

Image registration using entropic graph-matching criteria

Huzefa Neemuchwala^{*#}, Alfred Hero^{*†}, and Paul Carson^{*#}

Dept. of Biomedical Engineering^{*}, Dept. of EECS[†], and Dept. of Radiology[#]
The University of Michigan Ann Arbor, MI 48109

June 10, 2002

Abstract

Image registration requires the specification of a class of discriminatory image features and an appropriate image-dissimilarity measure. Entropic spanning graphs produce a consistent estimator of feature entropy and divergence. We compare direct estimators with non-parametric “plug-in” density estimators, on single pixels and independent image component feature vectors. We have also investigated a technique for minimum spanning tree construction with significantly lower memory and time complexity. On the basis of misregistration errors with decreasing SNR, the minimal graph entropy estimator can have better performance than indirect estimators. In general, misregistration errors are lower with higher dimensional ICA feature vectors as compared to single pixels.

1 Introduction

The focus application of this paper is co-registration of a pair of ultrasound images of the breast, called the reference and the secondary images, respectively. Accurate registration of 3D breast ultrasound image volumes is an essential part of whole breast imaging for detection of asymptomatic breast lesions. Such lesions are missed by community practitioners in up to 45% of women with dense breasts. Here we shall examine cases where the target image correspond to transformed versions of the reference image, e.g. rotation and translation. To date the most effective methods for medical image registration have been pixel and voxel based and include: color histogram matching, texture matching using cross correlation, and, more recently, mutual information maximiza-

tion on pixel coincidence histograms [2], e.g. as used in Radiology at the University of Michigan (the MIAMI-Fuse[®] registration algorithm). While these methods are adequate for some applications, there will always be cases of spurious image qualities such as high speckle to noise ratios, in ultrasound images which will limit registration accuracy and even prevent convergence.

In [1] these authors presented an approach to image registration which gets around the disadvantages of pixel based registration techniques such as the ones previously mentioned. The key to our approach was the inclusion of highly specific image features, such as ICA basis vectors extracted from ultrasound images, and use of a generalized information divergence matching criterion based on the Chernoff bound of detection theory. ICA is an iterative method which is closely related to the projection pursuit technique of non-linear regression and was applied to images by Olshausen, Hyvärinen and others [5, 7]. ICA bases are relevant and robust features of an image and capture non-local spatial information ignored by standard single pixel techniques.

The objective here is to assess the characteristics of the distribution of the feature vectors in the reference and target image, using discriminants such as entropy or other information divergence criteria. For example the mutual information method of image registration searches through a number of coordinate transformations to find the one that minimizes the entropy of joint pixel distributions. Entropic spanning graphs such as minimum spanning trees (MST’s) span a set of feature vectors and produce a consistent estimator of feature entropy and divergence. The MST is constructed over the set of joint feature vectors from the target and reference image to yield Jensen’s entropy difference. We compare and contrast this approach with non-parametric “plug-in” density estimation, using histograms, over feature vectors. In [1] we introduced

[†]Email contact: hero@eecs.umich.edu. This work was supported in part by PHS grant P01 CA85878

image independent vectors as feature vectors for registration and reported an increase in the discriminating ability of the mutual α -information matching criteria. Mutual alpha-information is a ranking discriminant applied to joint densities estimated from histograms. We study the discrimination properties of these estimators using both single pixels and independent image components.

2 α -Jensen Difference Function

The α -Jensen index function has been independently proposed by Ma [3] and He *et al* [6] for image registration problems. Let f_0 and f_1 be two densities and $\beta \in [0, 1]$ be a mixture parameter. The α -Jensen difference is the difference between α -entropies of the mixture

$$f = \beta f_0 + (1 - \beta) f_1 \quad (1)$$

and the mixture of the α -entropies of f_0 and f_1 :

$$\Delta H_\alpha(\beta, f_0, f_1) = H_\alpha(\beta f_0 + (1 - \beta) f_1) - [\beta H_\alpha(f_0) + (1 - \beta) H_\alpha(f_1)] \quad (2)$$

where $\alpha \in (0, 1)$. The α -Jensen difference is a measure of dissimilarity between f_0 and f_1 : as the α -entropy $H_\alpha(f)$ is concave in f it is clear from Jensens inequality that $\Delta H_\alpha(\beta, f_0, f_1) = 0$ iff $f_0 = f_1$ a.e.

