GRADIENT BASED OPTIMIZATION OF AN EMST IMAGE REGISTRATION FUNCTION

Mert R. Sabuncu, Peter J. Ramadge

Princeton University
Department of Electrical Engineering
Princeton, NJ 08544

ABSTRACT

This paper examines the problem of registering images using an information theoretic metric (e.g., entropy) estimated using a Euclidean Minimum Spanning Tree (EMST). The objective is to find an extremum of the metric with respect to a vector of free parameters. One of the major difficulties posed by such graph theoretic metrics is concurrently obtaining gradient information as the metric is computed. Obtaining the gradient is a first step in efficiently optimizing the metric. Our main contribution is to show how to obtain a gradient-based descent direction from the computation of the EMST metric. We also indicate how this can be used for optimizing image registration over a vector set of parameters and provide some preliminary experimental results.

1. INTRODUCTION

Registration or alignment of images is a fundamental problem in image processing. The ultimate goal is usually to compare or fuse the information contained in images captured at different times, by different sensors (multi-modal) or from different viewpoints. Typical applications include medical/biological image analysis, computer vision and military applications.

Recently, information theoretic registration metrics, e.g. mutual information of pixel intensities, have been proposed and extensively studied [10], [9]. These metrics are well-suited to multi-modal signal registration problems. Most commonly, the algorithms rely on histograms to compute the information theoretic metric. An alternative approach that has produced promising results is based on entropic spanning graphs [4]. However, for this approach it has not been clear how to concurrently obtain derivative information. Hence no efficient vector optimization method has been reported based on this type of entropy metric. The current paper focuses on this problem for the special case of Euclidean Minimum Spanning Trees (EMST) and indicates how to select a gradient-based descent direction for the EMST metric. This leads to a more efficient vector optimization algorithm. Although we restrict attention to the EMST metric, the basic principles of our approach should extend to other entropic graph functions.

The remainder of the paper is organized as follows. Section 2 provides some background on graph theoretic entropy estimators and Section 3 defines the EMST registration metric. Section 4 formulates the underlying optimization problem and subsection 4.1 proposes a gradient-based solution. Details of the proposed method applied to image registration are provided in Section 5 and some preliminary experimental results are presented in subsection 5.1.

2. ENTROPIC SPANNING GRAPHS

It is possible to estimate entropic measures of a probability density based on computing specific graphs (e.g. a minimum spanning tree) on independent samples from the density. In [11], Yukich and Redmond provide a general framework to obtain convergence results of some Euclidean length functionals for specific graphs. Based on this framework, Hero et al. [3] present the following result to estimate the Renyi entropy H_α of an underlying p.d.f: Let $Z_n = \{z_1, \dots, z_n\}$ be n iid samples drawn from a Lebesgue density f, $G(E, Z_n)$ be a graph with edges weighted by the Euclidean distance between their endpoints and $\mathcal{G}(Z_n)$ denote a family of graphs with a specific topological constraint (e.g. spanning trees, k-neighbor graphs, etc.). Define the graph weight function $W_\gamma(G) = \sum_{e \in E} |e|^\gamma$ and the minimum graph weight (MGW) function

$$W_\gamma(Z_n) = \min_{G \in \mathcal{G}(Z_n)} W_\gamma(G),$$

Then:

$$\lim_{n \to \infty} \log(\frac{W_\gamma(Z_n)}{n^\alpha}) = H_\alpha(f) + c \text{ a.s.}, \tag{1}$$

where c is a constant that depends only on $\mathcal{G}(Z_n)$, i.e., the topological constraint. Thus the MGW function can be used as an asymptotically unbiased and strongly consistent estimator of the α-Renyi entropy. When f is a smooth function, this method has a faster convergence rate compared

Research partially supported by **S**iemens **C**orporate **R**esearch, Princeton N.J.

to the popular histogram based "plug-in" estimation techniques [2] as $n \to \infty$ [3]. In the remainder of the paper, we restrict attention to graphs $\mathcal{G}$ that are spanning trees. In this case, the MGW function is equal to the EMST weight.

