4. Interest of our Ritz-Galerkin Formulation

The variational formulation proposed above consists in representing the objective function $\mathcal{E}$ as a function of the parameters stored in $\mathbf{Y}$ (the column vector defined in section 3.1). We then directly seek the vector $\mathbf{Y}$ that minimizes $\mathcal{E}$. An alternative approach often used in the literature [1, 2, 4], is to derive the Euler-Lagrange Partial Differential Equation (EL-PDE) and then find a solution by discretization. This can be achieved either using a finite difference method or by using the Petrov-Galerkin [8] formulation within a Finite-Element framework. It must be stressed that the EL-PDE is a necessary condition that may not be sufficient to lead to a minimum of the objective function $\mathcal{E}$. This limitation of EL-PDE formulations is clear when dealing with the registration of images with few or no localized features. In this case, one needs to register pairs of smooth object boundaries together. For example, when dealing with a binary image pair representing rounded objects, we only know that their edges should be matched together without knowing beforehand the detailed correspondence between individual edge-points. For illustration, we take two particular cases for the deformation penalty $\mathcal{D}$, namely:

$$\begin{aligned} \mathcal{D}_1 &= \int_{\Omega} \left[\frac{\lambda_1}{2} \left(\sum_i \partial_i U_i \right)^2 + \frac{\mu_1}{4} \sum_{i,j} (\partial_i U_j + \partial_j U_i)^2 \right] d\underline{x} \\ \mathcal{D}_2 &= \int_{\Omega} \left[\frac{\lambda_2}{2} \left(\sum_i \partial_i U_i \right)^2 + \frac{\mu_2}{4} \sum_{i,j} (\partial_i U_j - \partial_j U_i)^2 \right] d\underline{x} \end{aligned} \quad (24)$$

$\mathcal{D}_1$ is the Lamé elastic energy constraint [2] while $\mathcal{D}_2$ is the div-curl regularizer which was first proposed by Suter [10]. Each constraint is defined by two parameters (λ_1, μ_1) for $\mathcal{D}_1$ and (λ_2, μ_2) for $\mathcal{D}_2$. The two constraints yield very similar EL-PDE which can respectively be written as in [7]:

$$\begin{aligned} \mu_1 \Delta \underline{U} + (\lambda_1 + \mu_1) \nabla \nabla \cdot \underline{U} &= \gamma (T(\underline{x} + \underline{U}) - R(\underline{x})) \nabla T \\ \mu_2 \Delta \underline{U} + (\lambda_2 - \mu_2) \nabla \nabla \cdot \underline{U} &= \gamma (T(\underline{x} + \underline{U}) - R(\underline{x})) \nabla T \end{aligned} \quad (25)$$

where ∇T is evaluated at $\underline{x} + \underline{U}$, γ is a constant, Δ is the Laplacian operator and $(\nabla \nabla \cdot)$ is the grad-div operator. It is noted that if $\mu_2 = \mu_1$ and $\lambda_2 = \lambda_1 + 2\mu_1$ the two above EL-PDEs are identical. However, it is clear from Eq.(24) that, provided $\mu_1 = \mu_2 > 0$, $\mathcal{D}_1$ penalizes the symmetrical part $\frac{1}{2} [\partial_i U_j + \partial_j U_i]$ of the warping field Jacobian matrix $[\partial_i U_j]$ whereas $\mathcal{D}_2$ penalizes the anti-symmetric part $\frac{1}{2} [\partial_i U_j - \partial_j U_i]$. One therefore expects radically different spatial behaviours of the warping field obtained using $\mathcal{D}_1$ or $\mathcal{D}_2$. If one relies on the EL-PDE, this difference in behaviours cannot be captured because one would have identical PDEs to solve. To illustrate this point concretely, we have compared our PUFEM-based Ritz-Galerkin results for matching 2D binary image pairs with $\mathcal{D}_1$ and $\mathcal{D}_2$ penalties respectively. In order to benefit from validation from closed-form results, we take the particular cases $\mu_1 = \mu_2 = \mu$, $\lambda_1 = -\mu$, $\lambda_2 = \mu$. The special forms dealt with for $\mathcal{D}_1$ and $\mathcal{D}_2$ read:

$$\begin{aligned} \mathcal{D}_1 &= \int_{\Omega} \frac{\mu}{2} \left[(\partial_1 U_2 + \partial_2 U_1)^2 + (\partial_1 U_1 - \partial_2 U_2)^2 \right] d\underline{x} \\ \mathcal{D}_2 &= \int_{\Omega} \frac{\mu}{2} \left[(\partial_1 U_2 - \partial_2 U_1)^2 + (\partial_1 U_1 + \partial_2 U_2)^2 \right] d\underline{x} \end{aligned} \quad (26)$$

