

FAST GLOBAL IMAGE REGISTRATION USING RANDOM PROJECTIONS

Dennis M. Healy, Jr.

University of Maryland
Dept. of Mathematics
College Park, MD. USA

Gustavo K. Rohde

Carnegie Mellon University
Dept. of Biomedical Engineering
Pittsburgh, PA. USA

ABSTRACT

We describe a new method for efficiently computing the global optimum of least squares registration problems based on the recently developed theory of signal processing using random projections. The method is especially attractive for large scale registration problems where the goal is to register many images to a standard template. We test our new algorithm using real images of cells' nuclei and show our method can outperform more traditional dimension reduction methods such as projections onto lower dimensional B-spline function spaces.

Index Terms— Image registration, global optimization, random projection, B-splines.

1. INTRODUCTION

Image registration methods are routinely employed in imaging problems as a means through which to extract quantitative spatial information from two or more images. Amongst the many algorithms developed for image registration so far, methods based on image intensity values are especially attractive as they are easy to automate as solutions to optimization problems. The sum of squared differences, amongst other quantities, is typically used as the objective function in such optimization problems. The parameterization of the spatial transformation, as well as the size of the images, typically dictate the complexity of the problem. Pure translations, for example, can be computed efficiently, and globally, as the maxima of the cross correlation function between two images. Other parameters such as rotations, combined with scaling, shears, give rise to nonlinear functions which must be solved using iterative nonlinear optimization methods.

One of the biggest challenges in nonlinear, non convex, optimization problems has to do with guaranteeing convergence to a global solution of the problem. To date, no well established framework for finding a global solution of a registration problem exists. Multiple image resolutions can be used to solve the problem in an iterative fashion, from a coarse to fine scale, in the hope that local optima due to fine image detail can be avoided. Simplex-type methods are often used in the hope that their independence from local characteristics (local derivatives of the objective function with respect

to spatial transformation parameters) will render the method less likely to be confused by a local optima. Yet another approach is to solve the registration problem by initializing the optimization at several, perhaps randomly chosen locations, with the aim of avoiding at least some of the local optima that may be present. Finally, exhaustive searchers can also be used in instances where the parameter space is not too large. Of all these methods, only exhaustive searches can guarantee that the solution achieved is the global one.

Based on the work of Baraniuk and Wakin [1, 2] on random projections of signal manifolds we develop a novel image registration algorithm. The algorithm not only provides a fast way to compute the solution to the registration problem but can also guarantee with high probability that, in the sense of least squares, the solution encountered is a global solution. We begin this paper by reviewing the recently developed theory of random projections of image manifolds and then derive an algorithm that approximates the solution of the problem using a sampled version of a low dimensional embedding of the image manifold. We show the effectiveness of the algorithm using real images of cell nuclei and compare it to a similar algorithm based on orthogonal projections onto B-spline function spaces.

2. THEORY

Let $s(\mathbf{i})$, $\mathbf{i} \in \Omega = [0, 1, \dots, Q - 1]^d$ represent pixel values associated with a d dimensional real valued digital image with $N = Q^d$ pixels. We denote $\mathbf{s} \in \mathbb{R}^N$ to be the digital image organized in an N dimensional column vector. A continuous and differentiable representation for the image can be obtained by

$$\tilde{s}(\mathbf{x}) = \sum_{j \in \Gamma} b_j \phi_j(\mathbf{x}), \quad (1)$$

where $\phi_j(\mathbf{x})$ are C^r d dimensional basis functions chosen based on their interpolation or approximation properties and b_j , $j \in \Gamma \subset \mathbb{Z}$, are the coefficients of the linear expansion. A warped version of the digital image can be constructed by sampling (1) at coordinates $f_{\mathbf{p}}(\mathbf{i})$, $\mathbf{i} \in \Omega$, with \mathbf{p} some parameterization for the spatial transformation $f_{\mathbf{p}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$. We denote $\tilde{\mathbf{s}}_{\mathbf{p}}$ to represent the collection of image values $\tilde{s}(f_{\mathbf{p}}(\mathbf{i}))$,

$\mathbf{i} \in \Omega$ organized in a column vector. The spatial transformation $f_{\mathbf{p}}$ that aligns two images \mathbf{t} and \mathbf{s} can be computed as the solution of the following optimization problem:

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \|\mathbf{t} - \tilde{\mathbf{s}}_{\mathbf{p}}\| \quad (2)$$

where $\|\cdot\|$ is the standard discrete 2 norm for a finite dimensional vector.

