

Iconic Feature Registration with Sparse Wavelet Coefficients

Pascal Cathier

CEA, DSV, DRM, SHFJ, Orsay, F-91400 France

Abstract. With the growing acceptance of nonrigid registration as a useful tool to perform clinical research, and in particular group studies, the storage space needed to hold the resulting transforms is deemed to become a concern for vector field based approaches, on top of the traditional computation time issue. In a recent study we lead, which involved the registration of more than 22,000 pairs of T_1 MR volumes, this constrain appeared critical indeed. In this paper, we propose to decompose the vector field on a wavelet basis, and let the registration algorithm minimize the number of non-zero coefficients by introducing an L_1 penalty. This enables a sparse representation of the vector field which, unlike parametric representations, does not confine the estimated transform into a small parametric space with a fixed uniform smoothness : nonzero wavelet coefficients are optimally distributed depending on the data. Furthermore, we show that the iconic feature registration framework allows to embed the non-differentiable L_1 penalty into a C^1 energy that can be efficiently minimized by standard optimization techniques.

1 Introduction

While nonrigid registration has been around for decades, its use has long been restricted to field experts. In recent years, however, we observed a multiplication of its applications and a democratization of its use, probably lead by the growing popularity of voxel-based morphometry (VBM) studies [1] and computerized atlases [2,3,4]. We believe that with the growing confidence in nonrigid registration fueled by results obtained in such group studies, nonrigid registration is to become a popular tool.

Its massive use yet raises new problems. One of them, on which we focus in this article, is the problem of storage requirements. In a recent study we lead [5], which involved the registration of more that 22,000 pairwise registration of T_1 MRI data, the disk space required to hold all the resulting vector fields indeed appeared critical.

To reduce storage requirements, one can use parametric transform models, the most popular being the tensor product of B-splines [6,7]. By placing control points at every 4 voxels, the number of vector parameters drops to $1/4^3 = 1.6\%$ of the total number of voxels. However, this data independant decimation has a cost : transform smoothness is increased uniformly. By contrast, elastic materials typically have sharp deformations with rapidly varying Jacobian values at their boundaries, and smooth variations inside uniform areas.

Another solution is to compress vector fields posterior to registration. The compression has to be lossy — lossless compression of floating point sequences is notably inefficient. Compared to the parametric approach, this solution better distributes spatially the information that needs to be kept. However, the procedure is still done without knowledge of the data. One does not know what is lost during compression, and thus compression is necessarily done at conservative levels. Furthermore, vector fields resulting from registration may badly depend on the basis used for compression, depending on the regularization.

In this paper, we propose an algorithm that uses a wavelet decomposition of the vector field to register two images. The number of non-zero wavelet coefficients is minimized by including an L_1 penalty in the global energy. It is a well-known trick used e.g. for signal filtering or feature selection [8], and which has been used by [9] in the close field of optical flow estimation. Furthermore, we show that this L_1 penalty can be embedded into a smooth energy that can be minimized by standard optimization procedures. We finally present some preliminary results showing that registration can be achieved using only a very small number of non-zero wavelet coefficients.

2 Presentation of the Registration Algorithm

2.1 Wavelet-Based Iconic Feature Registration

In [10], we introduced the concept of iconic feature registration. The main characteristic of iconic feature registration (IFR) is its similarity with geometric landmark registration in its use of corresponding points to guide the alignment. As opposed to other intensity-based techniques, IFR algorithms compute intensity similarity between corresponding points, and estimate the transform from these pairings by optimizing a geometric criterion. This formulation of registration is potentially more powerful because it better takes into account localization errors. See Fig. 1 for a schematic illustration of the differences between the IFR and the standard formulations.

Although usually not recognized as such, IFR is a widespread technique. Block matching [2] falls into that category, as well as many other two-step techniques [11,12]. These are yet presented as *ad hoc* techniques lacking a solid variational framework. In [10], we proposed an energy minimization framework for IFR. The algorithm presented here relies on a generalization of this energy.

