A Modular Registration Algorithm for Medical Images

Silvia Bertoluzza[1] and Giulia Maggi[2]

[1] CNR IMATI "Enrico Magenes", Pavia, Italy
silvia.bertoluzza imati.cnr.it
[2] Dipartimento di Matematica, University of Pavia, Italy
giulia.maggi unipv.it

Abstract. The aim of this communication is to present the design of a code for image registration based on a modular structure which allows to easily combine and interchange different image models, transformation classes and image "distance" functionals, which are dealt with by independent modules which can be implemented transparently to each other. The code is tested by comparing several registration strategies on medical images.

Keywords: registration algorithm, medical imaging, optimization

1 Introduction

Diagnostic imaging is an invaluable tool in medicine today. With the increasing size and number of medical images available to the physician, the use of computers in facilitating their processing and analysis has become necessary. In particular, computer algorithms for the delineation of anatomical structures and other regions of interest are a key component in assisting and automating specific radiological tasks.

Automatic or semi-automatic delineation of the contours in an image is one of the typical tasks related to the image segmentation field. There exist several methods that can be used (see for example [3],[4],[5],[6],[7],[8]); the most suitable method for a given case depends clearly on the image data, the object imaged and the type of desired output information. We focused on atlas-guided approaches which are a powerful tool for medical image segmentation when a standard atlas or template is available ([9],[10],[11],[12],[13]). The use of the atlas based approach reduces the problem of the segmentation of an anatomical image to an image registration problem.

A huge variety of registration techniques has been proposed in the literature. Among the different ingredients that, combined together, make up a registration algorithm we can mention the way the images are modeled, the criterion that we use to measure whether two images are well registered or not, the way we allow the images to be deformed in the registration process, the particular

optimization strategy that is used to perform the registration. These ingredients can be combined together in many different ways, resulting in different, more or less robust, more or less efficient registration algorithms. Our aim is to design a code allowing to easily interchange these ingredient, so that many different registration techniques (corresponding to different combinations) are easily included in the code. In order to do so we will present a code with a modular structure constructed in such a way that each module takes care of one of the "ingredients" and is completely transparents to the other modules.

2 The Modular Algorithm

The problem that we want to solve is the following: given two monochromatic images $T : \Omega \to \mathbb{R}^+$ and $R : \Omega \to \mathbb{R}^+$ (where for the sake of simplicity we set $\Omega = (0,1)^2$), we look for a mapping $\vartheta = (\vartheta_1, \vartheta_2) : \Omega \to \Omega$ such that $T \circ \vartheta$ and R are as close as possible. In order to practically solve this problem we need to make a certain number of choices. We first need to choose a suitable *class of admissible mappings*. In a parametric approach this is equivalent to selecting a function $\Theta : \Omega \times \mathbb{R}^N \to \Omega$, with $\mathbb{R}^N$ playing the role of a parameter space. The set of admissible transformations will be

$$\mathfrak{M} = \{\vartheta_\alpha : \Omega \to \Omega : \ \vartheta_\alpha(x) = \Theta(x, \alpha), \quad \alpha \in \mathbb{R}^N\}.$$

Many choices are available in the literature, ranging from rigid transformations, to affine transformations to different classes of curved transformations (B-splines, Interpolating wavelets,...).

Another key element of the registration algorithm is the choice of the *optimization criterion* that will be underlying the algorithm itself. In other words the user will have to choose a criterion for measuring whether or not two images are close. Rigorously speaking we will have to choose a functional $d : \mathfrak{I} \times \mathfrak{I}$ ($\mathfrak{I}$ denoting the functional space which will be chosen to represent images, usually the space $L^2(\Omega)$ of finite energy functions or the space $L^\infty(\Omega)$ of bounded functions) and we will associate to a parameter vector $\alpha \in \mathbb{R}^N$ its cost $c(\alpha)$ defined as

$$c(\alpha) = d(T \circ \vartheta_\alpha, R).$$

The semi-continuos registration problem will then read: find $\alpha^\star \in \mathbb{R}^N$

$$\alpha^\star = \arg \min_{\alpha \in \mathbb{R}^N} c(\alpha). \tag{1}$$

The optimization problem can be solved by resorting to one of several possible optimization algorithms (whose choice possibly depends on the choice of the "distance functional" d). Whatever the choice, the user will have in general to provide methods to compute the cost function $c(\alpha)$ as well as its gradient $\nabla c(\alpha)$.