The respective EL-PDEs are identical: the grad-div terms in the left-hand side of Eqs.(25) vanish, leaving only the Laplacian $\Delta \underline{U}$ term. In the numerical experiments, we used for R and T images of size 512×512 . We took a hierarchy of arrays with 7 pyramid-levels. The array size ranged from 6×6 for the coarsest level to 321×321 for the finest. For all levels, the maximum polynomial order was $p = 2$ and the inter-node conformity constraint was $\mathcal{S}_2$ (Sobolev order $k = 2$ constraint).

Fig. 1, shows the results obtained for registering mutually rotated ellipses. It is clear from the above expression (26) that $\mathcal{D}_1$ vanishes for any similarity transformation. The rotation $\mathbf{R}$ is correctly captured when using the $\mathcal{D}_1$ penalty as seen in Fig. 1-(b). More generally, $\mathcal{D}_1$ vanishes for any angle-preserving Conformal Mapping [11] for which the Cauchy-Riemann equations $\partial_1 U_1 = \partial_2 U_2$ and $\partial_1 U_2 = -\partial_2 U_1$ are satisfied. As for the deformation penalty $\mathcal{D}_2$, it is minimal when the warping Jacobian is symmetrical. To deal analytically with this case, we look for a linear expression of the warping field in the form of a symmetric transformation matrix $\mathbf{S}$. Let $\underline{x}$ be the position

vector with the origin coinciding with the ellipse centroid. The equation of the ellipse in the reference image R can be expressed as $\underline{x}^T \mathbf{H} \underline{x} = 1$ where $\mathbf{H}$ is a 2×2 diagonal matrix with positive diagonal. The rotated ellipse (contour in template image) satisfies the equation $\underline{x}^T \mathbf{H}_{\theta} \underline{x} = 1$ where $\mathbf{H}_{\theta} = \mathbf{R}^T \mathbf{H} \mathbf{R}$. The symmetric matrix transformation $\mathbf{S}$ needed to register those two ellipses must satisfy the matrix equation $\mathbf{S} \mathbf{H} \mathbf{S} = \mathbf{H}_{\theta}$. By right-multiplying with matrix $\mathbf{H}$, we get $\mathbf{S} = (\mathbf{H}_{\theta} \mathbf{H})^{1/2} \mathbf{H}^{-1}$. As seen in Fig. 1-(c), using the $\mathcal{D}_2$ deformation penalty yields a pure shear transformation which coincides with the above closed form result. Fig. 2 shows the results obtained for registering a reference disk image with a Cassini oval [12] with a corresponding shape ratio $b/a = 1.1$. The unit disk to Cassini oval mapping has been very frequently used for testing and evaluating Numerical Conformal Mapping Methods [11]. The corresponding numerical problem is challenging. In contrast with the above case with ellipses, the deformation energy is taken in account only within the interior of the contour of the reference image (*i.e.* inside the reference image disk region). This is because, when using the $\mathcal{D}_1$ penalty, the mathematical ground truth for the warping field is just the conformal mapping which has known singularities outside the disk. To restrict the penalty to the disk region, we simply set the corresponding nodal penalties $\mathcal{D}^{(n)}$ to zero whenever the node centre lies outside the disk region. The result for the $\mathcal{D}_1$ penalty case depicted in Fig. 2-(b) behaves as predicted by the exact closed-form expression. The displacement vectors errors relative to ground-truth are found to be less than 2% of the disk radius. The $\mathcal{D}_2$ penalty case leads to a very different warping field (Fig. 2-c).

