NSF BES-02: FEATURE-BASED ULTRASOUND IMAGE REGISTRATION

A 3 year Grant Proposal to

NATIONAL SCIENCE FOUNDATION, DIVISION OF BIOENGINEERING AND ENVIRONMENTAL SYSTEMS PROGRAM IN BIOMEDICAL ENGINEERING AND RESEARCH TO AID PERSONS WITH DISABILITIES

October 2, 2001

PI: Alfred O. Hero,
Depts. of Electrical Engineering and Computer Science, BioMedical Engineering, and Statistics,
University of Michigan, Ann Arbor

co-PI: Paul Carson,
Dept. of Radiology,
University of Michigan, Ann Arbor

Project Summary

We propose to develop a powerful mathematical feature selection and volumetric registration method to detect and track low contrast lesions from 3D ultrasound scans of organs and tissues in the body. Our approach is based on a new class of feature matching classifiers and a new feature selection criterion which are specifically adapted to registering a primary image to a secondary image. The features are arranged in a hierarchical tree structured database which is used for detecting the coincidence of common features at common locations in the two images. Image registration is performed by searching over a sequence of spatial transformations of the secondary image for the one that optimizes the feature coincidences in common locations. The registration procedure is based on a new mutual-information image registration function which is very robust to small differences between the features, e.g. due to presence of a tumor or muscle tear in only one of the images. The procedure is implemented using fast, accurate and robust graph-theoretic estimators of mutual-information, specifically the k-means algorithm and the minimal spanning tree, and, unlike previous information-theoretic registration methods, does not suffer from the difficulties of joint density estimation. As a result of more accurate image registration, the algebraic differences between the registered images will reveal lesions which were previously hidden due to their low contrast and small size. We expect the outcome of this research to lead to improved quantification of medical conditions such as: healing, scaring or fibrosis of muscle tears and ischemias; and genesis, growth and necrosis of malignant breast tumors.

Use of image registration to aid in detection and tracking of breast cancer and muscolo-skeletal lesions is relatively new and not widespread. Only recently, with single voxel mutual information techniques, has such registration appeared to be very promising. The methods we propose to investigate are new and have not been previously applied to image registration. Based on preliminary results these methods appear to offer significant improvement on single pixel based methods for registration of ultrasound breast images. Our approach combines two very powerful new ideas: 1) improvement of registration accuracy and robustness by automated selection of robust, sensitive and parsimonious spatial features derived from adaptive quantization and independent components analysis (ICA); 2) use of registration functions which extend the mutual information to a more stable criterion which has provable detection theoretic optimality properties. These two ideas will be combined into a novel, computationally-tractable registration algorithm using divide-and-conquer processing, and validated with testing on actual ultrasound scans. The participation in this project of leaders in image processing and radiology gives us an unprecedented opportunity to make a leap forward in image registration for medical imaging.

1 Introduction

This is a proposal to develop new approaches to image registration for the purpose of detecting and tracking lesions from multiple ultrasound scans. While the methods to be developed will have wide applicability to many other types of imaging modalities and applications our methods will be tested and validated on ultrasound images of breast and musculo-skeletal tissues. Accurate registration of multiple medical image scans is essential for detection of functional and anatomical anomalies and assessment of the evolution of such anomalies over time. For example in whole breast imaging, registration of 3D breast ultrasound image volumes is essential to detecting small breast lesions; discrimination of malignant from benign lesions; detection of secondary masses; and quantifying response to treatment. Advances in image registration of breast images will have substantial societal impact as typical community practitioners currently miss asymptomatic breast lesions in up to 45% of women with dense breasts [29]. Accurate registration will aid diagnosis of benign and malignant masses by facilitating visual or automatic comparison with the structures in the same region at the time of a previous examination. Image volume registration may aid tracking of response to cancer therapy as well. Success in these goals should lead to earlier detection and diagnosis of many breast cancers, as well as reduced pain and uncertainty for those with suspicious lesions and more appropriate dosages and selections of therapies. Another medical area where our results will have great impact is in tracking the evolution of muscle tears and sprains, post-surgical recovery and other areas in sports medicine and physical rehabilitation.

Paul: need put some extra stuff in here about musculo-skeletal impact of registration.

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 maximization on pixel coincidence histograms [31] as used in Radiology at the University of Michigan and elsewhere. The variant used at UM, called MIAMI-Fuse© (mutual information for automatic multimodality image fusion), has been shown to significantly outperform the older correlation and template-matching methods. We and others have shown that mutual information-based, linear and nonlinear registration can be performed with a quality that will meet the needs of many medical applications of X-ray CT, magnetic resonance imaging (MRI), positron emission tomography (PET) and single photon emission tomography (SPECT). However, MI methods have proven overly sensitive to spurious image components and artifacts due to speckle, shear/compressive tissue deformation, and shadowing which are ubiquitous to modalities such as ultrasound (UL) imaging. As UL imaging is a cheap and widely available diagnostic imaging modality, improving UL image registration can have wideranging medical impact.