3. THE EMST REGISTRATION METRIC

Inspired by information theoretic registration techniques [9], and result (1), entropic graphs have recently been usefully employed for multi-modal registration [4],[8],[5]. The basic framework of this approach is as follows. For images, $I_i(x,y)$, $i = 1,2$, and d a positive even integer, define $f_i(x,y) = f(I_i(x',y') : (x',y') \in N(x,y)) \in \mathbb{R}^{d/2}$ where $N(x,y) \subset \mathbb{R}^2$ is called the "neighborhood" of (x,y) and f is the "feature" function. A trivial but useful example is $f_i = I_i(x,y)$ where f is the identity function and $N(x,y) = (x,y)$. In this case pixel intensity values are the "features". A more complex example that incorporates image gradient information is to select the feature function to be $f_i = (I_i(x,y), \nabla_x I_i(x,y), \nabla_y I_i(x,y)) \in \mathbb{R}^3$, where ∇_x (resp. ∇_y) denotes the partial difference with respect to x (resp. y). Information theoretic registration methods are based on treating samples of the joint signal $(f_1, f_2) \in \mathbb{R}^d$ as realizations of a random variable. Intuition suggests that samples from f_1 and f_2 should become "dependent" as the two images are correctly aligned. This might be measured, for example, by mutual information or joint entropy or, based on (1), by an entropic spanning graph [5], [8]. In particular, for a fixed feature function f and constant α, we define the EMST registration function to be the EMST weight $W_\gamma(\mathcal{F}_n)$, where $\mathcal{F}_n$ is a set of n samples of (f_1, f_2). To date, the major disadvantage of using the EMST metric is that it was not clear how gradient information could be obtained concurrently from the computation. We address this issue below.

4. OPTIMIZING THE EMST METRIC

Image registration problems typically consist of three major components. The *transformation space* determines the allowed spatial transformation applied to the images. This component is highly application dependent. Examples are rigid-body, affine and deformable transformations. The *registration function* quantifies the similarity between two images under a given transformation. Some examples are mutual information, correlation ratio, the EMST function, etc. The *optimization method* searches for the optimum transformation that maximizes the similarity between the images.

Consider the problem of aligning images $I_i(x,y)$, $i = 1,2$, using the EMST metric with fixed sampling locations $\Omega \subset \mathbb{R}^2$ and feature function $f(.)$. Let the spatial transformation $T : \Omega \mapsto \mathbb{R}^2$ be parameterized by m parameters, i.e. $T_{\mathbf{t}}$ where $\mathbf{t} = (t_1, \ldots, t_m)$. For example, a rigid-body transformation $T^R_{\mathbf{t}}$ has three parameters: two shifts (t_x, t_y) and a rotation θ and can be expressed as:

$$T^R_{\mathbf{t}}(x,y) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}. \quad (2)$$

Let $W(\mathbf{t})$ denote the EMST weight of

$$\mathcal{F}(\mathbf{t}) = \{(f_1(x,y), f_2(T_{\mathbf{t}}(x,y))) \colon \forall (x,y) \in \Omega\}.$$

Note that in the rest of the paper we will assume $\gamma = 1$. The choice of a different value for γ may change the registration result, but the methods discussed below extend in a straightforward way to these cases. In this framework, the registration problem boils down to the following optimization problem:

$$\mathbf{t_{opt}} = \arg\min_{\mathbf{t}} W(\mathbf{t}).$$

Typically this problem is non-convex. Thus finding the global optimum is a difficult task. For illustration, Figure 1 shows the profile of an EMST registration function with respect to the translation along the x-axis for the image pair in Figure 2.

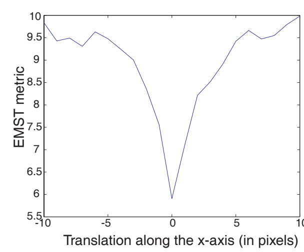

Fig. 1. Profile of a EMST registration function

4.1. Obtaining a Descent Direction

Descent optimization methods use a computed search direction that guarantees a local decrease in the value of the optimization function. Using the gradient or an estimate of the gradient is a common way to determine this direction. These methods usually assure convergence to a local extremum of the optimization function. To obtain a global extremum, multiple-start and multi-resolution approaches, [7], can be used to augment the basic method in an attempt to avoid getting stuck in an undesired local extremum.