There will then be the need to deforme the input image and to compute its gradient. Since the input images will, in practice, be available only in (discrete) digital form, we will then have to select an *image model* allowing to somehow evaluate a sampled image, as well as its gradient, at arbitrary points in Ω.

The idea underlying the modular design of our code is to separate as much as possible the implementation related to this three different components. We will then have three modules: a mapping module, an distance functional module and an image model module. As far as the evaluation of the cost functional is concerned, the actions of the three modules will be quite simple. Assume that the images T and R are sampled on a sampling grid $\mathcal{G}$.

- The distance module evaluates the functional d on two images sampled on the grid $\mathcal{G}$
- The image model module samples the image T on an arbitrary grid $\mathcal{X}$
- The mapping module evaluates the mapped grid $\mathcal{G}_\alpha = \vartheta_\alpha(\mathcal{G})$.

The evaluation of the cost functional is then performed in three steps:

- Given α call the mapping module and obtain $\mathcal{G}_\alpha$.
- Call the image model module and sample T on the grid $\mathcal{G}_\alpha$ (which is equivalent to sampling $T \circ \theta_\alpha$); the result is a digital image T_α, sampled on the grid $\mathcal{G}$.
- Call the distance module and evaluate $c(a) = d(T_\alpha, R)$.

Let us now consider the problem of evaluating the gradient of the cost functional. In order to see how such a computation can be decoupled in the independent action of the three modules, it is convenient to introduce the functional $\mathcal{J} : \mathfrak{I} \to \mathbb{R}$ defined as

$$\mathcal{J}(S) = d(S, R),$$

with R fixed. To fix the idea let $\mathfrak{I} = L^2(\Omega)$. We recall that the Fréchet derivative of $\mathcal{J}$ evaluated at S ($S \in L^2(\Omega)$), is the element $D\mathcal{J}[S]$ of $L^2(\Omega)$ such that for all $\Delta \in L^2(\Omega)$ we have

$$\mathcal{J}(S + \Delta) = \mathcal{J}(S) + \int_\Omega D\mathcal{J}[S](x)\Delta(x)\,dx + o(\|\Delta\|_{L^2(\Omega)}).$$

If we apply the chain rule to the cost function $c(\alpha) = \mathcal{J}(T \circ \vartheta_\alpha)$ we obtain the following expression for the gradient $\nabla c(\alpha) = (\partial_{\alpha_1} c(\alpha), \cdots, \partial \alpha_N c(\alpha))$:

$$\nabla c(\alpha) = \int_\Omega D\mathcal{J}[T \circ \vartheta_\alpha](x) \nabla T(\vartheta_\alpha(x)) \mathbb{J}_\alpha \Theta(x; \alpha)\,dx, \tag{2}$$

where $\nabla T(x) = (\partial_{x_1} T(x), \partial_{x_2} T(x))$ is the spatial gradient of the image T and $\mathbb{J}_\alpha \Theta(x; \alpha)$ is the Jacobian of the mapping function Θ with respect to the variable α

$$\mathbb{J}_\alpha \Theta(x, \alpha) = \begin{pmatrix} \partial_{\alpha_1} \Theta_1(x, \alpha) & \cdots & \partial_{\alpha_N} \Theta_1(x, \alpha) \\ \partial_{\alpha_1} \Theta_2(x, \alpha) & \cdots & \partial_{\alpha_N} \Theta_2(x, \alpha) \end{pmatrix}.$$

The actions that we will need the three modules to perform are then respectively

- The distance module will have to compute the Fréchet derivative $D\mathcal{J}[S]$ at a given image S (S sampled on the grid $\mathcal{G}$). The result must be itself an $L^2(\Omega)$ function, also sampled on $\mathcal{G}$.
- The image model module must sample the gradient of T at an arbitrary grid $\mathcal{X}$
- The mapping module must evaluate the Jacobian $\mathbb{J}_\alpha\Theta(x;\alpha)$ at the points of the grid $\mathcal{G}$

Then, the evaluation of the gradient of the cost functional c is performed as follows.