When the problem is nonlinear, non-convex, performing a global search is usually computationally prohibitive. Let \mathbf{p} belong to some (discrete) set $\Theta \subset \mathbb{R}^K$, $K < N$. Suppose that Θ is organized in a K dimensional grid, with each dimension having G discrete coordinates. A global search for the optimum parameters \mathbf{p}^* over the samples in Θ would then involve a nearest neighbor search over $\#\Theta = G^K$ N -dimensional vectors. One alternative to decrease the dimensionality of the image space (N) is to project $\mathbf{t} \in \mathbb{R}^N$ and $\mathbf{s} \in \mathbb{R}^N$ onto a lower dimensional space (say \mathbb{R}^M) via low resolution approximations of each image. This does not decrease the number of vectors ($\#\Theta$) over which one has to search, but does improve things in the sense that evaluations of the objective function involve inner products in a $M < N$ dimensional space. Such low resolution approximations can be computed as orthogonal projections onto low dimensional B-spline function spaces [3], for example.

In this work we introduce a new way of reducing the dimensionality of the space in which the cost function needs to be evaluated by using random projections instead. The motivation for our approach is as follows. Even though each digital image $\tilde{\mathbf{s}}_{\mathbf{p}}$ is a point in \mathbb{R}^N , the set of images $\tilde{\mathbf{s}}_{\mathbf{p}}$, for $\mathbf{p} \in \Theta$, does not fill the entire space. That is, for a fixed set of coefficients b_j , the model in (1), under spatial transformations $f_{\mathbf{p}}$, $\mathbf{p} \in \Theta$ is usually not capable of generating any arbitrary image. Thus each image $\tilde{\mathbf{s}}_{\mathbf{p}}$ must lie on a K dimensional manifold \mathcal{M} embedded in \mathbb{R}^N . Thus, in a sense, the information pertaining to the spatial organization of the image $\tilde{\mathbf{s}}$ under spatial transformation $f_{\mathbf{p}}$ lies in a space of significantly lower dimension than N . Therefore image registration could be achieved, in principle, by evaluating the objective function in a lower dimensional space so long as we find a lower (lower than N) dimensional embedding (projection) of \mathcal{M} which retains some properties in terms of distances between different points in \mathcal{M} .

2.1. Random projections of image manifolds

Naturally, the first requirement is that our embedding procedure be a one to one mapping. If not, one would run the risk of having two warped images $\tilde{\mathbf{s}}_{\mathbf{p}_1}$ and $\tilde{\mathbf{s}}_{\mathbf{p}_2}$ occupy the same point in \mathbb{R}^M . Whitney's easy embedding theorem provides a guideline for the number of dimensions required for the embedding of the K dimensional manifold \mathcal{M} :

Theorem 2.1 [4] *Let \mathcal{M} be a compact Hausdorff C^r K -dimensional manifold, with $2 \leq r \leq \infty$. Then there is a C^r embedding of \mathcal{M} in \mathbb{R}^{2K+1} .*

In addition, Baraniuk and Wakin [1, 2] observed that, assuming mild conditions on \mathcal{M} , a randomly chosen projection from \mathbb{R}^N to \mathbb{R}^{2K+1} , when restricted to \mathcal{M} , will also be invertible with high probability. We denote a random projection of an N dimensional vector image $\tilde{\mathbf{s}}_{\mathbf{p}}$ on \mathbb{R}^M as $\mathcal{P}\tilde{\mathbf{s}}_{\mathbf{p}}$.