Let $I : \Omega_I \in \mathbf{R}^3 \rightarrow \mathbf{R}$ and $J : \Omega_J \in \mathbf{R}^3 \rightarrow \mathbf{R}$ be two images to be registered. The energy we minimize depends on two maps T and M going from Ω_I to Ω_J that represent resp. the transform to be estimated, and a set of point matches:

$$E(M, T) = S(I, J \circ M) + d(M, T) + R(T), \tag{1}$$

where $S(I, J \circ M)$ is an intensity (dis)similarity energy between images I and J related by matches M , $R(T)$ is a regularization energy accounting for priors on the transform, and d is a geometric distance between the matches M and the transform T , accounting for geometrical noise originating from errors in our



Fig. 1. Schematic illustration of the difference between intensity feature based and standard intensity-based approaches. **First two images:** original images showing a large and a small circle. **Third image:** minimizing intensity differences results in a maximum overlap of the circles. **Last image:** an iconic-landmark based approach minimizes distances between iconic pairings, putting the smaller circle at a constant distance to the larger one, thus better distributing the alignment errors.

model of the transform, or to genuine differences between objects that we do not wish to recover.

A Statistical Inference Viewpoint. Energy (1) can be related to the conditional log-probability $\log p(M, T|I, J)$, since

$$\log p(T, M|I, J) = \log p(I, J|M, T) + \log p(M|T) + \log p(T) - \log p(I, J). \quad (2)$$

Assuming a uniform prior on images I and J , we identify $S(I, J \circ M)$ with $-\log p(I, J|M, T) = -\log p(I, J|M)$, $R(T)$ with $-\log p(T)$ and $d(M, T)$ with $-\log p(M|T)$. The minimization of (1) can thus be interpreted as a maximum a posteriori (MAP) estimation of the unknown transform T and matches M .

An alternative would be to maximize $\log p(T|I, J)$ directly and treat M as a hidden variable. This yields a robust registration algorithm taking into account the uncertainty over M for the estimation of T . However, the set of matches M makes a bad hidden variable, because its conditional distribution cannot be deduced analytically from T but has to be sampled. Standard algorithms such as expectation-maximization (EM) become very computational intensive in that context. It can nevertheless be tractable, esp. in two dimensions [13].

Cubic Spline Wavelets. Cubic spline wavelets are one of the most useful wavelets in image processing and computer graphics, especially for reconstruction tasks. This popularity is due to the fact that it is smooth, symmetric, and can be efficiently implemented in the spatial domain by a so-called lifting scheme [14].

Wavelets are 1D creatures. In higher dimensions, the wavelet transform is done through multiresolution analysis (MRA), which alternates unidimensional transforms along the canonical axes. MRA yields basis functions $\Psi_{\ell, i}$ that are tensor products of unidimensional wavelets and cubic B-splines. Details on this standard procedure can be found e.g. in [15].

In this work, 3D displacement fields representing M and T are decomposed on a cubic spline MRA wavelet basis $\Psi_{\ell, i}$. The transform T is written as

$$T(\mathbf{x}) = \sum_{\ell=1}^L \sum_{i=1}^{N_\ell} \Theta^{\ell,i} \Psi_{\ell,i}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^3, \Theta^{\ell,i} \in \mathbf{R}^3, \quad (3)$$

where L is the total number of multiresolution levels, which depends on the image size, N_ℓ is the number of basis functions at level ℓ , and $\Theta^{\ell,i}$ is the coefficient of the basis function $\Psi_{\ell,i}$. M is decomposed similarly, with coefficients $\theta^{\ell,i}$. The very same basis has already been used in the past to register images in [16].

Regularization. The registration is regularized using the following L_1 penalty on the wavelet coefficients:

$$R(T) = \lambda_R \sum_{\ell=1}^L \sum_{i=1}^{N_\ell} 2^\ell \|S\Theta^{\ell,i}\|_1, \quad (4)$$

where λ_R is a regularization strength parameter, $S = (s_{i,j}) \in \mathbf{R}^{3 \times 3}$ is a diagonal matrix where $s_{i,i}$ is the voxel size along the i -th axis, and $\|\mathbf{x}\|_1 = \sum_{i=1}^3 |x_i|$.

The 2^ℓ factor under the sum introduces a linear increase of the penalty with the frequency of the wavelets, which was chosen here to imitate a first-order penalty like linear elasticity. This is of course only an approximation, but it is quite satisfactory for registration purposes.