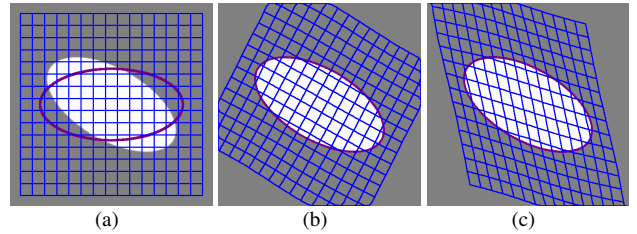


Figure 1. (a) Reference image contour (dark elliptic contour) and rectangular grid with template image (white object). The reference ellipse is rotated by an angle $\theta = \pi/6$ relative to that of the template. The deformed reference contour and reference rectangular grid are displayed together with template image after algorithm convergence using deformation penalty (b) $\mathcal{D}_1$, (c) $\mathcal{D}_2$.

5. Registering 2D Cardiac MR Images

In this experiment, we register pairs of 2D Cardiac MR short-axis images using different regularization constraints. To obtain a ground truth, we generate a synthetic warping field emulating deformations observed in Cardiac tagged

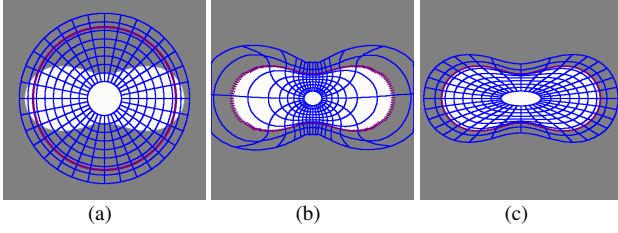


Figure 2. (a) Reference image contour (dark circular contour) together with reference polar grid and template Cassini-oval (white object). The deformed reference contour and reference polar grid are displayed together with template image after algorithm convergence using deformation penalty (b) $\mathcal{D}_1$, (c) $\mathcal{D}_2$.

MRI sequences. The template T is a frame extracted from a cine MR sequence, and the reference R is synthesized as:

$$R(\underline{x}) = T(\underline{x} + \underline{U}_{GT}(\underline{x})) \quad (27)$$

where $\underline{U}_{GT}$ is a synthetic, rotation-contraction displacement field (Fig. 3).



Figure 3. Synthetic warping field $\underline{U}_{GT}$ made up of rotation+contraction

To visualize the deformation fields, we adopt here also the method used in Fig. 2. We display a regular polar grid in the reference image and warp it using the displacement field $\underline{U}$ (Fig. 4-(a) and -(d), *i.e.* a point initially at position $\underline{x}$ is moved to the position $\underline{x} + \underline{U}(\underline{x})$) to generate a deformed grid in the template image.

We registered R and T using a 7-level pyramid. The coarsest level uses a 3×3 array while the finest uses a 129×129 array. For all levels, the maximum polynomial order was $p = 2$ and the inter-node conformity constraint was S_2 (Sobolev order $k = 2$ constraint). With a deformation cost $\mathcal{D}$ penalizing the divergence of the warping field $\mathcal{D} = \int_{\Omega} (\nabla \cdot \underline{U})^2$, the resulting displacement differs significantly from the ground truth (Fig. 4-(b) and -(e)). The fact that introducing a penalty on the warping field divergence (*i.e.* penalizing changes in area) produces departures from ground truth is not surprising because, in such 2D short axis cine sequences, the myocardial wall thickness changes appreciably as well as the apparent ventricular region. We also experimented with other deformation constraints and obtained similar results. In contrast, when taking no deformation constraints at all (*i.e.* $\mathcal{D} = 0$ and using only the intrinsic Sobolev conformity constraint S_2), the warping field

resulting from our algorithm is found to be very close to ground-truth (Fig. 4-(c) and -(f)).