The difficulty, however, is that, in general, the EMST metric is not differentiable. For example, consider the vertex set $\mathcal{V} = \{v_1, v_2, v_3\}$ with edges and parameterized weights: $|e_{12}| = 1 - 2t$, $|e_{23}| = t + 1$ and $|e_{13}| = -2 + 3t$. It's easy to show that at $t = 0^-$, the MST consists of e_{23} and e_{13}, whereas at $t = 0^+$, e_{12} and e_{13} belong to the MST. Thus $dW(0^-)/dt = 4$ and $dW(0^+)/dt = 1$. Since the left and right derivatives are not equal, the derivative of the EMST weight does not exist at $t = 0$.

Our main result, Theorem 1, resolves the above issue by providing a means to quickly determine a descent direction for the EMST metric. Let $E_{\mathbf{t_0}}$ denote the edges that belong to an EMST of $\mathcal{F}(\mathbf{t_0})$. Let $\mathbf{u}$ be an m-dimensional unit vector, $\nabla_{\mathbf{u}}$ denote the directional derivative and ∇ denote the m-dimensional gradient vector. Let $\|.\|$ designate the usual L_2-norm and $\|e(\mathbf{t_0})\|$ be the Euclidean length of an edge with endpoints in $\mathcal{F}(\mathbf{t_0})$.

Theorem 1 *If*

$$\sum_{e \in E_{\mathbf{t_0}}} \nabla_{\mathbf{u}} \|e(\mathbf{t_0})\| < 0, \tag{3}$$

then $\exists \epsilon > 0$ *such that*

$$W(\mathbf{t_0} + h\mathbf{u}) \leq W(\mathbf{t_0})$$

for all $0 \leq h \leq \epsilon$.

Proof: If $\sum_{e \in E_{\mathbf{t_0}}} \nabla_{\mathbf{u}} \|e(\mathbf{t_0})\| < 0$, by vector calculus $\exists \epsilon > 0$ such that:

$$\sum_{e \in E_{\mathbf{t_0}}} \|e(\mathbf{t_0} + h\mathbf{u})\| \leq \sum_{e \in E_{\mathbf{t_0}}} \|e(\mathbf{t_0})\| = W(\mathbf{t_0}), \tag{4}$$

for all $0 \leq h \leq \epsilon$. Since $W(\mathbf{t_0} + h\mathbf{u})$ is the weight of the *minimum* spanning tree at $\mathbf{t_0} + h\mathbf{u}$:

$$W(\mathbf{t_0} + h\mathbf{u}) \leq \sum_{e \in E_{\mathbf{t_0}}} \|e(\mathbf{t_0} + h\mathbf{u})\|. \tag{5}$$

Hence combining 4 and 5, we get $W(\mathbf{t_0} + h\mathbf{u}) \leq W(\mathbf{t_0})$.$\square$

It can be shown that, when nonzero,

$$\mathbf{u}_d = - \sum_{e \in E_{\mathbf{t_0}}} \nabla \|e(\mathbf{t_0})\|, \tag{6}$$

satisfies the condition in Equation 3 and therefore is a descent direction. Hence, computation of a descent direction simply involves the additional computation of the gradient of the edge weights. This requires a negligible amount of additional computation compared to the computation of the actual EMST weight and it can be done concurrently with that computation.

5. APPLICATION IN IMAGE REGISTRATION

As a concrete application of the above ideas, consider the problem of rigidly registering two images I_1 and I_2. To provide a fair comparison to the popular mutual information based algorithm we use pixel intensity values as features. For a given set of transformation parameters $\mathbf{t_0} = (t_{x0}, t_{y0}, \theta_0)$ and fixed set of sampling locations $\Omega \subset \mathbb{R}^2$ the registration metric is equal to the total weight of the EMST of $\{I(x,y) = (I_1(x,y), I_2^{\mathbf{t_0}}(x,y)) : \forall (x,y) \in \Omega\}$ where $I_2^{\mathbf{t_0}}(x,y) = I_2(T_{\mathbf{t_0}}^R(x,y))$. Let $e = (I(x_1,y_1), I(x_2,y_2))$ be an edge in the computed EMST. Define

$$\begin{aligned} \nabla\|e\| &= \left(\partial\|e\|/\partial t_x; \quad \partial\|e\|/\partial t_y; \quad \partial\|e\|/\partial\theta \right) \text{ and} \\ \nabla I_2^{\mathbf{t_0}}(x,y) &= (\partial I_2^{\mathbf{t_0}}(x,y)/\partial x; \; \partial I_2^{\mathbf{t_0}}(x,y)/\partial y). \end{aligned}$$