- Given α call the mapping module and evaluate $\mathbb{J}_\alpha\Theta(x;\alpha)$ at the points of the grid $\mathcal{G}$
- Call the image module and evaluate ∇T at the points of the deformed grid $\mathcal{G}_\alpha = \vartheta_\alpha(\mathcal{G})$.
- Call the distance module and evaluate $D\mathcal{J}[T_\alpha]$ at the grid points of $\mathcal{G}$
- Compute $\nabla c(\alpha)$ by numerically evaluating the integral (2)

$$\nabla c(\alpha) \sim \sum_{x_k \in \mathcal{G}} \delta^2 D\mathcal{J}[T_\alpha](x_k)\nabla T(\vartheta_\alpha(x_k))\mathbb{J}_\alpha^T\vartheta(x_k;\alpha),$$

  where δ^2 denotes the area of the pixels of the sampling grid $\mathcal{G}$

Since, as a result of its modularity, our code will be able to include a wide variety of different cost functionals, we decided to solve the optimization problem (1) by resorting to a Trust Region method (for which the assumptions on the cost functional are quite weak). More precisely we made use of the Matlab function `fminunc`, which we combined with the multilevel approach proposed by Thevenaz in [2].

3 The three modules

Let us now give a closer look to the three modules. Remark that the three modules are completely transparent to each other. As a result the choice of the class of admissible mappings, of the distance functional and of the image model can be carried out completely independently one from the other. One module can be replaced without any modifications to the other modules or to the overall optimization procedure. In the following for $\mathcal{X} = (x_1, \cdots, x_M)$ indicating a set of points, and for $g : \Omega \to \mathbb{R}^L$ indicating a vector or matrix valued function, we will write $g(\mathcal{X})$ to indicate the hypervector

$$g(\mathcal{X}) = (g(x_1), \cdots, g(x_N))$$

where each $g(x_i)$ is itself a vector or a matrix. In particular, letting $\mathcal{G}$ denote the sampling grid, containing M grid points (that is the image has a resolution of $M_1 \times M_2$ pixels with $M_1 M_2 = M$, writing $g(\mathcal{G})$ will stand for the sampling of g on $\mathcal{G}$.

3.1 The Mapping module

The mapping module deals with the parametric mapping Θ. It takes as input a parameter vector $\alpha \in \mathbb{R}^N$ and returns two outputs. The first one is the deformed grid $\mathcal{G}_\alpha = \vartheta_\alpha(\mathcal{G}) = \Theta(\mathcal{G}, \alpha)$. The second one is the Jacobian with respect to the variable α of Θ, evaluated at the points of the grid $\mathcal{G}$, that is

$$\mathbb{J}_\alpha \Theta(\mathcal{G}, \alpha) = \begin{pmatrix} \partial_{\alpha_1}\Theta_1(\mathcal{G}, \alpha) & \cdots & \partial_{\alpha_N}\Theta_1(\mathcal{G}, \alpha) \\ \partial_{\alpha_1}\Theta_2(\mathcal{G}, \alpha) & \cdots & \partial_{\alpha_N}\Theta_2(\mathcal{G}, \alpha) \end{pmatrix}.$$

Remark that in the quite frequent case that the dependence on α of the function Θ is linear, that is if Θ takes the form

$$\Theta(x, \alpha) = \sum_i \alpha_i \phi^i(x), \qquad \phi^i(x) = (\phi_1^i(x), \phi_2^i(x)),$$

then $\mathbb{J}_\alpha \Theta(\mathcal{G}, \alpha)$ is independent of α and takes the form

$$\mathbb{J}_\alpha \Theta(\mathcal{G}, \alpha) = \begin{pmatrix} \phi_1^1(\mathcal{G}) & \cdots & \phi_1^N(\mathcal{G}) \\ \phi_2^1(\mathcal{G}) & \cdots & \phi_2^N(\mathcal{G}) \end{pmatrix}.$$

In the tests that we present in the following we chose the space of admissible mappings to be the space of Interpolating wavelets ([26, 27]) as proposed by David Donoho.

3.2 The Image Model module

The image model module takes care of applying the deformation constructed by the mapping module to the actual image and of evaluating its gradient at the points of the deformed grid. Sampling the deformed image $T \circ \vartheta_\alpha$ at the points of the grid $\mathcal{G}$ is equivalent to sampling T at the points of the deformed grid $\mathcal{G}_\alpha$. The Image module takes in input an arbitrary grid $\mathcal{X}$ and returns the result of resampling the template image T (which is sampled on the grid $\mathcal{G}$), as well as its gradient ∇T, at the points of $\mathcal{X}$. This reduces to performing some kind of interpolation and of numerical differentiations. In our code we implemented three options: B-spline interpolation, bicubic interpolation, and average interpolation. The latest seems particularly well suited to deal with the resampling of images, since it is based on expressing the image as a Fourier series where the coefficient corresponding to each pixel is the mean value of the image on the pixel itself

3.3 Distance Measure

The distance measure module actually computes the "distance" (we recall that the functional d involved in the definition of the cost functional c might not be an actual distance) of the deformed template image and the reference image. It takes as input an image S (sampled at the grid $\mathcal{G}$) and returns the value $\mathcal{J}(S) = d(S, R)$. It also returns $\mathcal{J}[S](\mathcal{G})$, that is the Fréchet derivative of $\mathcal{J}$ in

S (which is, we recall, a function in $L^2(\Omega)$), sampled at the grid points of the sampling grid $\mathcal{G}$.