The use of an L_1 penalty is a well-known trick to optimize the number of non-zero parameters [8]. This property comes from the fact that this norm has a gradient discontinuity at its minimum, zero. When added to a smooth similarity criterion, the minimum escapes from zero only when the similarity strongly suggests so. A downside of the L_1 norm is its anisotropy. The (non-squared) L_2 norm would be a better choice in that respect, but at the price of worse compression ratios—coefficients in all directions have to be simultaneously small to be zeroed.

Coupling M and T . The relationship between matches M and the transform T —or more precisely, the conditional probability $p(M|T)$ —is the key feature of iconic feature registration. It is also the harder to model since it encompasses many sources of geometrical uncertainties and model inadequacies.

In the following, this relationship is described by an Euclidean distance between the wavelet coefficients of M and T :

$$d(M, T) = \lambda_d \sum_{\ell=1}^L \sum_{i=1}^{N_\ell} 2^{-\ell} \|S(\theta^{\ell,i} - \Theta^{\ell,i})\|_2^2, \quad (5)$$

with $\|\mathbf{x}\|_2^2 = \sum_{i=1}^3 x_i^2$. The factor $2^{-\ell}$ relates to the size of the spatial support of the wavelet — if cubic splines were orthogonal, this formula would be simply proportional to $\|M - T\|_2^2$.

2.2 Iconic Feature Registration Without Alternating Minimization

IFR algorithms, following their geometrical counterparts like the iterative closest point (ICP) algorithm, usually proceed by alternating two steps: a matching

step, where optimal matching pairs are sought, and a fitting step, where the transform is updated. These steps correspond resp. to the minimization of (1) w.r.t. M and T .

Sometimes, however, we may derive a closed-form formula for the optimal transform $T^*(M)$ that depends on M when both the relationship $p(M|T)$ between matches and the transform and the prior $p(T)$ on the transform are simple enough. Replacing T by $T^*(M)$ in (1) then yields an energy that depends on M only, and we are left optimizing the matches M . When the optimal M^* is found, the optimal transform is simply given by $T^*(M^*)$.

Avoiding the alternating minimization constraint can be computationally attractive. Alternating minimization forces each minimization step to lie within given subspaces of the parameter space, which forbids looking in directions given by linear combinations of gradients typically used by fast optimization procedures. Convergence to the minimum is thus linear, i.e. slow.

Optimizing on M has yet some drawbacks. The matches are typically not constrained much, meaning that we may encounter many local minima. Also, a lot of time may be wasted finding the finer details of M at the highest frequencies, while these details, being heavily penalized by (4), do not impact the transform. This can be avoided by including a stopping criterion holding on $T^*(M)$.

Application to Our Algorithm. From equations (4) and (5), we deduce a closed-form formula for $T^*(M)$. Noting $\Theta_j^{\ell,i}$ and $\theta_j^{\ell,i}$ the j -th component, $j \in \{1, 2, 3\}$, of resp. $\Theta^{\ell,i}$ and $\theta^{\ell,i}$, we have

$$\Theta_j^{\ell,i} = \begin{cases} \theta_j^{\ell,i} - K/s_{j,j} & \text{if } s_{j,j}\theta_j^{\ell,i} > K \\ -\theta_j^{\ell,i} + K/s_{j,j} & \text{if } s_{j,j}\theta_j^{\ell,i} < -K \\ 0 & \text{otherwise} \end{cases}, \quad \text{with } K = 2^{2\ell-1} \frac{\lambda_R}{\lambda_d}. \quad (6)$$

Replacing T with $T^*(M)$ into (1), our energy becomes

$$E(M, T^*(M)) = S(I, J \circ M) + \lambda_d \sum_{\ell=1}^L \sum_{i=1}^{N_\ell} \sum_{j=1}^3 2^\ell f_K \left(s_{j,j} \theta_j^{\ell,i} \right), \quad (7)$$

with

$$f_K(x) = \begin{cases} x^2 & \text{if } |x| < K \\ 2K|x - K| + K^2 & \text{if } |x| \geq K \end{cases}. \quad (8)$$

Therefore, in spite of the L_1 constraint on the wavelet coefficients of T , it turns out that the objective function on M is smooth, as f_K has a continuous derivative everywhere — it happens to be Huber’s M-estimator penalty function. Standard optimization algorithms can be used for its optimization.