Then it's easy to show that:

$$\begin{aligned} \nabla\|e(\mathbf{t_0})\| &= 1/\|e(\mathbf{t_0})\| * (I_2^{\mathbf{t_0}}(x_1,y_1) - I_2^{\mathbf{t_0}}(x_2,y_2)) \\ &\quad *(\nabla I_2^{\mathbf{t_0}}(x_1,y_1) - \nabla I_2^{\mathbf{t_0}}(x_2,y_2))^T * R, \end{aligned}$$

where $R = \begin{pmatrix} 1 & 0 & y_1 - y_2 \\ 0 & 1 & x_2 - x_1 \end{pmatrix}$. Plugging this in Equation 6 gives us a descent direction for each iteration of the formulated optimization problem.

We have implemented an iterative optimization scheme for the EMST image registration metric based on the above. In our implementation we employ a "pyramid approach" to avoid "getting trapped" in local minima. Thus the algorithm starts with registering coarse (low resolution) representations of the images. We use Gaussian blurring and uniform subsampling to obtain these low resolution images. The alignment results obtained from the coarse level are then used to initialize the registration at the next level. The algorithm terminates once full resolution is reached.

For a given set of transformation parameters $\mathbf{t_0}$ we can compute a descent direction $\mathbf{u}_d$ by considering only the edges that belong to the EMST $E(\mathbf{t_0})$. For any other $\mathbf{t}$, the EMST metric by definition satisfies $W(\mathbf{t}) \leq \sum_{e \in E(\mathbf{t_0})} \|e(\mathbf{t})\|$ and moving in a direction that decreases the total weight of $E(\mathbf{t_0})$ is equivalent to decreasing an upper bound on the EMST function. Hence, in our approach we propose to "update" the EMST every k steps instead of recomputing it at every iteration. This accelerates the registration algorithm by reducing the total number of times we run the MST algorithm. To compute the MST we use Kruskal's algorithm [6] preceded by a Delaunay triangulation [1]. The time complexity of this implementation using pixel intensities as features is $O(N \log N)$, where N is the total number of pixels. Note that, once the EMST is computed, the gradient computation takes $O(N)$ time. Hence, each iteration of the proposed algorithm has a time complexity of $O(N \log N)$. The following is the pseudo-code for the implementation:

0 Initialize transformation parameters with zero: $\mathbf{t_0} = \mathbf{0}$

1 Starting from the lowest resolution, at each level of the image pyramid:

- Initialize transformation parameters: $\mathbf{t} = \mathbf{t_0}$.
- For $iter = 1 \ldots maxIter$
  - if ($iter\%k == 1$) compute the EMST of samples from $(I_1, I_2^{\mathbf{t}})$ and corresponding $\mathbf{u}_d$.

- Update transformation parameters: $\mathbf{t} = \mathbf{t} + stepSize * \mathbf{u}_d$.

- $\mathbf{t_0} = \mathbf{t}$.

2 Warp floating image with $\mathbf{t}$.

5.1. Experimental results

We present some preliminary experimental results obtained using the pair of images[1] shown in Figure 2. Since both images are artificially generated, ground truth for the alignment is known. Prior to the experiment, both images were corrupted with i.i.d Gaussian noise with a variance 0.1 times the maximum value of the original image and one of the images was transformed using a rigid body transformation with known parameters (t_x, t_y, θ). Tables 1 and 2 provide results obtained using the proposed algorithm and a popular mutual information (MI) based algorithm [10] with three levels in the image pyramid (translation in pixels, rotation angle in degrees). The RMS error between the correct and recovered values and the average recovered transformation parameters are listed for three different cases. These results were obtained by averaging over ten trials for each case. Note that in all three cases both algorithms yield a registration with sub-pixel accuracy. The MI-based algorithm, however, has a higher time complexity ($O(N^2)$), where N is the total number of pixels). Our implementation of the proposed algorithm was approximately 10 times faster with compatible parameter values. Both algorithms were implemented in C as consistently as possible.

