Sparse Bayesian Registration

Loic Le Folgoc, Hervé Delingette, Antonio Criminisi, Nicholas Ayache

HAL Id: hal-01006605

https://inria.hal.science/hal-01006605

Sparse Bayesian Registration

Loïc Le Folgoc[1], Hervé Delingette[1], Antonio Criminisi[2], and Nicholas Ayache[1]

[1] Asclepios Research Project, INRIA Sophia Antipolis, France
[2] Machine Learning and Perception Group, Microsoft Research Cambridge, UK

Abstract. We propose a Sparse Bayesian framework for non-rigid registration. Our principled approach is flexible, in that it efficiently finds an optimal, sparse model to represent deformations among any preset, widely overcomplete range of basis functions. It addresses open challenges in state-of-the-art registration, such as the automatic joint estimate of model parameters (*e.g.* noise and regularization levels). We demonstrate the feasibility and performance of our approach on cine MR, tagged MR and 3D US cardiac images, and show state-of-the-art results on benchmark datasets evaluating accuracy of motion and strain.

1 Introduction

Non-rigid image registration is the ill-posed task of inferring a deformation u from a pair of observed (albeit typically noisy), related images I and J. Classical approaches propose to minimize a functional which weighs an image similarity criterion $\mathcal{D}$ against a regularizing (penalty) term $\mathcal{R}$:

$$\arg\min_{\mathrm{u}} \mathcal{E}(\mathrm{u}) = \mathcal{D}(I, J, \mathrm{u}) + \lambda \cdot \mathcal{R}(\mathrm{u}) \tag{1}$$

Prior knowledge to precisely model the space of plausible deformations or the regularizing energy is generally unavailable. Adequate choices thus result from a compromise between computational efficiency, numerical stability of the optimization schemes and extent of the spanned space. Furthermore, the optimal regularizing trade-off λ is unknown. Cross-validation presupposes availability of either a set of ground-truth transforms or a criterion to measure the 'goodness' of a transform. Even then, high variability in images of the same modality can render the use of any single value of λ unsuitable. In the end, user expertise acquired via lengthy trial-and-error processes often remains necessary to improve the quality of registration.

Recent advances pave the way to go beyond this state of things. Bayesian formulations of registration allow to infer from the data ([1]) or to integrate over ([2,3]) such parameters as the regularization penalty λ. Such probabilistic frameworks additionally let us derive uncertainty estimates over the solution, which may prove instrumental in improving the accuracy and robustness of downstream medical imaging applications. These methods remain computationally intensive, and typically achieve tractability *via* a sound use of theoretical or numerical approximations; they also introduce a number of additional user-defined hyperparameters. On a different note, sparsity-inducing priors provide a venue for the use of rich, over-complete representations of deformations,

at little added computational cost, while preserving the stability of solvers. L_1 regularizers were recently proposed in the context of registration [4]; as a drawback, they so far render both the joint estimation of model parameters and the uncertainty estimation impractical.

In this paper, we further build on the tools from the Sparse Bayes and Automatic Relevance Determination fields so that we retain the computational and numerical advantages of sparse formulations, while jointly estimating the most adequate deformation model and its parameters thanks to a probabilistic setting. As a result, the need for manual interaction and expertise is reduced. Our contributions are as follows:

1. We extend a state-of-the-art tool for sparse regression [5,6] in scope, so that it may now handle multivalued outputs and generic quadratic priors.
2. We develop a registration framework that makes use of sparse regression to jointly select the most appropriate deformation model and estimate its parameters.
3. We demonstrate our approach by tracking the cardiac motion on cine MR, tagged MR and $3D$-US images.

2 Registration Setting

At a high-level, our approach casts the registration process into two easier sub-tasks: the search for (voxel) pairwise correspondences, followed by the estimate of a smooth displacement *via* the tools of statistical regression. This can be interpreted as minimizing the quadratic approximation of (1) w.r.t. the displacement field u, near a local minimum of the similarity term.