Minimization. In the following, we minimize $E(M, T^*(M))$ using a conjugate gradient descent. The computation of the partial derivatives $\partial_{\theta_j^{\ell,i}} S(I, J \circ M)$ typically requires the convolution of partial derivatives w.r.t. the displacement of a single point $\partial_{M(\mathbf{x})} S(I, J \circ M)$ with the basis functions $\Psi_{\ell,i}$. These partial derivatives are not given by the analysis transform because cubic spline wavelets are not orthogonal : the partial derivatives have to correspond to convolutions

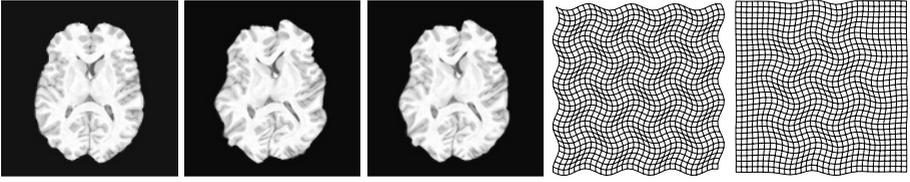


Fig. 2. Synthetic registration experiment. **From left to right:** the original volume, the deformed volume, the original volume registered onto the deformed volume, the synthetic deformation, and the recovered deformation. The transform is recovered where the images bear sufficient information.

with cubic spline wavelets, not their duals. While not standard, it is not difficult to build a direct wavelet lifting transform corresponding to these convolutions.

The energy is minimized in a multiresolution fashion. A Gaussian pyramid of images I and J is constructed. At each level of the pyramid, the transform is estimated, starting from the top. When convergence is reached, the estimated transform is propagated to the next level, and registration is resumed.

3 Experiments

In all the following experiments, volumes are $256 \times 256 \times 256$ T_1 -weighted MRI. We used the Gaussian-weighted local correlation coefficient as the similarity measure S [17]. At each level of the Gaussian pyramid, the basis of the last multiresolution level was systematically ignored, in part to address the problem mentioned in section 2.2 that fine details of M are long to compute yet have the least impact on T .

3.1 A Synthetic Experiment

We first present a synthetic experiment where the ground truth is known. We deformed a MR volume of a brain with the sinusoidal deformation presented in Fig. 2. We manually tuned the λ_R and λ_d parameters until the results were visually satisfying. The registration took about 5 hours to complete. Results in Fig. 2 show that the algorithm is able to recover the transform pretty well, except of course for areas lying in the background. Yet to achieve this result, only 262,144 non-zero wavelet coefficients were used. That represents a vector for about 0.53% of the voxels.

3.2 Registration of Different Brains

The next experiment deals with volumes of the brain of two different individuals. The images have been corrected for non-uniform bias, and the brains have been segmented and aligned affinely. Parameters λ_R and λ_d have been chosen manually to correspond to a conservative regularization, which is necessary for multisubject brain registration. The registration took about five hours to complete the task on a 1.8MHz Pentium. Figure 3 shows slices of the original volume

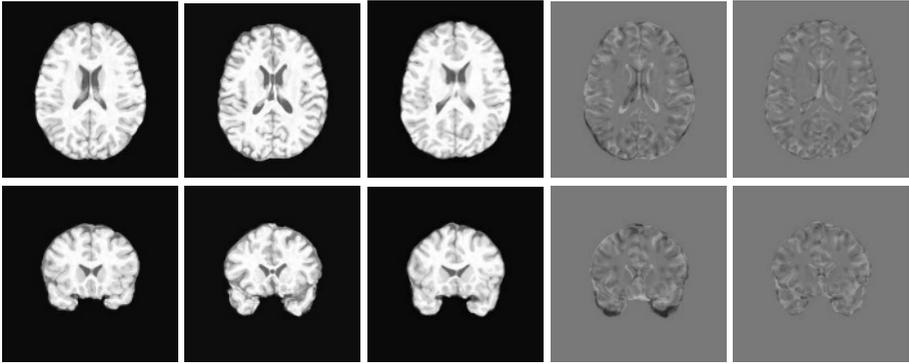


Fig. 3. Registration of two different brains. **From left to right:** The source volume; The target volume; The registered volume; Intensity difference between the source and the target volumes; and between the registered and the target volumes. Regularization strength was set high enough to prevent over-registration of brain sulci.

pair and of the registered volume, as well as intensity differences between volumes before and after registration.

