Beyond Mutual Information: A Simple and Robust Alternative

Eldad Haber[1] and Jan Modersitzki[2]

[1]Department of Mathematics, Emory University, Atlanta USA
[2]Institute of Mathematics, University of Lübeck, 23560 Lübeck, Germany
Email: haber@mathcs.emory.edu, modersitzki@math.uni-luebeck.de

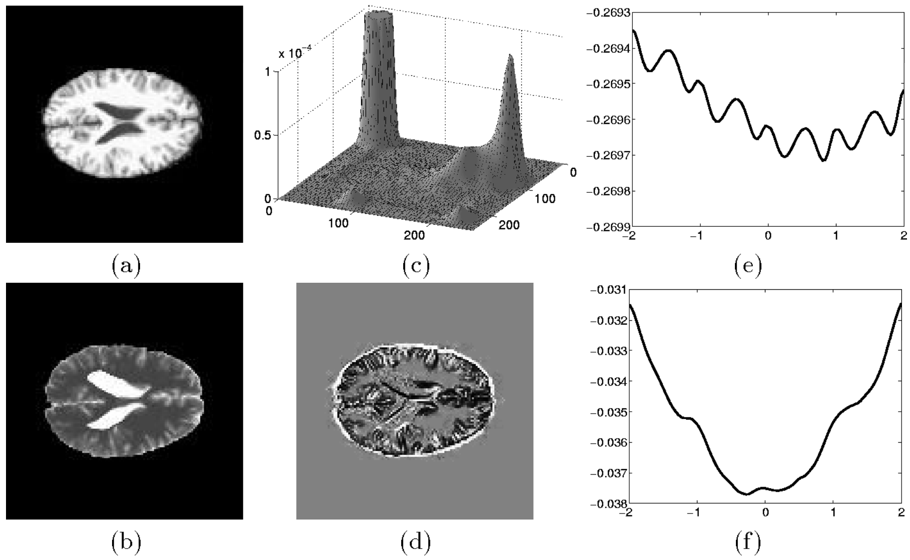
Abstract. One of the remaining challenges in image registration arises for multi-modal images taken from different imaging devices and/or modalities. Starting in 1995, mutual information has shown to be a very successful distance measure for multi-modal image registration. However, as it is well-known, mutual information also has a number of drawbacks. Here, we present an alternative image similarity measure which is also capable to handle multi-modal images but better suited for optimization. As we show, the alternative normalized gradient based approach is deterministic, much simpler, easier to interpret, fast and straightforward to implement, faster to compute, and also much more suitable to optimization.

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 open challenges in image registration arises for multi-modal images taken from different imaging devices and/or modalities; see Figure 1 for an example. 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 [1, 2]. Starting in 1995, mutual information has shown to be a successful distance measure for multi-modal image registration and is to be considered as the state-of-the-art technique.

However, mutual information has a number of well-known drawbacks; cf. e.g., [4, 5]. Firstly, mutual information is highly non-convex and has typically many local minima; see for example the discussion in [6, §6.6] and Section 2. Therefore, the non-convexity and hence non-linearity of the registration problem is enhanced by the usage of mutual information. Secondly, as it has its foundation in information theory, mutual information has a naturally discrete nature. However, fast and efficient registration schemes rely on powerful optimization techniques and thus on smooth functions. Thirdly, since mutual information is defined via the typically unaccessible joint density of the gray value

Fig. 1. Original BrainWeb [3] T1 (a) and T2 (b) weighted magnetic resonance images; (c) the joint density approximation for R and T, (e) mutual information versus shift, (d) the normalized gradient field for R and T, (f) normalized gradient field versus shift.



distribution, approximations of the density are required. These approximations typically involve some very sensitive smoothing parameters (e.g. a binning size or a Parzen window width). Fourthly, mutual information completely decouples the gray value from the location information. Therefore judging the output of the registration process is difficult. Finally, because of the previous difficulties, there is not a unique or even common implementation for mutual information and its derivatives. 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, [4]. As a result, a practical implementation of mutual information is highly non-trivial.