This two-step process is in turn embedded in a multi-resolution pyramidal scheme: each resolution level runs on a smoothed, subsampled version of the images at the previous level.

2.1 Data-Matching Strategy

The similarity energy $\mathcal{D}(I, J, \mathrm{u})$ relates a fixed image I and a moving image J warped according to some candidate displacement field u. Defining an image as a set of intensity measurements regularly sampled at points x_i, $1 \leq i \leq N$, and taking the quadratic approximation of $\mathcal{D}(I, J, \mathrm{u})$ around one of its local minima yields:

$$\arg\min_{\mathrm{u}} \sum_{1 \leq i \leq N} (\bar{\mathrm{u}}_i - \mathrm{u}(\mathrm{x}_i))^{\intercal} \mathbf{H}_i (\bar{\mathrm{u}}_i - \mathrm{u}(\mathrm{x}_i)) + \lambda \cdot \mathcal{R}(\mathrm{u}) \qquad (2)$$
$$\text{s.t.} \quad \partial_{\mathrm{u}} \mathcal{D}(I, J, \mathrm{u})\big|_{\bar{\mathrm{u}}} = 0.$$

$\bar{\mathrm{u}}$ is a local minimizer of $\mathcal{D}(I, J, \mathrm{u})$ and $\bar{\mathrm{u}}_i \triangleq \bar{\mathrm{u}}(\mathrm{x}_i)$. $\mathbf{H}_i$ is the Hessian of $\mathcal{D}(I, J, \mathrm{u}_i)$ evaluated at $\bar{\mathrm{u}}$. In this paper, we use the sum of squared differences (SSD) as a similarity energy and thus approximate $\mathbf{H}_i$ with the structure tensor at $\mathrm{y}_i = \mathrm{x}_i + \bar{\mathrm{u}}_i$, that is to say $\mathbf{H}_i = \sum_{p \in \mathcal{W}} \nabla I(\mathrm{y}_i + p) \nabla I(\mathrm{y}_i + p)^{\intercal}$. To compute the local estimate of displacement $\bar{\mathrm{u}}$, we adopt a block matching strategy where the optimization is done using the L-BFGS-B algorithm independently for each voxel, which is highly parallelizable. The structure

matrix $\mathbf{H}_i$ captures local structures such as boundaries and edges. $\boldsymbol{n}^{\mathsf{T}}\mathbf{H}_i\boldsymbol{n}$ relates to our confidence in $\bar{u}_i$ along the direction $\boldsymbol{n}$. These 'weights' $\mathbf{H}_i$ naturally account for the inhomogeneous spread of informative features in the image.

2.2 Representation of Displacements

For this work, we constrain the displacement field $u(x)$ to be expressed over a dictionary of radial basis functions $\{\phi_k, 1 \leq k \leq M\}$, specifically Gaussian kernels $\phi_k(x) = K_{S_k}(x_k, x)\,\mathbf{I}$, of varying variance S_k and centered at a given voxel x_k:

$$u(x) = \sum_{1 \leq k \leq M} \phi_k(x)\boldsymbol{w}_k \tag{3}$$

The kernel width S_k spans a user-predefined set of values, which allows for a redundant, multiscale representation. Therefore we benefit both from a compact representation *via* larger kernels, and from the ability to capture finer local details *via* smaller kernels.