This 3 year grant proposal describes a new and powerful approach to image registration which uses stable higher order image features in addition to single pixel or single voxel gray scale levels. In so doing we can capture complex spatial information that is ignored in the current state of the art single pixel mutual information and correlation-based registration techniques. Further, this spatial information can permit more accurate and robust image and image volume registration which overcomes the limitations of previous mutual-information based registration techniques. We propose to apply a matching method which is provably optimal, in terms of minimizing the misclassification probability. We propose to perform feature selection using state-of-the-art inductive learning methods including randomized feature classification trees and randomized independent component analysis (ICA). We propose to evaluate the goodness of match between a pair of images from the co-occurences of their spatially localized features. Finally, we propose to develop fast methods for maximizing the number of feature co-incidences based on minimal spanning graph algorithms and a new registration function as a feature matching criterion: the α-mutual information.

2 Background

Precise registration of image volumes to allow comparison of ultrasound studies separated by months or years should greatly facilitate detection of changes and assessment of their potential malignant growth patterns. New techniques developed by our ultrasound (UL) group in Basic Radiological Sciences (see http://www.ultrasound.med.umich.edu/) are ready for evaluation and refinement in clinical studies and include: registration of serial studies in single and multiple modes, creation of expanded 3D fields of view, a moderately fast, manual method of segmenting ellipsoidal volumes, quantification of vascularity in regions of interest, and improved UL imaging by compounding (combining multiple views for improved image contrast to noise ratio) using image volume based registration (IVBaR)[7-10]. Work is also progressing on computer aided diagnosis (CAD) of UL masses [11], with potential links between UL and x-ray CAD. What is needed is a much more accurate and robust registration method to provide the needed confidence in the registration for studies with a wider range of image quality. Segmentation of artifacts, such as acoustic shadows and reverberations, is particularly important, for the artifacts can be the dominant "information" in some cases, and yet are affected strongly by subtle changes in the UL view Figure 1 shows examples of three typical but very different breast scans which illustrate the challenge of image registration for this application.

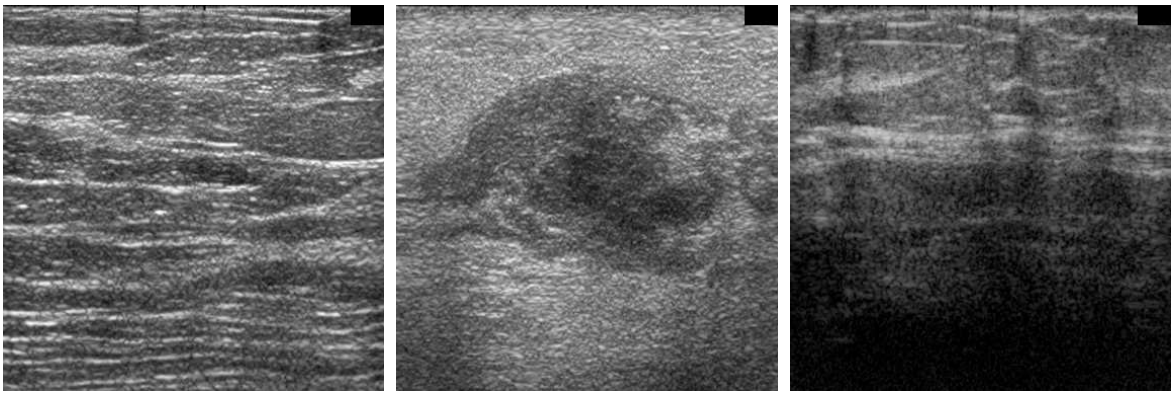

Figure 1: 2D slices from three 3D ultrasound breast scans. From top to bottom are: case 151, case 142 and case 162. The image slice chosen from Case 151 (image slice 40) had significant connective tissue structure. Case 142 (slice 35) had a distinct malignant tumor, while Case 162 (slice 60) was degraded due shadowing.

Registration of an image to a database of rotated and translated exemplars constitutes an important image retrieval and indexing application which arises in biomedical imaging, digital libraries, georegistration, and other areas. Let X_0 be a the reference image and consider a secondary image X_1 and sequence of deformed secondaries $X_i = T_i(X_1)$, $i = 1, 2, \ldots$. Let Z_0, Z_i be feature vectors extracted from X_0, X_i and define the joint histogram $f_{0,i}(z_0, z_i)$ and the marginal histograms $f_0(z_0)$, $f_i(z_i)$. The similarity between features Z_0 and Z_i can be gauged by the difference between $f(z_0, z_i)$ and the product $f(z_0)f(z_i)$ which measures statistical dependence. Alternative measures of dependency are cross correlation, higher order cross-moments like joint kurtosis, and mutual-information (MI) methods based on the entire feature histogram.

