

Intensity Gradient Based Registration and Fusion of Multi-modal Images

Eldad Haber¹ and Jan Modersitzki²

¹ Mathematics and Computer Science, Emory University, Atlanta, GA, USA
haber@mathcs.emory.edu

² Institute of Mathematics, Lübeck, Germany
modersitzki@math.uni-luebeck.de

Abstract. A particular problem in image registration arises for multi-modal images taken from different imaging devices and/or modalities. Starting in 1995, mutual information has shown to be a very successful distance measure for multi-modal image registration. However, mutual information has also a number of well-known drawbacks. Its main disadvantage is that it is known to be highly non-convex and has typically many local maxima.

This observation motivate us to seek a different image similarity measure which is better suited for optimization but as well capable to handle multi-modal images. In this work we investigate an alternative distance measure which is based on normalized gradients and compare its performance to Mutual Information. We call the new distance measure Normalized Gradient Fields (NGF).

1 Introduction

Image registration is one of today's challenging medical image processing problems. The objective is to find a geometrical transformation that aligns points in one view of an object with corresponding points in another view of the same object or a similar one. An important area is the need for combining information from multiple images acquired using different modalities, sometimes called fusion. Typical examples include the fusion of computer tomography (CT) and magnetic resonance (MRI) images or of CT and positron emission tomography (PET). In the past two decades computerized image registration has played an increasingly important role in medical imaging (see, e.g., [2,10] and references therein).

One of the challenges in image registration arises for multi-modal images taken from different imaging devices and/or modalities. In many applications, the relation between the gray values of multi-modal images is complex and a functional dependency is generally missing. However, for the images under consideration, the gray value patterns are typically not completely arbitrary or random. This observation motivated the usage of mutual information (MI) as a distance measure between two images cf. [4,18]. Starting in 1995, mutual information has shown to be a successful distance measure for multi-modal image registration.

Therefore, it is considered to be the state-of-the-art approach to multi-modal image registration.

However, mutual information has a number of well-known drawbacks; cf. e.g., [14,13,12,17,20]. Firstly, mutual information is known to be highly non-convex and has typically many local maxima; see for example the discussion in [10], [20], and Section 3. Therefore, the non-convexity and hence non-linearity of the registration problem is enhanced by the usage of mutual information. Secondly, mutual information is defined via the joint density of the gray value distribution, and therefore, approximations of the density are required. These approximations are nontrivial to compute and typically involve some very sensitive smoothing parameters (e.g. a binning size or a Parzen window width, see [16]). Finally, mutual information decouples the gray value from the location information. Therefore judging the output of the registration process is difficult. These difficulties had stem a vast amount of research into mutual information registration, introducing many nuisance parameters to help and bypass at least some of the difficulties; see, e.g. [14].

These observations motivate us to seek a different image similarity measure which is capable to handle multi-modal images but better suited for optimization and interpretation. In this paper we investigate an alternative distance measure which is based on normalized gradients. As we show, the alternative approach is deterministic, simpler, easier to interpret, fast and straightforward to implement, faster to compute, and also much more suitable to optimization.

The idea of using derivatives to characterize similarity between images is based on the observation that image structure can be defined by intensity changes. The idea is not new. In inverse problems arising in geophysics, previous work on joint inversion [9,21,7] discussed the use of gradients in order to solve/fuse inverse problems of different modalities. In image registration, a more general framework similar to the one previously suggested for joint inversion was given in [6] and [14]. Our approach is similar to [14] however, we have some main differences. Unlike the work in [14] we use only gradient information and avoid using mutual information. Furthermore, our computation of the gradient is less sensitive and allow us to deal with noisy images. This allows us to use simple optimization routines (e.g. Gauss Newton).

The rest of the paper is organized as follows: In Section 2 we shortly lay the mathematical foundation of image registration. Section 3 presents an illustrative example showing some of the drawbacks of mutual information. In Section 4 we discuss the proposed alternative image distance measure. Finally, in Section 5 we demonstrate the effectiveness of our method.

2 The Mathematical Setting

Given a reference image R and a template image T , the goal of image registration is to find a “reasonable” transformation, φ , such that the “distance” between the reference image and a deformed template image is small.