2.3 Regularization Framework

While any quadratic energy $\mathcal{R}(u)$ suits our framework, we specifically consider those that exploit the R.K.H.S. structure of the space $\mathcal{H}$ spanned by Gaussian kernels of width $S \leq \min_k\{S_k\}$; namely $\mathcal{R}(u) = \|Du\|_{\mathcal{H}}^2$, with Du a(ny) differential operator acting on u. Indeed, mathematical bridges between the theory of translation invariant kernels and that of Fourier analysis allow us to derive *analytic* expressions for quantities of the form

$$(D\phi_k | D\phi_l)_{\mathcal{H}} = \int_{\xi} \widehat{D\phi_k}(\xi)^* \widehat{D\phi_l}(\xi)\, \widehat{K}_S^{-1}(\xi)d\xi\,. \tag{4}$$

The properties of Gaussian kernels under Fourier transform, multiplication (by Gaussian kernels) and summation reduce (4) to an analytic expression evaluated *via* basic calculus operations and exponentiation. In other words, our framework comes with a computationally efficient, analytic implementation of *e.g.* a thin-plate energy ($D = \nabla$), a bending energy ($D = \nabla^2$) or a 'compressibility' penalty ($D = \mathrm{div}$).

3 Sparse Bayes Regression

Solving (2) w.r.t. u constitutes a canonical regression problem: from a finite set of noisy observations $\bar{u}_i$ at points x_i (with confidence $\beta\mathbf{H}_i$), and further assuming a prior over u proportional to $\exp{-\lambda\mathcal{R}(u)}$, find an optimal reconstruction of the complete signal u. To introduce sparsity in such a formulation, we resort to a mechanism first suggested in the seminal work [5,6]. We hereby present the few core ideas of our algorithm, whereas additional details are left to the electronic appendix. Consider the effect of adding a penalty term $\boldsymbol{w}^{\mathsf{T}}\mathbf{A}\boldsymbol{w} \triangleq \sum_k \boldsymbol{w}_k^{\mathsf{T}}\mathbf{A}_k\boldsymbol{w}_k$ on the magnitude of the $\boldsymbol{w}_k$'s in Eq. (2). The problem remains quadratic with additional diagonal weights:

$$\arg\min_{\boldsymbol{w}} (\bar{\boldsymbol{u}} - \Phi\boldsymbol{w})^{\mathsf{T}} \beta\mathbf{H} (\bar{\boldsymbol{u}} - \Phi\boldsymbol{w}) + \boldsymbol{w}^{\mathsf{T}} (\lambda\mathbf{R} + \mathbf{A})\, \boldsymbol{w} \tag{5}$$

where we resorted to block notations: w and $\bar{u}$ are respectively the concatenation of the w_k and $\bar{u}_i$, $\Phi_{ik} \triangleq \phi_k(\mathrm{x}_i)$, $\mathbf{H} \triangleq \mathrm{diag}(\mathbf{H}_i)$ and $\mathbf{A} \triangleq \mathrm{diag}(\mathbf{A}_k)$. The minimizer of (5) is given by $\boldsymbol{\mu} = \Sigma\Phi^\intercal\beta\mathbf{H}\bar{u}$, with $\Sigma = (\Phi^\intercal\beta\mathbf{H}\Phi + \lambda\mathbf{R} + \mathbf{A})^{-1}$. In fact $\boldsymbol{\mu}$ coincides with the expectation of the posterior distribution of w conditionally to the model $\mathbf{A}, \beta, \lambda, \mathbf{R}$, and Σ with its covariance.

Sparsity will ultimately stem from the fact that if a penalty $\mathbf{A}_k$ becomes 'infinite', the corresponding coefficients in w and Σ will in turn be null, effectively pruning the corresponding basis from the solution. Automatic selection of the optimal basis penalties $\mathbf{A}_k$ and of the model parameters β, λ can be achieved either by maximizing the evidence $l(\mathbf{A}) \triangleq \log p(\bar{u}|\mathbf{A}, \beta, \lambda)$; or *via* an Expectation-Maximization (EM) scheme, by maximizing $\mathbb{E}_{w\sim\mathcal{N}(\boldsymbol{\mu},\Sigma)}[\log p(\bar{u}, w|\mathbf{A}, \beta, \lambda)]$. Tipping proposed an efficient *evidence*-based procedure to estimate the $\mathbf{A}_k$'s for his Relevance Vector Machine [6], in the case of *scalar* regression with $\lambda = 0$, which we extend to the fully general setting.