The simple single-pixel mutual-information method introduced by Viola and Wells [31] is an example of the general MI approach. Their technique has been refined and adapted by several medical image registration groups including a group at The University of Michigan (UM) Dept. of Radiology [24, 23, 19]. In all of these groups the MI-based technique have supplanted previous correlation methods due to its superior performance. The UM method, called MIAMI Fuse© functions as follows. A randomly or intelligently chosen set of voxels, called control points, are selected in the secondary image and associated

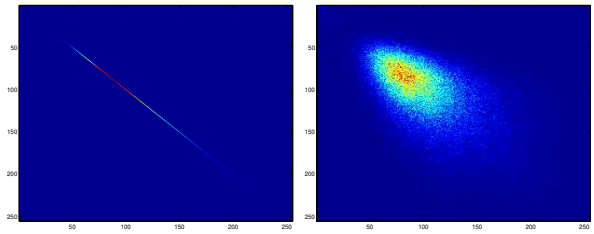

Figure 2: Single-pixel gray-level histogram $\hat{f}_{0,i}$ for registering a single slice to rotated rotated version of itself (relative rotation angle = 5%). The slice is taken from the ultrasound breast image shown in first panel of Fig 1. Horizontal and vertical axes range over the gray levels (0-255). Observe that a straight line is obtained (panel at right) when perfect registration occurs.

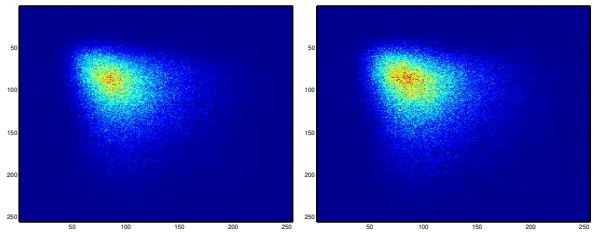

Figure 3: Same as in Fig. 2 except that primary and secondary images are taken from adjacent slices separated by 2mm. Due to the presence of speckle we do not obtain a straight line even when images are registered (left panel).

with knot positions on a thin-plate spline. The bending resistance of the thin plate spline cause a smooth deformation of the entire voxel grid on the secondary image as the control points are moved around. A sequence of deformations is defined by successive movements of the control point positions giving rise to the sequence of deformed secondary images $X_i = T_i(X_1)$, $i = 1, 2, \ldots$ obtained by interpolation of X_1 onto the deformed grid. Let X_0 and X_i be $N \times N$ images (N^2 pixels). In single-pixel registration techniques such as MIAMI-fuse©, for each iteration i a set of feature pairs $\{(Z_0(k), Z_i(k))\}_{k=1}^{N^2}$ are defined as the successive pairs of gray levels (typically ranging from 0-255) of the lexicographically indexed primary and deformed secondary images, respectively. A two dimensional gray level empirical histogram $\hat{f}_{0,i}(z_0, z_i)$, $z_0, z_i \in \{0, \ldots, 255\}$, is then generated from these feature pairs $\{(Z_0(k), Z_i(k))\}_{k=1}^{N^2}$ yielding an extended point cloud in the plane if the images are not registered and a straight line if they are perfectly registered (see Figure 2). A good feature matching criterion would be a measure of the degree of independence between Z_0 and Z_i, or equivalently the degree to which the $\hat{f}_{0,i}(z_0, z_i)$ can be factored into the product $\hat{f}_0(z_0)\hat{f}_i(z_i)$, where $\hat{f}_0(z_0)$ and $\hat{f}_i(z_i)$ are the marginals. The Shannon MI is one such feature matching criterion

$$\mathrm{MI}(\hat{f}_{0,i}) = \sum_{z_0} \sum_{z_i} \ln \hat{f}_{0,i}(z_0, z_i)/(\hat{f}_0(z_0)\hat{f}_i(z_i)). \tag{1}$$

Registration occurs when this registration function is maximized, and the search over deformations T_i for this maximum is performed by an iterative optimization algorithm.

When the image volumes are identical except for rotation, translation and small deformations, the MI will exhibit a very sharp peak over $i = 1, 2, \ldots$ for that deformation T_i which achieves perfect registration of the images, as in the perfectly registered example shown in the left panel of Fig 2. However the assumption that there exists a perfect registration is unrealistic in practical cases where cancer detection, diagnosis and treatment response are the goal. In the latter case, the reference and secondary images are

genuinely somewhat different from each other due to biological changes and differences in the imaging, such as speckle, positioning of the tissues during compression, and angle dependence of UL scattering from tissue boundaries. In addition, the tissues are distorted out of a given image plane as well as within it leading to poor performance of slice-by-slice (2D) registration. As a result, even registered images will not be perfectly correlated (see Fig. 3) and the MI peak will not be as sharp as in the idealized situation. The combination of fully 3D feature registration techniques and more sensitive feature matching measures that we propose in this grant proposal can overcome these problems.