In this paper we use both affine linear and nonparametric registration. In the linear registration, we set the transformation φ to

$$\varphi(\gamma, \mathbf{x}) = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \gamma_5 \\ \gamma_6 \end{pmatrix}. \tag{1}$$

Given a distance measure \mathcal{D} , the registration problem is to find a minimizer γ of

$$f(\gamma) := \mathcal{D}\left(R(\cdot), T(\varphi(\gamma, \cdot))\right). \tag{2}$$

For nonparametric registration we set $\varphi = \mathbf{x} + \mathbf{u}(\mathbf{x})$ and use regularization minimizing

$$f(\gamma) := \mathcal{D}\left(R(\mathbf{x}), T(\mathbf{x} + \mathbf{u})\right) + \alpha\mathcal{S}(\mathbf{u}). \tag{3}$$

where $\alpha\mathcal{S}(\mathbf{u})$ is a regularization term such as elastic (see [10]).

3 An Illustrative Example

To emphasize the difficulty explained above, we present an illustrative example. Figure 1 shows a T1 and a T2 weighted magnetic resonance image (MRI) of a brain. Since the image modalities are different, a direct comparison of gray values is not advisable and we hence study a mutual information based approach.

Figure 1c) displays our approximation to the joint density which is based on a kernel estimator, where the kernel is a compactly supported smooth function. Note that the joint density is completely unrelated to the spatial image

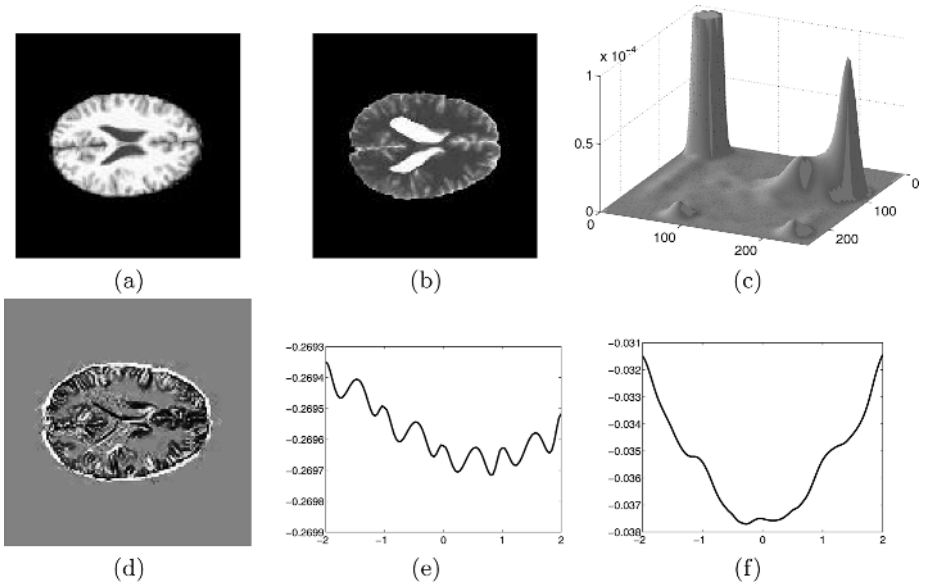


Fig. 1. Distance measures versus shifts; (a)Original BrainWeb [3] T1 (b)Original BrainWeb [3] T2 (c) the joint density approximation for the images, (d) the normalized gradient field $d^d(T, R)$ for the images, (e) negative mutual information versus shift, (f) normalized gradient field versus shift

content (though there is some interpretation [11]). We now slide the template along the horizontal axis. In the language of equation (1), we fix $\gamma_{1,\dots,5}$ and obtain the transformed image by changing γ_6 . Figure 1e) shows the negative mutual information versus the shift ranging from -2 to 2 pixels.

This figure clearly demonstrates that mutual information is a highly non-convex function with respect to the shift parameter. We have experimented with different interpolation methods and all have resulted in similar behavior. In particular, the curve suggests that there are many pronounced local minima which are closed in value to the global minimum. Our particular example is not by any means pathologic. Similar, and even less convex curves of Mutual Information appear also in [20] pp293 who used different interpolation.

Figure 1d) displays a typical visualization of our alternative NGF distance between R and T (discussed in the next section). Note that for the alternative distance measure, image differences are related to spatial positions. Figure 1f) shows the alternative distance measure versus the shift parameter. For this particular example, it is obvious that the alternative measure is capable for multi-modal registration and it is much better suited for optimization.