3.1 Automatic Determination of $\mathbf{A}_k$

We maximize the evidence $l(\mathbf{A}) = \int_w p(\bar{u}|w, \beta)p(w|\mathbf{A}, \lambda)dw$. The distribution of $\bar{u}$ conditionally to the model is a zero-mean Gaussian with covariance $\mathbf{C}$ given by (6). The evidence thus reduces to the sum of two antagonistic terms, $-1/2\{\log|\mathbf{C}| + \bar{u}^\intercal\mathbf{C}^{-1}\bar{u}\}$, that respectively induce complexity control (sparsity) and data fidelity.

$$\mathbf{C} = (\beta\mathbf{H})^{-1} + \Phi(\mathbf{A} + \lambda\mathbf{R})^{-1}\Phi^\intercal \tag{6}$$

The process of setting the $\mathbf{A}_k$'s can be seen as fitting a covariance model $\mathbf{C}$ to the data $\bar{u}$: since part of the data is explained 'for free' by the contribution $(\beta\mathbf{H})^{-1}$ of the noise to $\mathbf{C}$, only a few degrees of freedom need be active and sparsity is achieved.

3.2 A Fast Greedy Iterative Scheme

The procedure relies on efficient updates *via* rank-one matrix identities to iteratively build a solution starting with no base (all $\mathbf{A}_k$'s set to ∞): at each iteration, we take a single action among the addition of a previously inactive basis, or the update or deletion of an active one. The contribution of a given basis to the evidence can be singled out in the form of Eq. (7). We leave the maximization of $l(\mathbf{A}_k)$ to the technical appendix.

$$l(\mathbf{A}_k) = \log|\mathbf{A}_k + \boldsymbol{\kappa}_k| - \log|\mathbf{A}_k + \boldsymbol{\kappa}_k + \boldsymbol{s}_k| + \boldsymbol{q}_k^\intercal\{\mathbf{A}_k + \boldsymbol{\kappa}_k + \boldsymbol{s}_k\}^{-1}\boldsymbol{q}_k \tag{7}$$

The action that leads to the largest gain in evidence is implemented, and the current estimate of the solution $w^* = \boldsymbol{\mu}$ is in turn updated. The evidence therefore increases monotonically: convergence towards a local minimum is guaranteed, and monitored *via* the gain in evidence. The asymptotic complexity of the algorithm is of the same order as the specific case of the regular RVM [6] ($\lambda = 0$).

3.3 Automatic Determination of the Model Parameters

We jointly estimate the model parameters *via* an EM algorithm. For the regularizing trade-off λ, this leads to the maximization problem (8), which involves a single SVD of

$\mathbf{A}^{-1/2}\mathbf{R}\mathbf{A}^{-1/2}$ followed by a Newton optimization in log-scale.

$$\lambda^* = \arg\max_\lambda -\frac{\lambda}{2}\text{tr}(\Sigma\mathbf{R}) + \frac{1}{2}\log|\mathbf{A} + \lambda\mathbf{R}| - \frac{\lambda}{2}\boldsymbol{\mu}^{\mathsf{T}}\mathbf{R}\boldsymbol{\mu} \qquad (8)$$

For the estimation of the noise level β^{-1} in the context of registration, we follow Simpson *et al* [1] in introducing a virtual decimation factor α that accounts for the local covariance in image measurements. We thus estimate both the virtual decimation and the noise level (9) from the residual image, once per resolution level in the pyramid.

$$N \cdot \beta^{*-1} = \|I - J \circ (\text{Id} + \mathrm{u})^{-1}\|^2 + \text{tr}(\Sigma\Phi^{\mathsf{T}}\mathbf{H}\Phi) \qquad (9)$$