The algorithm did a decent job at aligning brain volumes, while preserving unmatched structures such as the details of the cortical sulci. However, the registration used only 242,296 non-zero wavelet coefficients. Reported to the entire volume, this is a vector for about 0.48% of the voxels.

4 Conclusion

This article presented registration from a new point of view: the storage space of the generated deformations, which we believe will become more and more problematic with the growing popularity of nonrigid registration in group studies. We proposed a solution based on describing the transform in terms of wavelets which minimizes the number of non-zero coefficients.

Much work has still to be done to confirm the behavior of the approach. The impact of the anisotropy of the L_1 constraint should be evaluated. Extensive studies should also give an idea of compression ratios we can expect depending on the imaging modalities and the desired regularity and accuracy of the transform. Finally, computation speed-up may be obtained by tuning the optimization precision depending on the multiresolution level, avoiding spending too much time finely estimating the highest frequencies of the matches that do not impact much the transform.

References

1. Ashburner, J., Friston, K.F.: Voxel-based morphometry — The methods. *NeuroImage* **12** (2000) 1427 – 1442
2. Collins, D.L., Evans, A.C.: ANIMAL validation and applications of nonlinear registration based segmentation. *Int. J. of Pattern Recognition and Artificial Intelligence* **11** (1997) 1271 – 1294

3. Thompson, P., Toga, A.W.: Detection, visualization and animation of abnormal anatomic structures with a deformable probabilistic atlas based on random vector field transformations. *Medical Image Analysis* **1** (1997) 271 – 294
4. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3D statistical deformation models of the brain using nonrigid registration. *IEEE Trans. Med. Im.* **22** (2003)
5. Cathier, P., Fung, G., J.-F., M.: Population classification based on global brain features inferred from massive T1 MRI registration. In: Proc. of HBM. (2006)
6. Szeliski, R., Coughlan, J.: Spline-based image registration. *Int. J. of Comp. Vision* **22** (1997) 199–218
7. Kybic, J., Unser, M.: Fast parametric elastic image registration. *IEEE Trans. Im. Proc.* **12** (2003) 1427 – 1442
8. Tibshirani, R.: Regression selection and shrinkage via the lasso. *J. Royal Stat. Soc.* **1** (1995) 267 – 288
9. Ng, L., Solo, V.: Optical flow estimation using wavelet zeroing. In: Proc. of ICIP. (1999) III:722 – 726
10. Cachier, P., Bardinet, E., Dormont, D., Pennec, X., Ayache, N.: Iconic feature based nonrigid registration: The PASHA algorithm. *Comp. Vision Im. Underst.* **89** (2003) 272 – 298
11. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis* **2** (1998) 243–260
12. Hayton, P.M., Brady, M., Smith, S.M., Moore, N.: A non-rigid registration algorithm for dynamic breast MR images. *Artificial Intelligence* **114** (1999) 124–156
13. Singh, A., Allen, P.: Image-flow computation: An estimation-theoretic framework and a unified perspective. *CVGIP: Im. Underst.* **56** (1992) 152 – 177
14. Daubechies, I., Sweldens, W.: Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.* **3** (1998) 247 – 269
15. Mallat, S.G., ed.: *A wavelet tour of signal processing*. Academic Press (1998)
16. Wu, Y.T., Kanade, T., Li, C.C., Cohn, J.: Image registration using wavelet-based motion model. *Int. J. of Comp. Vision* **38** (2000) 129 – 152
17. Cachier, P., Pennec, X.: 3D non-rigid registration by gradient descent on a Gaussian-windowed similarity measure using convolutions. In: Proc. of MMBIA (2000) 182 – 189