The α -Jensen difference can be motivated as an index function as follows. Assume two sets of labeled feature vectors $Z_0 = \{Z_0^{(i)}\}_{i=1, \dots, n_0}$ and $Z_1 = \{Z_1^{(i)}\}_{i=1, \dots, n_1}$ are extracted from images X_0 and X_1 , respectively. Assume that each of these sets consist of independent realizations from densities f_0 and f_1 respectively. Define the union $Z = Z_0 \cup Z_1$ containing $n = n_0 + n_1$ unlabeled feature vectors. Any consistent entropy estimator constructed on the unlabeled $Z^{(i)}$'s will converge to $H_\alpha(\beta f_0 + (1 - \beta) f_1)$ as $n \rightarrow \infty$ where $\beta = \lim_{n \rightarrow \infty} n_0/n$. This motivates the following consistent minimal-graph estimator of Jensen difference for $\beta = n_0/n$:

$$\Delta \hat{H}_\alpha(\beta, f_0, f_1) = \hat{H}_\alpha(Z_0 \cup Z_1) - [\beta \hat{H}_\alpha(Z_0) + (1 - \beta) \hat{H}_\alpha(Z_1)] \quad (3)$$

where $\alpha \in (0, 1)$, $\hat{H}_\alpha(Z_0 \cup Z_1)$ is the minimal graph entropy estimator constructed on the n point union of both sets of feature vectors and $\hat{H}_\alpha(Z_0)$, $\hat{H}_\alpha(Z_1)$ are constructed on the individual sets of n_0 and n_1 feature vectors, respectively. We can similarly define the density-based estimator of Jensen difference based on the entropy

estimates of the form constructed on $Z_0 \cup Z_1$, Z_0 and Z_1 . For some indexing problems the marginal entropies $\{H_\alpha(f_i)\}_{i=1}^K$ over the database are all identical so that the indexing function $\{H_\alpha(\beta f_0 + (1 - \beta) f_i)\}_{i=1}^K$ is equivalent to $\{\Delta H_\alpha(\beta, f_0, f_i)\}_{i=1}^K$

3 Minimum Spanning Tree and Renyi Entropy

The minimum spanning tree (MST) is a graph-theoretic technique, which determines the dominant skeletal pattern of a point set by mapping the shortest path of nearest neighbor connections. Given a set $Z_n = \{z_1, z_2, \dots, z_n\}$ of n , i.i.d vectors Z_i in \mathbf{R}^d each with density f , a spanning tree is a connected acyclic graph which passes through all coordinates associated with Z_n . In this graph all n points are connected by $n - 1$ edges $\{e_i\}$. For a given real weight exponent $\gamma \in (0, 1)$ the minimum spanning tree is the spanning tree which minimizes the total edge weight

$$L(Z_n) = \min_{e \in T} \sum_e \|e\|^\gamma \quad (4)$$

where $\|e\|$ denotes Euclidean (L2) norm of the edge. The overall length of the MST can be used to construct a strongly consistent estimator of Lebesgue continuous densities [3].

The MST length $L_n = L(Z_n)$ is plotted as a function of n in Fig. 1 for the case of uniformly and non-uniformly distributed points for $\gamma=1$. It is intuitive that the length of the MST spanning the more concentrated non-uniform set of points increases at a slower rate than does the MST spanning the uniformly distributed points. This fact motivates the application of MST to test randomness of a set of points. The length function when normalized by \sqrt{n} produces sequences that converge within a constant factor to the alpha entropies with $\alpha=1/2$, as illustrated in Fig. 1. Furthermore by changing the value of γ in Eqn.4, one can change the convergent limit to the α -entropy for $\alpha = (d - \gamma/d)$. The MST is an entropic spanning graph as its normalized log-length converges (a.s.) within a constant to an alpha-entropy. Specifically, Renyi entropy

$$\hat{H}_\alpha(Z_n) = 1/(1 - \alpha)[\ln L(Z_n)/n^\alpha - \ln \beta_{L,\gamma}] \quad (5)$$

is an asymptotically unbiased and almost surely consistent estimator of the α -entropy of f where $\beta_{L,\gamma}$ is a constant bias criterion independent of f . In image registration,

when two images are properly matched, corresponding regions of interest should overlap and the resulting joint probability distribution is highly concentrated. Thus the Renyi entropy of the overlapped images should achieve the minimum. This would be reflected in the length of the MST. A transformation that minimizes Renyi entropy can be calculated, since misregistration would increase the dispersion of the joint probability distribution.