4 A Simple and Robust Alternative

The alternative multi-modal distance measure is based on the following simple though general interpretation of similarity:

two image are considered similar, if intensity changes occur at the same locations.

An image intensity change can be detected via the image gradient. Since the magnitude of the gradient is dependent upon the modality of the image, it would be unwise to base an image similarity measure on gradient magnitude. We therefore consider the normalized gradient field, i.e., the local orientation, which is purely geometric information.

$$\mathbf{n}(I, \mathbf{x}) := \begin{cases} \frac{\nabla I(\mathbf{x})}{\|\nabla I(\mathbf{x})\|}, & \nabla I(\mathbf{x}) \neq \mathbf{0}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

For two related points \mathbf{x} in R and $\varphi(\mathbf{x})$ in T or, equivalently, \mathbf{x} in $T \circ \varphi$, we look at the vectors $\mathbf{n}(R, \mathbf{x})$ and $\mathbf{n}(T \circ \varphi, \mathbf{x})$. These two vectors form an angle $\theta(\mathbf{x})$. Since the gradient fields are normalized, the inner product (dot-product) of the vectors is related to the cosine of this angle, while the norm of the outer product (cross-product) is related to the sine. In order to align the two images, we can either minimize the square of the sine or, equivalently, maximize the square of the cosine. Thus we define the following distance measures

$$d^c(T, R) = \|\mathbf{n}(R, \mathbf{x}) \times \mathbf{n}(T, \mathbf{x})\|^2; \quad \mathcal{D}^c(T, R) = \frac{1}{2} \int_{\Omega} d^c(T, R) \, d\mathbf{x}, \tag{5}$$

$$d^d(T, R) = \langle \mathbf{n}(R, \mathbf{x}), \mathbf{n}(T, \mathbf{x}) \rangle^2; \quad \mathcal{D}^d(T, R) = -\frac{1}{2} \int_{\Omega} d^d(T, R) \, d\mathbf{x}, \tag{6}$$

Note that from an optimization point of view, the distances \mathcal{D}^c or \mathcal{D}^d are equivalent.

The definition of the normalized gradient field (4) is not differentiable in areas where the image is constant and highly sensitive to small values of the gradient field. To avoid this problem we define the following regularized normalized gradient fields

$$\mathbf{n}_{\mathcal{E}}(I, \mathbf{x}) := \frac{\nabla I(\mathbf{x})}{\|\nabla I(\mathbf{x})\|_{\mathcal{E}}}, \quad \|\nabla I(\mathbf{x})\|_{\mathcal{E}} := \sqrt{\nabla I(\mathbf{x})^{\top} \nabla I(\mathbf{x}) + \mathcal{E}^2}. \quad (7)$$

In regions where \mathcal{E} is much larger than the gradients the maps $\mathbf{n}_{\mathcal{E}}(I, \mathbf{x})$ are almost zero and therefore do not have a significant effect on the measures \mathcal{D}^c or \mathcal{D}^d , respectively. Indeed, the choice of the parameter \mathcal{E} in (7) answers the question, “what can be interpreted as a jump?”. As suggested in [1], we propose the following *automatic choice*:

$$\mathcal{E} = \frac{\eta}{V} \int_{\Omega} |\nabla I(\mathbf{x})| \, d\mathbf{x}, \quad (8)$$

where η is the estimated noise level in the image and V is the volume of the domain Ω .

The measures (5) and (6) are based on local quantities and are easy to compute. Another advantage of these measures is that they are directly related to the resolution of the images. This property enables a straightforward multi-resolution approach. In addition, we can also provide plots of the distance fields d^c and d^d , which enables a further analysis of image similarity; see, e.g., Figure 1d). Note that in particular $d^c = 0$ everywhere if the images match perfectly. Therefore, if in some areas the function d^c takes large values, we know that these areas did not register well.

5 Numerical Experiments

In our numerical experience, we have used various examples with results along the same lines. Since it is impossible to show every result we restrict ourselves to three illustrative and representative examples.

In the first example we use the images in Figure 1. We take the T1 image and generate transformed versions of the image by using an affine linear transformation (1). We then use the transformed images to recover γ and the original image. The advantage of this synthetic experiment is that it is controlled, i.e., we know the exact answer to the registration problem and therefore we are able to test our algorithm under a perfectly controlled environment. We randomly choose γ and then run our code. Repeating the experiment 100 times we find that we are able to converge to the correct γ with accuracy of less than 0.25 pixels.