6. Summary and Conclusions

We have proposed a mesh-free Hierarchical Partition of Unity Finite Element Method for solving variational problems related to non-rigid multidimensional image registration. Our objective function includes an image mismatch penalty $\mathcal{M}$ and a deformation penalty $\mathcal{D}$ that are both expressed as a sum of local terms involving one node at a time. In contrast to B-splines which have a built-in smoothness, our warping field smoothness is controlled by including a special term S_k to the objective function to minimize. This term is non-local, it is defined as the sum of inter-node order k Sobolev discontinuity measures $S_k^{(m,n)}$. This enforces inter-node smoothness for the warping field and all its partial derivatives up to a predefined order k . The proposed approach brings great flexibility since one can control the degree of smoothness of the warping field by selecting suitable values for k , for the degree p of the polynomials used and for the inter-element separations h_i used for the finest level of the PUFEM hierarchy.

Prior knowledge on the spatial behaviour of the warping field can be injected through the deformation penalty $\mathcal{D}$ in addition to the S_k constraint. Experiments were made for various deformation penalties $\mathcal{D}$ in the form of quadratic function of first order warping field derivatives. Our scheme proved effective and results agreed closely with ground-truth (when available).

Another mode of operation of our scheme is to skip totally the deformation energy $\mathcal{D}$. In this case, the regularity of the warping field is controlled exclusively by the Sobolev non-conformity penalty S_k . This mode is more attractive in cases where we wish to extract the warping field from the images without introducing any artificial bias. The regularity constraint S_k is needed only to reduce the influence of imaging artefacts or noise. Our experiments with real and synthetic Cardiac MR temporal sequences are promising in this respect.

Acknowledgement

The authors gratefully acknowledge St Catharina Hospital Eindhoven for providing the Cardiac MR images. Many thanks to Pierre Villon for fruitful discussions and to Benoit Mory, Cecil Dufour, Jean-Michel Rouet and ICCV reviewers for helpful remarks on the manuscript.