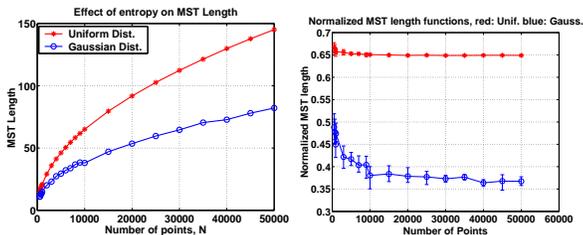


Figure 1: Length functions L_n of MST (left) and L_n/\sqrt{n} (right) as a function of n for the uniform and normal distributed points in Figure.

As contrasted with density based estimates of entropy, the MST estimator enjoys the following properties: it has a faster asymptotic convergence rate, especially for non smooth densities and for low dimensional feature spaces; it completely by-passes the complication of choosing and fine-tuning parameters such as histogram bin size, density kernel width, complexity and adaptation speed; the α parameter in the α -entropy function is varied by varying the inter-point distance measure used to compute the weight of the MST [3]. On the other hand the need for combinatorial optimization may be a bottleneck for a large number of feature samples. To overcome this problem, we suggest the following technique.

3.1 Computational Acceleration of the Kruskal MST Algorithm

The Kruskal Algorithm [4] is widely believed to be the fastest algorithm to solve the MST problem for sparse graphs. Edges, sorted by their weights, are maintained in a list and the algorithm grows the tree an edge at a time. Cycles are avoided within the tree by discarding edges that connect two sub-trees already joined through a prior established path. The time complexity of the algorithm is of $O(E \log E)$ where E is the initial number of edges in the graph. The memory requirement is of $O(E)$.

In the present application, for lack of a better starting

point, the initial estimate of the MST has all the possible edges within the point set. This results in N^2 edges for N points; a time requirement of $O(N^2)$ and a memory requirement of $O(N^2)$. The number of points in the graph is the total number of pixels participating in the registration from the two images. If each image has $M \times P$ pixels, the total number of points in the graph is $2 \times M \times P \approx 150,000$ for ultrasound images of size 256×256 . Desktop processors cannot fulfill memory requirements of the standard Kruskal algorithm in this case. Even with larger machines, the algorithm has a forbidding time requirement for tree construction.

A significant acceleration can be obtained however by a process of sparsification of the initial graph before tree construction. A selection criterion is imposed on the edges, which ensures that only those edges likely to occur in the final MST are included in the original graph. While constructing the edge list, a disc is placed on each point under consideration. As seen in Fig. 3, only those edges with lengths smaller than disc radius are accepted into the list. The edge-length sort algorithm, within Kruskal's algorithm, now has to sort $O(N)$ number of edges. For approximately uniform distributions, a constant disc radius is optimal for all areas within the distribution. Moreover, for non uniform distributions, the disc radius may be changed to adapt according to the underlying distribution. This can be achieved by selecting the distance of the k^{th} -nearest neighbor (kNN) as the disc radius for a given point. Further reduction in time and memory requirements can be obtained by first rank-ordering vertex coordinates along an arbitrary dimension. Now, it is not necessary to compute all N^2 edge-lengths. Only those edges that have lengths less than the disc radius in the dimension of ordering need to be considered (Fig. 3). Thus N^2 edge length computations can also be avoided. If during tree construction the algorithm runs out of edges, expanding the radius of the disc reaps-in additional edges. Fig. 2 shows bias of the modified MST algorithm as a function of the radius parameter.

It is straightforward to prove that the resultant tree is a minimum spanning tree.

Proof: We start by assuming that the Kruskal algorithm is an optimal algorithm to construct MST [4], which is based on the following lemma.

Lemma: If Kruskal algorithm does not include an edge $e = (x, y)$, then at the time that the algorithm considered e , there was already a path from x to y in

the algorithms' partial solution (where x and y are the endpoints of e).

There are 2 conditions to be considered here.