In the second example we use the images from Viola’s Ph.D thesis [18]. We compare the results of both MI and our NGF registration. The difference between the MI registration and the NGF registration was less than 0.25 of a pixel, thus we conclude that the methods give virtually identical minima. However,

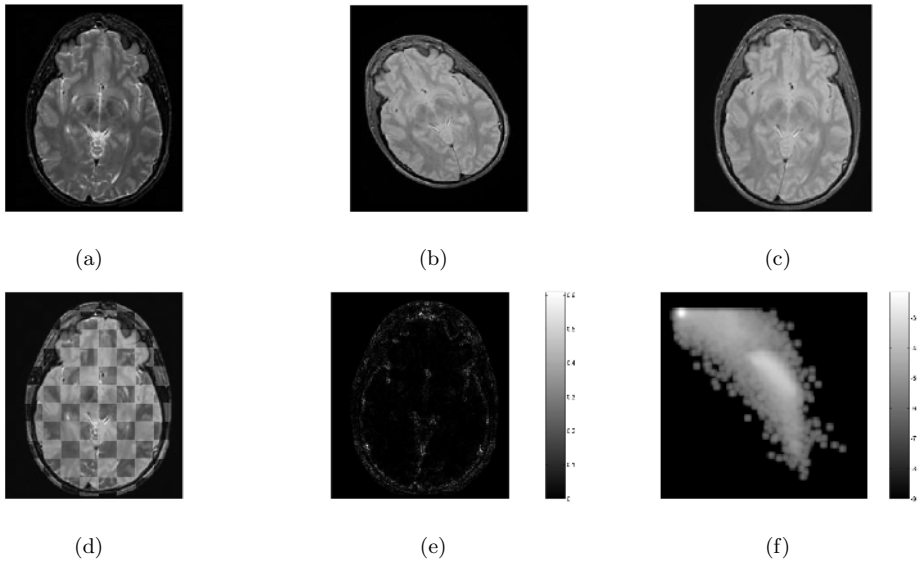


Fig. 2. Experiments with Viola’s example; (a) reference R , (b) template T , (c) registered T , (d) overlay of T and R (20^2 pixels checkerboard presentation), (e) cross product $\mathbf{n}_T \times \mathbf{n}_R$, (f) joint density at the minimum

to obtain the minima using MI we needed to use a random search technique to probe the space which is rather slow. The global MI minima has the value of about -9.250×10^{-2} while the guess $\gamma = 0$ has the value of about -9.115×10^{-2} . Thus the “landscape” of the MI function for this example is similar to the one plotted in Figure 1.

In comparison, our NGF algorithm used only 5 iteration on the finest grid. The registration was achieved in a matter of seconds and no special space probing was needed to obtain the minima. The value of the NGF function at $\gamma = 0$ was -4.63×10^1 while at the minima its value was -2.16×10^2 thus our minima is much deeper compared with the MI minima. The results of our experiments are also presented in Figure 2.

Another advantage of our method is the ability to quickly evaluate the registration result by looking at the absolute value of the cross-product $|\mathbf{n}_T \times \mathbf{n}_R|$. This product has the same dimension as the images and therefore can be viewed in the same scale. If the match is perfect then the cross product should vanish and therefore any deviation from zero implies a imperfect fit. Such deviations are expected to be present due to two different imaging processes, the noise in the images and possible additional nonlinear features. Figure 2e) shows the cross product for the image matched above. It is evident that the matching is very good besides a small number of locations where we believe the image to be noisy.