4 Experimental Results

For all of our experiments, the multiscale representation consists of three levels of progressively finer, isotropic Gaussian kernels, of respective variance $S_1 = 18^2$ mm^2, $S_2 = 9^2$ mm^2 and $S_3 = 4^2$ mm^2. The pyramidal scheme starts with only the coarsest bases at the coarsest pyramid level; a finer basis type is added for each subsequent pyramid level. Therefore at the finest level in the pyramid, all scales S_k are available and jointly optimized upon. Gaussian basis centers coincide with the voxel centers. Furthermore, we use the compressibility penalty as a regularizer (section 2.3) for the 3DTAG dataset, and a bending energy for the other modalities. Note also that our registration scheme does not make use of pre-segmentations of regions of interest, that can be challenging or otherwise impractical to obtain.

4.1 Synthetic 3D Ultrasound Cardiac Dataset

We first demonstrate our approach on synthetic 3D US data from the STACOM 2012 registration challenge [7]. The datasets consist of 4 B-mode image sequences generated from the same mechanical simulation at varying signal-to-noise ratios (SNRs). A dense ground truth, in terms of motion and strain, is available from the simulated meshes.

Fig. 1b evidences a slight tendency to underestimate the radial strain that is consistent with the level of accuracy reported in Fig. 1a. Of note for this dataset is that part

(a) Accuracy Benchmark   (b) Strain (AHA segments)   (c) Noise level vs. SNR
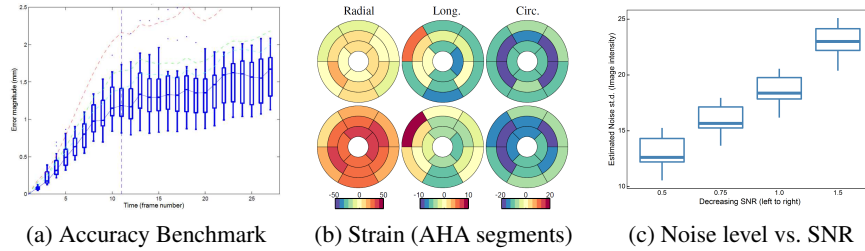
Fig. 1: (Left) Accuracy over time for decreasing SNRs, from blue to cyan, green and red. (Middle) Strain at End-Systole averaged over AHA segments: estimated (top) and ground truth (bottom). (Right) Boxplots of the estimated noise level per SNR dataset.

of the mesh of interest falls outside of the image domain (around 1.5cm beyond); yet our regularization strategy ensures that the visible portion of the motion drives the part of the mesh left-out: this typically yields a maximum error of 6mm and a median error of 2.5mm for this unseen region. Fig. 1c reveals a consistent increase in the (automatically) estimated noise level as the signal-to-noise ratio worsens. Lastly, we believe that the impact of temporal and spatial speckle patterns with regard to the Block Matching procedure would be worth investigating thoroughly, so as to improve the quality of voxel-wise correspondences in 3D US imaging prior to the regularization.

4.2 Tagged Magnetic-Resonance Imaging Benchmark

Next we validate our approach on real data from the STACOM 2011 registration challenge [8], consisting of 3D+T tagged MR sequences. Ground-truth is available in the form of 2 sets of 12 landmarks manually followed over time (by 2 observers), equally distributed on the 4 walls at 3 ventricular levels. The median accuracy at End-Systole (ES) was benchmarked for state-of-art contestants. Fig.2a summarizes challengers' results along with ours; our approach achieves best on this benchmark.

The Block Matching strategy appears to adequately capture the texture of the tags and provides good voxel-wise correpondences. As evidenced in Fig.2b, the regression step in turn estimates an appropriate regularity level, which results in smooth deformations and strain maps: longitudinal, circonferential and radial strain values are consistent with the literature and reveal a thickening of the myocardium along the radial orientation, and contraction in other directions. Note that strain values are simply computed from the analytical expression of the transform (Eq. (3)) without resorting to finite differences approximations.