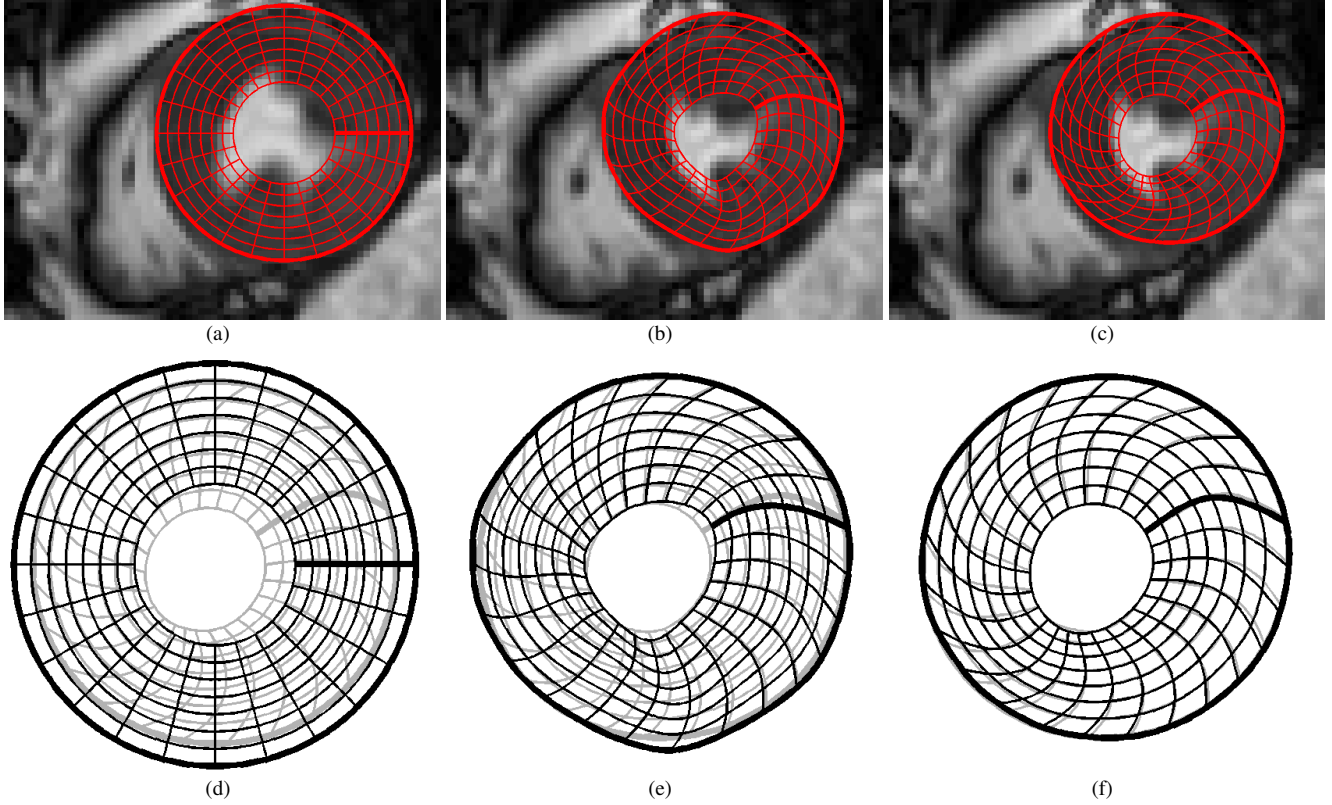


Figure 4. (a) Reference image and initial polar grid covering the myocardium. (b) Template image and deformed grid, with the constraint penalty $\mathcal{D} = \int_{\Omega} (\nabla \cdot \underline{U})^2$. (c) Same as (b) but with intrinsic internode Sobolev S_2 constraint only. (d) Initial regular grid (black) and ground truth deformed grid (gray). (e) Computed deformed grid (black) and ground truth (gray), with the elastic constraint penalty $\mathcal{D} = \int_{\Omega} (\nabla \cdot \underline{U})^2$. (f) Same as (e) with Sobolev S_2 constraint only, evaluated warping field (black-grid) is close to ground-truth (gray).

References

- [1] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, New-Yor, 2004.
- [2] E. Haber and J. Modersitzki, "Multilevel Method for Image Registration", *SIAM Journal of Scientific Computing*, vol. 27, no. 5, pp. 1594-1607, 2006.
- [3] S. S. Beauchemin and J. L. Barron, "The Computation of Optical Flow", *CM Computing Surveys*, vol. 27, no. 3, pp. 433-467, 1995.
- [4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping", in *Proc. 8th European Conference on Computer Vision*, Springer LNCS 3024, vol. 4, pp. 25-36, T. Pajdla and J. Matas Editors, vol. 4, pp. 25-36, Prague, 2004.
- [5] M. Ferrant, S. K. Warfield., C. R. G. Guttmann, R. V. Mulkern, F. A. Jolesz and R. Kikinis, "3D Image Matching Using a Finite Element Based Elastic Deformation Model", in *Proceedings of the MICCAI Conference*, C. Taylor and A. C. F. Colchester Editors, Cambridge UK, pp. 202-209, 1999.
- [6] C. Ó S. Sorzano, P. Thévenaz and M. Unser, "Elastic Registration of Biological Images Using Vector-Spline Regularization", *IEEE Transaction on Biomedical Engineering*, vol. 52, no. 4, pp. 652-663, 2005.
- [7] T. Corpetti, D. Heitz, G. Arroyo, E. Mémin and A. Santa-Cruz, "Fluid Experimental Flow Estimation Based on Optical-Flow Scheme", *Experiments in Fluids*, vol. 40, pp. 80-97, 2006.
- [8] I. Babuška and J. M. Melenk, "The Partition of Unity Method", *International Journal of Numerical Methods in Engineering*, vol. 40, no. 4, pp. 727-758, 1997.
- [9] J. M. Melenk and I. Babuška, "The Partition of Unity Finite Element Method: Basic Theory and Applications", *Computer Methods in Applied Mechanical Engineering*, vol. 139, no. 1-4, pp. 289-314, 1996.
- [10] D. Suter, "Motion Estimation and Vector Splines", in *Proceedings of the CVPR Conference*, Seattle USA, pp. 939-942, 1994.
- [11] T. K. Delillo, "The accuracy of Numerical Conformal Mapping Methods: A Survey of Examples and Results", *SIAM Journal on Numerical Analysis*, vol. 31, no. 3, pp. 788-812, 1994.
- [12] E. W. Weisstein, "Cassini Ovals", *A Wolfram Web Resource*, <http://mathworld.wolfram.com/CassiniOvals.html>.