The previous two examples demonstrate the ability of our new algorithm to successfully perform parametric registration. Nevertheless, when the number of parameters is very large then stochastic optimization techniques stall and

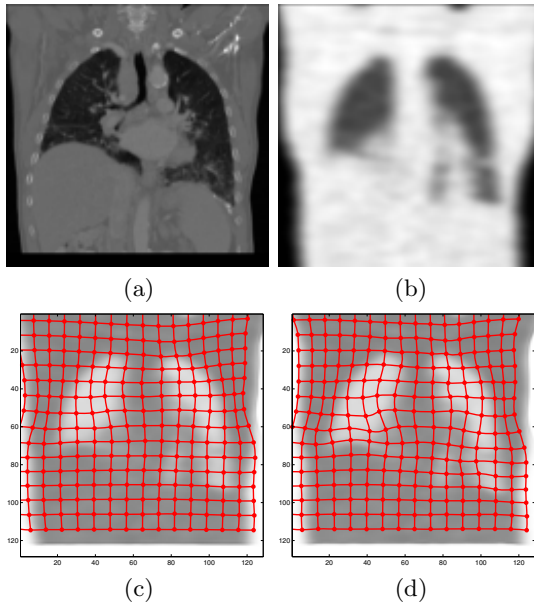


Fig. 3. Experiments with PET/CT images: (a) reference R , (b) template T , (c) MI registered T and deformed grid, (d) NGF registered T and deformed grid

differentiable smooth distance measure functions are much more important. We therefore test our algorithm on a PET/CT registration, see Figure 3. We perform an elastic registration of CT and a PET images (transmission) of a human thorax. For this example, the differences between the MI and NGF approach are very small; cf. Figure 3. For the NGF no special effort was needed in order to register the images. For MI we needed a good starting guess to converge to a local minima.

References

1. U. Ascher, E. Haber and H. Haung, *On effective methods for implicit piecewise smooth surface recovery*, *SISC* **28** (2006), no. 1, 339–358.
2. Lisa Gottesfeld Brown, *A survey of image registration techniques*, *ACM Computing Surveys* **24** (1992), no. 4, 325–376.
3. C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans, *BrainWeb MR simulator*, Available at <http://www.bic.mni.mcgill.ca/brainweb/>.
4. A. Collignon, A. Vandermeulen, P. Suetens, and G. Marchal, *multi-modality medical image registration based on information theory*, Kluwer Academic Publishers: *Computational Imaging and Vision* **3** (1995), 263–274.
5. J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, SIAM, Philadelphia, 1996.
6. M. Droske and M. Rumpf, *A variational approach to non-rigid morphological registration*, *SIAM Appl. Math.* **64** (2004), no. 2, 668–687.

7. L.A. Gallardo and M.A. Meju, *Characterization of heterogeneous near-surface materials by joint 2d inversion of dc resistivity and seismic data*, Geophys. Res. Lett. **30** (2003), no. 13, 1658–1664.
8. G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), 215–223.
9. E. Haber and D. Oldenburg, *Joint inversion a structural approach*, Inverse Problems **13** (1997), 63–67.
10. J. Modersitzki, *Numerical methods for image registration*, Oxford, 2004.
11. H. Park, P. H. Bland, K. K. Brock, and C. R. Meyer, *Adaptive registration using local information measures*, Medical Image Analysis **8** (2004), 465–473.
12. Josien PW Pluim, J. B. Antoine Maintz, and Max A. Viergever, *Image registration by maximization of combined mutual information and gradient information*, IEEE TMI **19** (2000), no. 8, 809–814.
13. J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, *Interpolation artefacts in mutual information based image registration*, Proceedings of the SPIE 2004, Medical Imaging, 1999 (K.M. Hanson, ed.), vol. 3661, SPIE, 1999, pp. 56–65.
14. ———, *Mutual-information-based registration of medical images: a survey*, IEEE Transactions on Medical Imaging **22**, 1999, 986–1004.
15. L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics : computational issues in nonlinear science, 1992, 259–268.
16. R. Silverman, *Density estimation for statistics and data analysis*, Chapman & Hall, 1992.
17. M. Unser and P. Thévenaz, *Stochastic sampling for computing the mutual information of two images*, Proceedings of the Fifth International Workshop on Sampling Theory and Applications (SampTA'03) (Strobl, Austria), May 26-30, 2003, pp. 102–109.
18. Paul A. Viola, *Alignment by maximization of mutual information*, Ph.D. thesis, Massachusetts Institute of Technology, 1995.
19. G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.
20. Terry S. Yoo, *Insight into images: Principles and practice for segmentation, registration, and image analysis*, AK Peters Ltd, 2004.
21. J. Zhang and F.D. Morgan, *Joint seismic and electrical tomography*, Proceedings of the EEGS Symposium on Applications of Geophysics to Engineering and Environmental Problems, 1996, pp. 391–396.