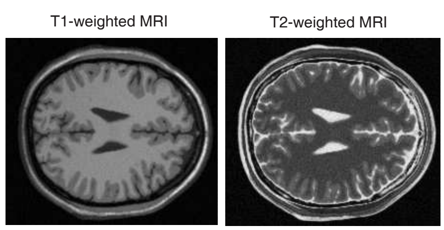

Fig. 2.

6. DISCUSSION

In this paper, we have presented some new results for optimizing graph-theoretic registration functions using gradient information. Initial experimental results indicate that the proposed algorithm can yield robust and accurate registration in a multi-modal application. Although these results are preliminary, the comparison to the standard MI algorithm is encouraging. Extension of the proposed method to non-linear warpings and higher dimensional features is in progress and more complete experimental results will be reported in the near future.

[1]Images obtained from http://www.bic.mni.mcgill.ca/brainweb/

correct param.			algo. output param.			RMS Error		
t_x	t_y	θ	t_x	t_y	θ	t_x	t_y	θ
-1.0	4.0	0	-0.86	4.02	-0.11	0.18	0.06	0.43
8.0	2.0	0	8.10	1.71	-0.31	0.31	0.39	0.50
1.0	4.0	-5.0	1.08	3.88	-4.79	0.54	0.67	0.59

Table 1. Registration Results: Proposed Algorithm. Run time: 12.7 sec.(averaged over 30 trials)

correct param.			algo. output param.			RMS Error		
t_x	t_y	θ	t_x	t_y	θ	t_x	t_y	θ
-1.0	4.0	0	-0.72	4.45	-0.15	0.35	0.45	0.16
8.0	2.0	0	8.14	1.89	0.46	0.15	0.10	0.47
1.0	4.0	-5.0	0.76	3.34	-4.39	0.45	0.96	0.85

Table 2. Registration Results: MI-based Algorithm. Run time: 120.5 sec. (averaged over 30 trials)

7. REFERENCES

[1] D. Eppstein, "Spanning trees and spanners," *Tech. Report 96-16*, Dept. of Info. and Com. Sci., Uni. of California, Irvine, May 1996.

[2] J. Beirlant, E.J. Dudewicz, L. Gyorfi and E.C. van der Meulen. "Nonparametric entropy estimation: an overview," *International Journal Math. Stat. Sci.*, vol.**6**, pp. 17-39, 1997.

[3] A.O. Hero, B. Ma, O. Michel, J.D. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Proc. Mag.*, vol **19**, no **5**, pp. 85–95, 2002.

[4] H.F. Neemuchwala, A.O. Hero and P.L. Carson, "Image registration using alpha-entropy measures and entropic graphs," *European Journal on Signal Processing*, 2002.

[5] B. Ma, A.O. Hero, J.D. Gorman and O. Michel, "Image Registration with Minimum Spanning Tree Algorithm," *IEEE International Conf. on Image Processing '00*, vol.**1**, pp.481-484, Vancouver, BC, Canada, Sept. 2000.

[6] J.B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem" *Proc. Amer. Math. Soc.*, vol.**7**, pp.48-50, 1956.

[7] A.A. Cole-Rhodes, K.L. Johnson, J. LeMoigne, I. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using stochastic gradient," *IEEE Transcations on Image Processing*, vol. **12**, no. **12**, December 2003.

[8] M.R. Sabuncu, P.J. Ramadge, "Spatial Information in Entropy based Image Registration." *International Workshop on Biomedical Image Registration '03*, Philadelphia, July 2003.

[9] J.P.W. Pluim, J.B.A. Maintz and M.A. Viergever, "Mutual Information based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, 2003.

[10] P.A. Viola, W.M. Wells III, H. Atsumi, S. Nakajima, R. Kikinis, "Multi-modal Volume Registration by Maximization of Mutual Information," *Medical Image Analysis*, vol. **1**, no. **1**, pp. 5–51, 1996.

[11] C. Redmond, J.E. Yukich, "Asymptotics for Euclidean functionals with power weighted edges," *Stochastic processes and their applications*, vol. **6**, pp. 289–304, 1996.