4.3 Cardiac Cine Magnetic-Resonance Imaging Dataset

The STACOM 2011 challenge [8] additionally provides cine MR (3D+T) cardiac volunteer data, along with End-Diastole (ED) segmentations of the Left Ventricle. For every volunteer, the motion is tracked over the cardiac cycle and resulting mesh deformations and volume curves are reported in Fig. 3. Cine MR cardiac sequences typically include basal slices, with visible outflow tracts and apparent topology changes which make the estimation of a regular motion difficult. To cope with this challenge, we allow the noise
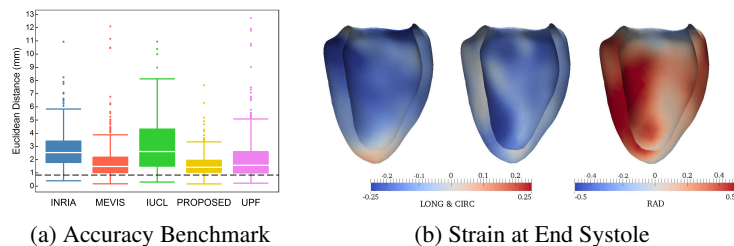

(a) Accuracy Benchmark   (b) Strain at End Systole

Fig. 2: (Left) Box-plots of tracking errors. Doted black line represents average inter-observer variability. (Right) Strain computed from the 3DTAG data of volunteer V9.
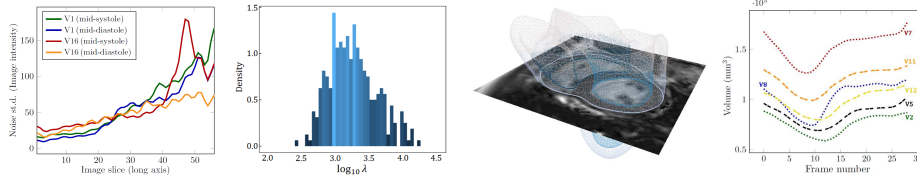
Fig. 3: (From Left to Right) Spatial and inter-image variability of the noise level estimate. Variability in the estimated regularization, illustrated as a density plot over all the datasets. Exemple 2D slice at ES, obtained by propagating the reference segmentation from ED *via* the registration output. Example volume curves for volunteers.

level β^{-1} to vary spatially along the long-axis, and derive independent estimates per slice according to Eq. (9). Fig 3a shows that our algorithm properly captures this specificity of basal slices via an increased noise standard deviation, which in turns allows to preserve the smoothness of the estimated displacement without affecting the overall accuracy of the tracking. Moreover, Fig 3a and Fig. 3b attest to the high variability of the optimal model parameters β^{-1} and λ, both within an image sequence or between different volunteer datasets, which otherwise render their manual estimation via a trial-and-error approach cumbersome.

4.4 Uncertainty estimates

Our regression framework provides us not only with a pointwise estimate of the displacement field, parameterized by the weights $\boldsymbol{w}^* = \boldsymbol{\mu}$, but also a covariance matrix Σ interpretable in terms of uncertainty over $\boldsymbol{w}$. What is more, such information can be translated into a spatially meaningful estimate of variance in the registration.

Fig 4 (Right) displays such variance estimates in the form of a tensor map, where each single tensor fully captures directional information about the uncertainty at a given point in space. The color scheme encodes the orientation of the second eigenvector of the tensor. Indeed, the first principal direction of uncertainty is, to an overwhelming majority over the image, oriented approximately along the long-axis. This observation is in agreement with the drastically lower resolution of cine MR sequences in this direction. The direction of lowest uncertainty roughly coincides with the gradient of intensity, as edges and boundaries prove to be informative features to drive the registration. Lastly the uncertainty is higher in homogeneous regions, such as blood cavities.