In cases where a significant amount of noise is present, or when dealing with a set of images which cannot be described exactly by spatial transformations of one single image, additional requirements may be necessary for the minimization over the reduced space (\mathbb{R}^M) equate to minimization over the original image space (\mathbb{R}^N). This amounts to requiring that distances between specific points on and near the manifolds be the roughly equivalent in both spaces. In the discrete setting described above, the Johnson-Lindenstrauss lemma (JL) [1] provides a useful guideline for M :

Lemma 2.2 [Johnson-Lindenstrauss] *Let Ψ be a finite collection of points in \mathbb{R}^N . Fix $0 < \epsilon < 1$ and $\beta > 0$. Let \mathcal{P} be a random orthoprojector from \mathbb{R}^N to \mathbb{R}^M with*

$$M \geq \left(\frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \right) \ln(\#\Psi)$$

If $M \leq N$, then, with probability exceeding $1 - (\#\Psi)^{-\beta}$, the following statement holds: For every $\mathbf{s}, \mathbf{t} \in \Psi$,

$$(1 - \epsilon) \sqrt{\frac{M}{N}} \leq \frac{\|\mathcal{P}\mathbf{s} - \mathcal{P}\mathbf{t}\|}{\|\mathbf{s} - \mathbf{t}\|} \leq (1 + \epsilon) \sqrt{\frac{M}{N}}$$

For the registration problem discussed above, $\#\Psi = \#\Theta = G^K$, while \mathcal{P} can be an $M \times N$ matrix whose entries are independent identically distributed instances of a random variable with distribution $\mathcal{N}(0, 1)$ (and whose rows are orthogonal). Other randomly constructed matrices (e.g. Bernoulli), some of which are sparse, can also be used [1]. The requirements of the JL lemma can be refined by realizing that in our problem the points \mathbf{s} and \mathbf{t} lie on a low dimensional smooth manifold in \mathbb{R}^N [1].

The notion of random projections of manifolds generated by applying spatial transformations to an image is illustrated in Figure 1. The top row shows the image (Gaussian 'blob') used in the experiment. The image was translated along a two dimensional sinusoidal path shown as the black curve on the same image. Since only one parameter is varied to translate the image along the path, the manifold generated by the translated versions of the image is one dimensional ($K = 1$). The bottom left panel shows a random projection of the manifold using only two coefficients. As shown in this panel, the projection is not one to one. The bottom right panel shows a projection where $M = 2K + 1 = 3$: the embedding is non self intersecting.

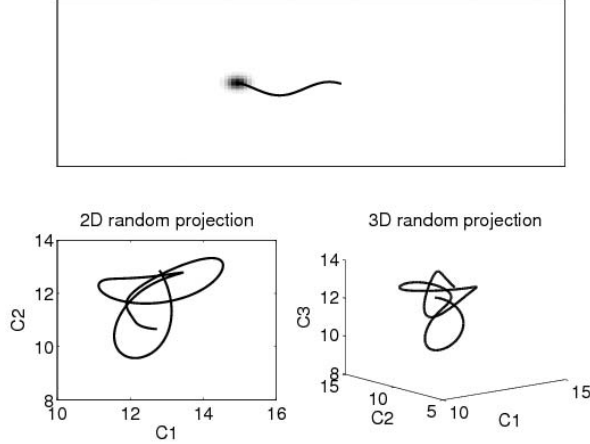


Fig. 1. Random projection of manifold generated by translating a 2D image along a sinusoidal path. The bottom panels show the random projection of the manifold in 2 (left) and 3 (right) dimensions.

3. ALGORITHM

Consider the problem of registering a series of images s^1, s^2, \dots to a reference image t . This operation is commonly performed in the analysis of large quantities of images, where a common reference frame is beneficial. Examples include large scale computational anatomy, registration of large time series of images, and others. Note that each image in the series can be made continuous through equation (1). Let Λ represent a set of M dimensional points generated by randomly projecting each $\tilde{t}_{\mathbf{p}}, \mathbf{p} \in \Theta$ with a fixed projection matrix \mathcal{P} . Note that $\#\Lambda = \#\Theta = G^K$. We refer to this the computation of the set of points Λ as the *training phase*. Note that this is a computationally expensive procedure, but since it only has to be performed once an exhaustive sampling of the spatial transformation parameters to a desired level of accuracy can be used.