In this paper we investigate an alternative distance measure which is based on normalized gradients using properties of differential geometry. As we show, the alternative approach is deterministic, much simpler, easier to interpret, fast and straightforward to implement, faster to compute, and also much more suitable to optimization.

2 An illustrative example

To emphasize the difficulty explained above, we present an illustrative example. Figure 1 shows a reference (a) and a template (b) image in different modalities. 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 an approximation to the joint density which is based on a kernel estimator; see [7] for details. Note that the joint density is completely unrelated to the spatial image content. We now slide the template image along the horizontal axis. Figure 1e) shows the 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. In particular, the curve suggests that there are many pronounced local minima which are closed in value to the global minima.

Figure 1d) displays a typical visualization of our alternative 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.

3 A simple and robust alternative to mutual information

Given a reference image R and a template image T, the goal of image registration is to find a "*reasonable*" transformation such that the "*distance*" between the reference image and a deformed template image is small. As described in [6], there are basically two registration approaches. Since our interest is the discussion of distance measures, we focus on so-called parametric image registration which is easier to explain.

Given a distance measure $\mathcal{D}$ and some basis functions $\phi_1, \ldots, \phi_m$, the registration problem is to find a minimizer γ of

$$f(\gamma) := \mathcal{D}[R(\mathbf{x}), T(\phi(\gamma, \mathbf{x}))], \quad \text{where} \quad \phi(\gamma, \mathbf{x}) = \textstyle\sum_{k=1}^{m} \gamma_k \phi_k(\mathbf{x}). \tag{1}$$

Our alternative multi-modal distance measure is based on observation that two image are considered to be similar, if intensity changes occur at the same locations. Intensity change can be detected via the image gradient. However, since the magnitudes of changes might be related to the imaging devices, and are not related to image differences, it is not advisatory to directly base a distance measure on gradients. We thus consider a regularized normalized gradient field

$$\boldsymbol{n}_{\mathcal{E}}(I, \mathbf{x}) := \tfrac{\nabla I(\mathbf{x})}{||\nabla I(\mathbf{x})||_{\mathcal{E}}}, \tag{2}$$

where, for $\mathbf{x} \in \mathbb{R}^d$ we set $||\mathbf{x}||_{\mathcal{E}} = \sqrt{\sum_{\ell=1}^{d} x_\ell^2 + \mathcal{E}^2}$ and $\nabla I := (\partial_1 I, \ldots, \partial_d I)^\top$. In regions where $\mathcal{E}$ is much larger than the gradients the maps $\boldsymbol{n}_{\mathcal{E}}(I, \mathbf{x})$ are almost zero and therefore do not have a significant effect. However, in regions where $\mathcal{E}$ is much smaller than the gradients, the regularized maps are close to the non-regularized ones and these regions make a substantial difference in the calculation of the distance measures. The parameter $\mathcal{E}$ therefore answers the question, "what can be interpreted as a jump?", and can be computed by the following *automatic choice*:

$$\mathcal{E} = \tfrac{\eta}{V} \textstyle\int_{\Omega} |\nabla I(\mathbf{x})| \, d\mathbf{x}, \tag{3}$$

where η is the estimated noise level in the image and V is the volume of the domain Ω; see also [8].

For two related points $\mathbf{x}$ in R and $\phi(\mathbf{x})$ in T or, equivalently, $\mathbf{x}$ in $T \circ \phi$, we look at the vectors $\boldsymbol{n}(R, \mathbf{x})$ and $\boldsymbol{n}(T \circ \phi, \mathbf{x})$. 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.

This observation motivate the following distance measures, which are equivalent from an optimization point of view:

$$\mathcal{D}^c(T,R) = \tfrac{1}{2}\int_\Omega d^c(T,R)\; d\mathbf{x}, \qquad d^c(T,R) \;=\; \|\boldsymbol{n}(R,\mathbf{x}) \times \boldsymbol{n}(T,\mathbf{x})\|^2, \qquad (4)$$

$$\mathcal{D}^d(T,R) = -\frac{1}{2}\int_\Omega d^d(T,R)\; d\mathbf{x}, \quad d^d(T,R) \;=\; \langle \boldsymbol{n}(R,\mathbf{x}), \boldsymbol{n}(T,\mathbf{x})\rangle^2 . \qquad (5)$$