Within the code we implemented a wide class of different "distance" measures, starting from the simplest Least Squares measure

$$d_{LS}(S,R) = \int_\Omega |S(x) - R(x)|^2 \, dx \sim \sum_{x_k \in \mathcal{G}} \delta^2 |S(x_k) - R(x_k)|^2,$$

up to the cost functional based on Mutual Information, a quite complex and expensive but extremely robust measure of the closeness of two images, based on statistical model for images which treats the generation of images as a random process assumed to be homogeneous in space

$$d_{MI}(S,R) = -\sum_{\kappa,\iota} h(\iota,\kappa) \log\left(\frac{h(\iota,\kappa)}{h_T(\iota)\cdot h_R(\kappa)}\right)$$

h_R and h_T being the histograms of R and T respectively, and h being the joint histogram of the two images (the histograms are evaluated using the approach of [1]).

Other distances that we implemented in the code are the SSIM distance ([19]), the distance in Besov spaces, and a new distance, which can be seen as a generalization of the SSIM distance, obtained by applying a bandwise divisive renormalization to the Besov distance.

4 Numerical Tests

We performed extended tests of the different combinations of the distance functions, image interpolation scheme and algorithm for optimization presented above. We tested the different combinations on some real life medical images provided by Istituto dei Tumori of Milan. In particular we considered 512×512 medical images that have been selected from two set of CT studies, of the same patient (intra-subject registration) acquired from different CT machines.

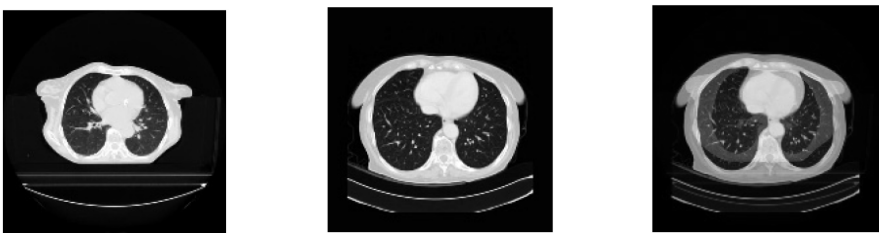

Fig. 1. Reference image, Template image and their superposition.

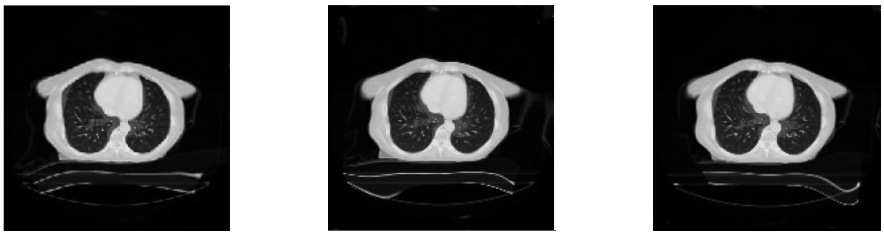

Fig. 2. Results by applying Mutual Information distance measure, SSIM ($p = 10$) based metric and Normalized Besov norm (p=3).

5 Conclusion

We propose a modular design resulting in a very flexible algorithm for image registration, which we test on real life medical images. The design allows us to combine different distance measure, image model and mapping classes by implementing such components of the algorithm as independent modules.
Thanks to this design we were then able to compare a wide range of different algorithms, ranging from Mutual Information based ones, which behave in an extremely good way, but at an extremely high computational cost, to Least Square, to some new method based on the SSIM and on a generalized SSIM metrics, which still display a quite good behaviour at a much smaller cost, providing therefore a good compromise between the robustness of the Mutual Information and the computational efficiency of Least Squares.