(1) If point p_i is included in the tree, then the path of its connection to the tree has the lowest weight amongst all possible non-cyclic connections. To prove this is trivial. The disc criterion includes lower weight edge before considering an edge with a higher weight. Hence, if a path is found by imposing the disc, that path is the smallest possible non-cyclic path. The non-cyclicity of the path is ensured in the Kruskal algorithm through a standard Union-Find data set.

(2) If a point p_i is not in the tree, it is because all the edges between p_i and its neighbors considered using the disc criterion of edge inclusion have lead to a cyclic path (using the lemma above). Expanding the disc radius would then provide the path which is lowest in weight and non-cyclic.

If the disc radius is underestimated the tree cannot complete without first including more edges into the list. If it is overestimated a surplus of edges will result in the edge list, however the final tree will have the required $N - 1$ edges only. It has been observed empirically that the optimal disc radius includes roughly between 10-20 edges from neighboring points. This number increases with the dimensionality of the data. The number of edges E thus reduces from N^2 to roughly $N \times 10$. The memory requirement of the modified algorithm is of $O(E)$. The time requirement now optimizes to $O(E \log E)$, where E is a fraction of N^2 for large N . Fig. 3 compares the performance of the standard Kruskal algorithm with our modified algorithm. The time and memory requirements are tremendously reduced. Several approaches to further decrease the time-memory complexity of the Kruskal algorithm have been studied by the authors but for lack of space cannot be described here. The authors have successfully constructed minimum spanning trees with up to 1 million points in 8-dimensional space using variants of the techniques discussed above.

4 Registration Results

Three test cases were chosen from the breast database. These will be referred to as Case 1, Case 2 and Case 3, shown in [1]. The image slice chosen from Case 1 contained connective tissue structure which are characterized through the edges in the image. Case 2 contained a dis-

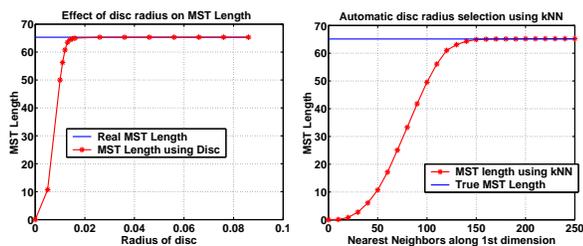


Figure 2: Bias of the $n \log n$ MST algorithm as a function of radius parameter (left) and as a function of the number of nearest neighbors (right)

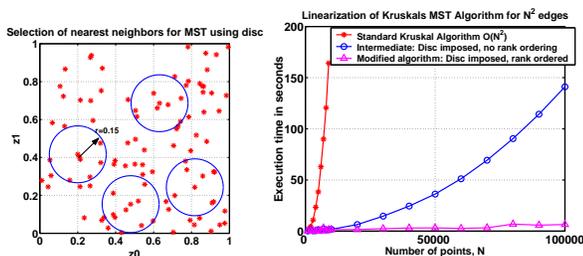


Figure 3: Acceleration of Kruskals MST algorithm from $n^2 \log n$ to $n \log n$ (left) and Comparison of Kruskal's MST to our $n \log n$ MST algorithm (right)

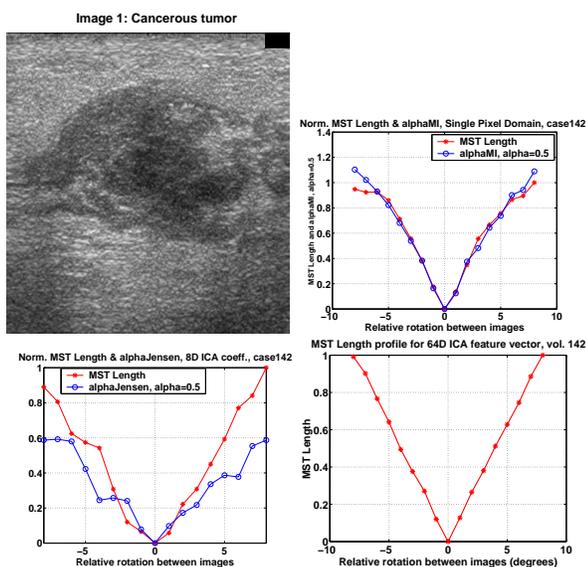


Figure 4: Top row: Ultrasound breast scan case 2 (left). MST and histogram based entropy estimation for single pixel features (right). Bottom row: MST and histogram based entropy estimation for 8D ICA features (left) and 64D ICA feature vectors (right)

tinct malignant tumor (seen in Fig. 4), while Case 3 was degraded due shadowing. We simulated the decorrelating effect of speckle by registering a slice to a rotated version of a proximal but different slice (approximately 2mm away along the depth of the scan).