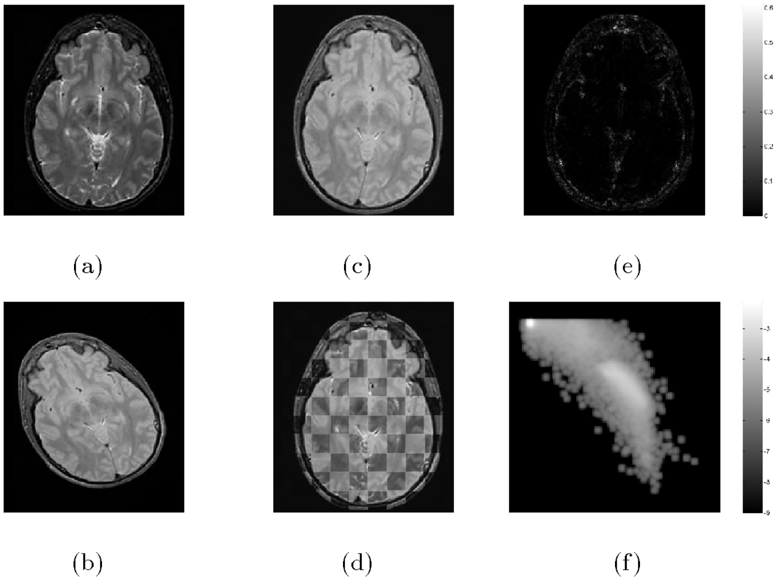
To find the image deformation we need to minimize $f(\gamma)$ (cf. (1)) for $\mathcal{D}^c$ or $\mathcal{D}^d$. Since this function is twice differentiable with respect to γ, we are able to use a Newton type method. Note that the distance measure has a least-squares from. Therefore a natural optimization algorithm is the Gauß-Newton method.

4 Numerical Experiments

In the second example we use the images from Viola's Ph.D thesis [2]. In the original work a few thousands of iterations of stochastic optimization algorithm where needed to achieve registration using MI as a distance measure. Here, we have used a more efficient implementation of mutual information [7] to obtain competitive results. We then compare the results of both registrations. The difference between the MI registration and the normalized gradient field (NGF) registration was less than 0.25 of a pixel, thus we conclude that the methods give virtually identical minima. However, to obtain the minima using MI we needed to use a random search technique to probe the space. This technique requires the estimation of many joint density distribution and therefore it is rather slow. When probing the space we have found many local minima. Furthermore, the local minima and the global minima tend to have roughly the same magnitude. The global 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 15 iteration on the coarse grid which is 22×24 and 5 iterations on each finer 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.

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.



References

1. Collignon A, Vandermeulen A, Suetens P, Marchal G. 3D Multi-Modality Medical Image Registration Based on Information Theory. Kluwer Academic Publishers: Computational Imaging and Vision 1995;3:263–274.
2. Viola PA. Alignment by Maximization of Mutual Information. Ph.D. thesis. Massachusetts Institute of Technology; 1995.
3. Cocosco CA, Kollokian V, Kwan RKS, Evans AC. BrainWeb MR simulator; 2004. Available at http://www.bic.mni.mcgill.ca/brainweb/.
4. Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging 1999;22:986–1004.
5. Pluim JPW, Maintz JBA, Viergever MA. Interpolation artefacts in mutual information based image registration. In: Hanson KM, editor. Proceedings of the SPIE 2004, Medical Imaging, 1999. vol. 3661. SPIE; 1999. p. 56–65.
6. Modersitzki J. Numerical Methods for Image Registration. Oxford University Press; 2004.
7. Haber E, Modersitzki J, Heldmann S. Computational Methods for Mutual Information Based Registration. Department of Mathematics and Computer Science. Emory University, Atlanta GA 30322; 2004. Submitted to Inverse Problems.
8. Ascher U, Haber E, Haung H. On effective methods for implicit piecewise smooth surface recovery. Dept of Math. & CS.. Emory University, Atlanta GA 30322; 2004.