References

1. Parzen, E.: On estimation of a probability density function and mode. Annals of Mathematical Statistics 33, 1065–1076 (1962)
2. Thévenaz P., Unser M.: Optimization of Mutual Information for Multiresolution Image Registration. In IEEE Transactions on Image Processing, vol. 9, no. 12, pp. 2083–2099 (2000)
3. Khoo V.S., Dearnaley D.P., Finnigan D.J., Padhani A., Tanner S.F., Leach M.O.: Magnetic Resonance Imaging (MRI): considerations and applications in Radiotherapy treatment planning. Radiother. Oncol. (1997)
4. Larie S.M., Abukmeil S.S.: Brain abnormality in schizophrenia: a systematic and quantative review of volumetric magnetic resonance imaging studies. J. Psych., 172:110–120 (1998)
5. Muller-Gartner H.W., Links J.M. et al.: Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. J. Cereb. Blood Flow Metab., pp. 571–583 (1992)
6. Taylor P.: Invited review: computer aids for decision-making in diagnostic radiology. A literature review. British J. Radiology, pp. 945–947 (1995)
7. Wort A.J.,Makris N., Caviness V.S., Kennedy D.N.: Neuroanatomical segmentation in MRI: technological objectives. Int. J. Patt. Rec. Art. Intel. (1997)

8. Zijdenbos A.P., Dawant B.M., Margolin R.A., Palmer A.C.: Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Transactions on Medical Imaging, Vol. 13, No. 4, pp. 716–724 (1994)
9. Rohlfing T., Brandt R., Menzel R., Russakoff D.B., Maurer Jr. C.R.: Quo Vadis, Atlas-Based Segmentation?. The Handhajnaletal of Medical Image Analysis–Volume III: Registration Models. Kluwer Academic / Plenum Publishers, pp. 435–486 (2005)
10. Qatarneh S.M., Noz M.E., Hyodynmaa S., et al.: Evaluation of a segmentation procedure to delineate organs for use in construction of a radiation therapy planning atlas. Int. J. Med. Inform.;69:39–55 (2003)
11. Klein S.T., Van der Heide U., Lips I., et al.: Automatic segmentation of the prostate on 3D MR images by atlas matching using localized mutual information. Med. Phys. (2008)
12. Commowick O., Grégoire V., Malandain G.: Atlas-based delineation of lymphonode levels in head and neck computed tomography images. Radiother. Oncol. (2008)
13. Isambert A., Grégoire V., Bidault F., et al.: Atlas-based automatic segmentation (ABAS) of head and neck (H& N) structures in conformal radiotherapy: Atlas development and preliminary results in clinical context. In: ESTRO (2007)
14. Hajnal J.V., Hill D.L.G., Hawkes D.J.: Medical Image Registration. CRC Press, ISBN 0-8493-0064-9 Boca Raton London New York Washington DC (2001)
15. Zitova B., Flusser J.: Image registration methods: a survey. Image and Vision Computing 21, pp. 977-1000 (2003)
16. Van Den Elsen P.A., Pol E.J.D., Viergever M.A.: Medical image matching: a review with classification. IEEE Engineering in medicine and biology, 12(1), pp. 26–39 (1993)
17. Maintz J.B.A., Viergever M.A.: A Survey of Medical Image Registration. Medical Image Analysis, Vol.2, n. 1, pp. 1–36, Oxford University Press (1998)
18. Brown Gottesfeld L.: A survey of image Registration Techniques. ACM Computing surveys 24, pp. 325–376 (1992)
19. Brunet D., Vrscay E.R., Wang Z.: A class of image metrics based on the structural similarity quality index. International Conference on Image Analysis and Recognition, Burnaby, BC, Canada, June (2011)
20. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing 13(4)(2004)
21. Bruckner T., Lucht R., Brix G.: Comparison of rigid and elastic matching of dynamic magnetic resonance mammographic images by mutual information. Medical Physics (2000)
22. Bro-Nielsen M.: Medical Image Registration and Surgery Simulation. PhD thesis Department of Mathematical Modelling, Technical University of Denmark, July (1997)
23. Cohen I., Cohen L., Ayache N.: Using deformable surfaces to segment 3D images and infer differential structures. In CVGIP: Image Understanding, September (1992)
24. Dubuc S.: Interpolation through an iterative scheme. In: J. Math. Anal. and Appl. 114, pp. 185–204 (1986)
25. Deslauriers G., Dubuc S.: Interpolation dyadique. Fractals, dimensions non-entires et applications, pp. 44–55. Masson, Paris (1987)
26. Donoho D.: Interpolating Wavelet Transform. Technical Report, Stanford University, (1992) `http://www-stat.stanford.edu/~donoho/Reports/1992/interpol.pdf`

27. Bertoluzza S., Naldi G.: A wavelet collocation method for the numerical solution of partial differential equations. Appl. and Comput. Harm. Anal. 3 1–9 (1996)