In the next phase of the algorithm each image s^j is registered to t by first searching for the nearest neighbor of $\mathcal{P}s^j$ in Λ . This is an exhaustive search but is occurs over points in \mathbb{R}^M rather than \mathbb{R}^N . Once the nearest neighbor is found, the set of parameters \mathbf{p} which generated the nearest neighbor to $\mathcal{P}s^j$ becomes known. A registered pair of images can be computed by $s(\mathbf{i}), \tilde{t}(f_{\mathbf{p}}(\mathbf{i}))$ or, alternatively, by $\tilde{s}(f_{\mathbf{p}}^{-1}(\mathbf{i})), t(\mathbf{i})$, where f^{-1} represents the function inverse.

3.1. Extension to multiscale manifold representations

When compared to global exhaustive searching, the random projection-based registration algorithm above reduces the computational complexity (including memory requirements) from $O(G^K N)$ to $O(G^K M)$, with $M \sim O(\ln(G^K))$. However the number of points G^K can be large and cumbersome to

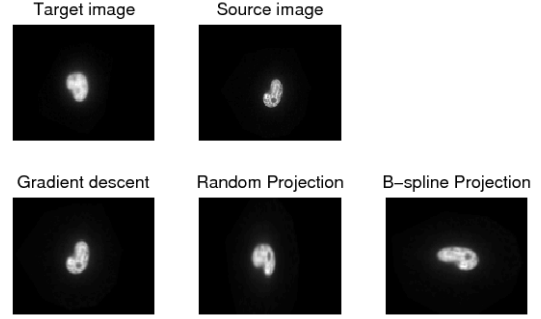


Fig. 2. Comparison of registration results obtained using a gradient descent optimization, versus gradient descent initialized using nearest neighbors of B-spline and random projections.

work with (e.g. may still be too large to store in memory). An additional improvement could be obtained via multiscale representations not of the image data, as is routinely done [5], but of the manifold generated by image t and its spatial transformations [6, 7]. An approximate registration could be obtained by the algorithm above using a coarse representation of the manifold data $\mathcal{P}\tilde{t}_{\mathbf{p}}, \mathbf{p} \in \Theta$ (thus reducing G drastically). Once an approximate solution to $\mathbf{p}_{app.}$ is obtained, a more accurate one can be computed by repeating the procedure replacing $\tilde{s}(\mathbf{i})$ with $\tilde{s}(f_{\mathbf{p}_{app.}}^{-1}(\mathbf{i}))$ and using a finer, but more localized around the identity transformation, sampling of the registration parameter space Θ . Alternatively, the refinement can also be computed using a standard gradient descent minimization method.

4. RESULTS

We test the nearest neighbor algorithm proposed above to register a set of 87 real microscopy images of HeLa cell nuclei [8]. We compare random projections to orthogonal projections on cubic B-spline function spaces [3]. In both cases the projection-based nearest neighbor algorithm was used to compute an estimate of an affine two dimensional transformations without shear (two translations, one rotation, two scalings): that is $K = 5$. Both the random and B-spline function space projections were of dimension $M = 20$ while $\#\Theta = 8.1 \times 10^6$. Multiscale manifold representations were not used. The estimates of the registration using random and B-spline projections were refined using a standard gradient descent approach. We also compare the result of the projection based registration methods with a registration based solely on the gradient descent minimization of (2) without global searching.

Figure 2 displays results of the registration obtained using different methods. As can be seen in this example, a simple gradient descent fails to register the source and target images

satisfactorily since it becomes ‘stuck’ in a local optima. Visual inspection of the result obtained with the gradient descent initialized based on the results obtained using nearest neighbors of random projections reveals this method produced a much superior match for this image. Finally, the result obtained by using a nearest neighbor search of projections on cubic B-spline function spaces produces an erroneous result suggesting that such drastic dimension reductions computed using the B-spline projection framework do not result in isometric manifold representations.