Fig. (4) show the representative profiles of the objective function for case 2 under different transformations. For single pixel features, the estimates using direct and indirect estimation methods are approximately equal. For higher dimensional ICA feature vectors, we observe a progressive linearization of the profile with increasing dimensionality for the MST based direct estimation techniques. Computing the histogram estimate of α -MI is exponentially complex, since the number of cells increase exponentially with dimensionality. On the other hand, the bias of the estimate increases exponentially as dimensionality, for a fixed number of cells. MST based estimates however, do not have this drawback and can be computed using the full 64 dimensions of the ICA features.

Additive noise and speckle typically degrade registration accuracy. We tested registration accuracy for single pixel and ICA features, and the discriminating criteria under increasing noise conditions. Fig. (5) plots registration errors versus increasing levels of additive (truncated) Gaussian noise to the images. The resultant registration peak shifts from the perfect alignment position (0 degrees relative rotation), to some arbitrary value depending on the SNR, the registration features and entropy estimation techniques. In general, the performance of higher order ICA features is seen to be better than those of single pixel features. Also, for higher order ICA features, the MST method of entropy estimation demonstrate a greater robustness to additive noise than histogram based methods.

5 Acknowledgements

To Sun Chung PhD, Computer Science Department, St. Thomas University for the valuable comments to accelerate the MST algorithm.

References

[1] H.Neemuchwala, A.O. Hero, and P.Carson, "Feature coincidence trees for registration of ultrasound

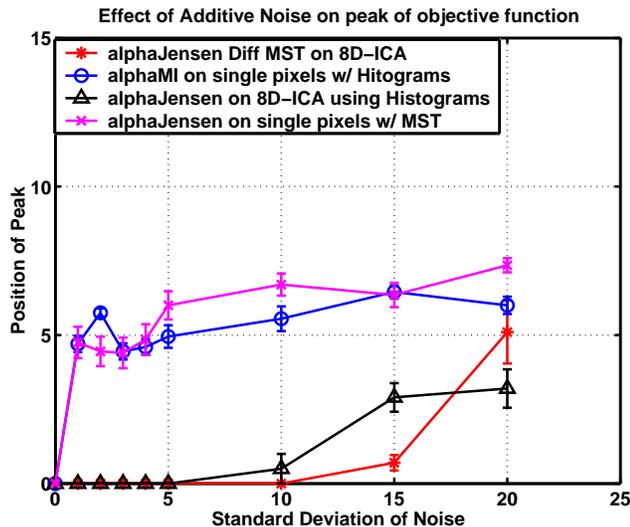


Figure 5: Effect of additive noise on the peak of the registration objective function

breast images”, *Proc.of IEEE Int. Conf. on Image Proc., Thessaloniki, Greece*, Oct 2001.

- [2] P.Viola and W.Wells, “Alignment by maximization of mutual information”, in *Proc. of 5th Int. Conf. on Computer Vision*, MIT, volume 1, pp.16-23, 1995.
- [3] A. O. Hero, B. Ma, O. Michel and J. Gorman, ”Applications of entropic spanning graphs,” to appear in *IEEE Signal Proc. Magazine (Special Issue on Mathematical Imaging)* Oct. 2002.
- [4] J. B. Kruskal, “On the shortest subtree of a graph and the traveling salesman problem”, *Proc. American Math. Societ*, vol. 7, 48-50, 1956.
- [5] A. Hyvarinen and E. Oja, “Independent component analysis: algorithms and applications”, *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 1999.
- [6] Y.He, A.Ben-Hamza and H. Krim, “An information divergence measure for ISAR image registration”, *Signal Processing*, submitted, 2001.
- [7] M.Lewicki and B. Olshausen, “Probabilistic framework for the adaptation and comparison of image codes”, *J. Opt. Soc. Am.*, vol. 16, no. 7, pp.. 1587-1601, 1999.