Alternatively, for complete 3D+T sequences, displacement fields can be efficiently sampled at each time step t from the estimated posterior Gaussian distribution over $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \Sigma_t)$; such 3D+T samples can in turn be used to approximate the variance on integral quantities of relevance. Fig 4 (Left) illustrates such processing on the LV volume over the cycle. Note that while segmentation errors and mesh misalignment obviously accounts for a significant part of the uncertainty, our confidence bars only reflect uncertainty in the registration itself. Finally, all of our uncertainty estimates preserve the spatial covariance in the signal, unlike previous work [1] where concerns for computational tractability forced the approximation of Σ as a diagonal matrix.
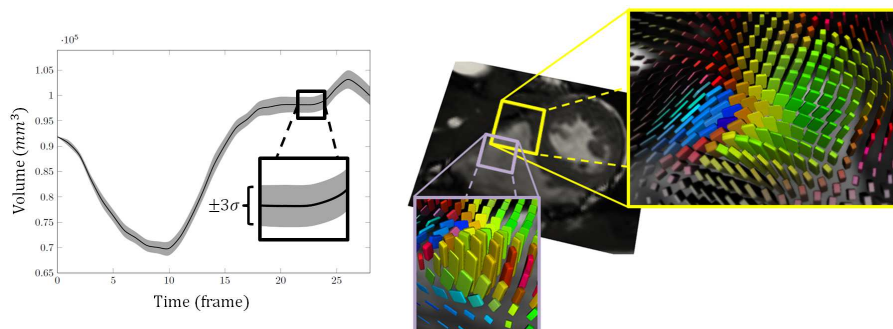
Fig. 4: (Left) LV volume with its associated confidence interval, estimated on the cine-MR data of volunteer V9. (Right) Spatial uncertainty visualized as a tensor map, with higher uncertainty in homogeneous regions and along contour lines.

5   Conclusion

We proposed a framework for registration that allows for the automatic determination of the relevant deformation model and of its hyperparameters, and demonstrated its performance on cardiac images of various modalities. Our work relies on statistical regression tools from the Sparse Bayes field, and extends them to handle multivalued output and any (quadratic) regularizing energy. As a generic machine learning tool, this work opens up many interesting perspectives for regression and classification problems. As a registration framework, its flexibility alleviates the need for manual supervision and makes the motion tracking process easier. As an added benefit, we have access to quantities that are directly interpretable in terms of uncertainty in the output.

References

1. Simpson, I.J., Schnabel, J.A., Groves, A.R., Andersson, J.L., Woolrich, M.W.: Probabilistic inference of regularisation in non-rigid registration. NeuroImage **59**(3) (2012) 2438–2451
2. Janoos, F., Risholm, P., Wells, W.: Bayesian characterization of uncertainty in multi-modal image registration. Biomedical Image Registration (2012) 50–59
3. Richard, F.J., Samson, A.M., Cuénod, C.A.: A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences. Stat Comput **19**(4) (2009)
4. Shi, W., Jantsch, M., Aljabar, P., Pizarro, L., Bai, W., Wang, H., ORegan, D., Zhuang, X., Rueckert, D.: Temporal sparse free-form deformations. MedIA **17**(7) (2013) 779–789
5. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. JMLR **1** (2001)
6. Tipping, M.E., Faul, A.C., et al.: Fast marginal likelihood maximisation for sparse bayesian models. In: Workshop on artificial intelligence and statistics. Volume 1., Jan (2003)
7. De Craene, M., Marchesseau, S., Heyde, B., Gao, H., Alessandrini, M., Bernard, O., Piella, G., Porras, A., Saloux, E., Tautz, L., et al.: 3d strain assessment in ultrasound (STRAUS): A synthetic comparison of five tracking methodologies. TMI (2013)
8. Tobon-Gomez, C., De Craene, M., McLeod, K., Tautz, L., Shi, W., Hennemuth, A., Prakosa, A., Wang, H., Carr-White, G., Kapetanakis, S., et al.: Benchmarking framework for myocardial tracking and deformation algorithms: An open access database. MedIA (2013)