In Figure 3 we plot the joint distribution of least squares error between the target image and each image registered using the standard gradient descent optimization method and the random projection assisted one. The nearest neighbor algorithm based on orthogonal projections onto cubic B-splines function spaces produced results of very poor quality. These are omitted for brevity. In Figure 3, the identity line is also plotted in order to make it more clear that, more often than not, the error produced by the random projection registration method is lower than the error produced by the standard method. However, in some cases the error produced by the standard method is lower. One possible explanation is that due to the influence of noise and finite step sizes in the numerical optimization, the gradient descent algorithm is not arbitrarily accurate (in our implementation translations are calculated to 0.1 pixels for example). This variability in accuracy of the registration parameters will result in variability on the final objective function value. Indeed many values in Figure 3 seem to be near the identity line, meaning that in those instances registration with both methods produced roughly equivalent results.

As for the few instances where the random projection algorithm produced a result significantly worse than the result obtained, two factors could be to blame. These are: not enough resolution in the parameter space ($\#\Theta$) so that a local optimum is found instead of a global optimum, and too few random projections. The first problem can be addressed by increasing $\#\Theta$ or by using multi-resolution manifold representations as mentioned earlier. The second issue can be addressed by increasing M so that, as much as possible, the projection operation results in an (approximate) isometry.

5. SUMMARY AND CONCLUSIONS

We have proposed an image registration algorithm based on random projections of manifolds generated by spatial transformations of a fixed image. The method is especially suited for situations where the goal is to register a series of images to a standard template. For low dimensional manifolds the method can significantly reduce the computational complexity of global searches, allowing for robust registration of a series of images.

We have demonstrated the method can improve the quality of the registration of a series of 2D images of nuclei of

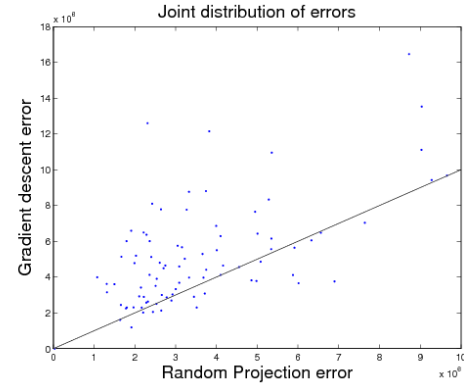


Fig. 3. Joint distribution of registration error between a standard gradient descent method and a random projection assisted gradient descent method.

cells as compared to a standard gradient descent implementations. In comparison, a similar algorithm devised based on orthogonal projections onto cubic B-spline function spaces performed very poorly and suggests such drastic dimension reduction with standard basis functions (B-splines, and others) is not appropriate for the purpose of registering two images.

6. REFERENCES

- [1] R.G. Baraniuk and M.B. Wakin, “Random projections of smooth manifolds,” *Preprint*, 2006.
- [2] M.B. Wakin and R.G. Baraniuk, “Random projections of signal manifolds,” in *Proc. of ICASSP 06*. IEEE, 2006, pp. 941–944.
- [3] M. Unser et al., “Enlargement or reduction of digital images with minimum loss of information,” *IEEE Transactions on Image Processing*, vol. 4, pp. 247–258, 1995.
- [4] M. W. Hirsch, *Differential Topology*, vol. 33 of *Graduate Texts in Mathematics*, Springer, 1976.
- [5] P. Thévenaz et al., “A pyramid approach to subpixel registration based on intensity,” *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, January 1998.
- [6] I.U. Rahman et al., “Multiscale representations for manifold-valued data,” *Multiscale Modeling and Simulation*, vol. 4, pp. 1201–1232, 2005.
- [7] R.R. Coifman et al, “Diffusion maps,” *Applied and computational Harmonic Analysis*, vol. 21, pp. 5–30, 2006.
- [8] M. V. Boland et al., “A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells,” *Bioinformatics*, vol. 17, pp. 1213–1223, 2001.