3 Proposed Methods

We take a novel approach to image registration that gets around the disadvantages of previous single pixel- or voxel-based registration techniques. The key to our approach is the inclusion of higher order image features combined with generalized information divergence criteria for feature matching. We will use inductive learning to select the most relevant and robust features (curves, edges, textures and simple spatial relations) from a large set of previously collected scans, e.g. of breast or Achilles tendon. These features are organized in an efficient hierarchical database, called a randomized feature tree. To register a secondary image to a reference image, coincidence histograms of the features captured by the feature tree for the reference and secondary images are constructed. The coincidence histograms are used to register the images by optimizing a new class of registration functions over the space of deformations of the secondary image. This class of registration functions includes the mutual α-information and the α-Jensen difference which will be estimated using a combination of feature density estimation and a novel minimal graph matching technique developed by Hero and co-workers [22, 11]. Our framework specializes to the standard (MIAMI-Fuse©) registration method when the features are single-pixel gray levels and $\alpha = 1$ for which the mutual α-information is equivalent to Shannon MI. The generality of our approach has the following advantages: 1) use of optimized feature matching measures can lead to much more stable registration function having much higher discrimination capability; and 2) use of higher order features can capture non-local spatial information which is ignored by previous algorithms.

3.1 Optimal Feature Matching Criteria

We adopt a systematic decision-theoretic framework for selecting the registration function. In this framework we assume that the feature point cloud, e.g. either panel of Fig. 3, is a random realization of some unknown underlying joint density $f_{0,i}$. A good registration function will satisfy the following criteria: 1) high average accuracy over an ensemble of feature realizations from joint density $f_{0,i}$, corresponding to accuracy averaged over a large representative image database; 2) high accuracy for a single realization of feature points from $f_{0,i}$, corresponding to accuracy for a single pair of images; 3) robustness to outliers, i.e. spurious features that would normally confound registration; 4) fast algorithm implementation.

Optimal registration functions which satisfy the above criteria will be developed and studied in a systematic three step process. First we study theoretically optimal *omniscient* registration functions which have access to the ensemble statistics of the image database, i.e., the exact feature density function $f_{0,i}$. These ensemble statistics will depend on the medical imaging application. Then, we explore methods of optimal empirical estimation of an omniscient registration function based on a single pair of images. Optimal estimation strategies will also depend on the medical imaging application. Finally we will explore fast and robust implementations of these empirical registration function estimates using graph matching techniques.

3.1.1 Omniscient Registration Functions

The above considerations have led us to propose the class of α-information criteria ($\alpha \in [0,1]$) and their close relatives, the α-Jensen differences.

As above let $f_{0,i}(z_0, z_i)$ be the true (ensemble) joint density between feature vectors of the primary and deformed secondary images. The (ensemble) α-information is defined as

$$D_\alpha(f_{0,i} \parallel f_0 f_i) \quad = \quad \frac{1}{\alpha - 1} \log \int f_{0,i}^{\alpha}(z_0, z_i) f_0^{1-\alpha}(z_0) f_i^{1-\alpha}(z_i) dz_i dz_0. \tag{2}$$

When specialized to various values of α the α-information reduces to well known information measures such as the Battacharya-information ($\alpha = 1/2$), and Shannon MI used in MIAMI-Fuse© ($\alpha = 1$). The α-information is a special case of "f-information" which has been investigated in a recent paper [28] for medical image registration and also generalizes the mutual-information method of Viola and Wells [31]. However, as shown by Hero and his co-workers [11], the α-divergence is the most appropriate member of the f-information class as it is directly related to the minimum attainable decision error incurred when trying to discriminate between an independent (factorizable) and a dependent (nonfactorizable) feature density $f_{0,i}(z_0, z_i)$ based on a finite number of feature samples. In particular, large deviations theory asserts that there exists a theoretically optimal value of α that optimizes the discrimination performance. This value of α is generally not equal to the value $\alpha = 1$ corresponding to the standard Shannon MI and depends on the class of feature densities $f_{0,i}$ under consideration.

A closely related quantity to be investigated is the (ensemble) α-Jensen difference. The α-Jensen difference is of interest due to the existence of lower complexity empirical estimators using minimal spanning graphs over the feature pairs $\{(Z_0(k), Z_i(k))\}_{k=1}^{N^2}$. The α-Jensen difference is

$$\triangle H_\alpha(\beta, f_0, f_1) \triangleq H_\alpha(\beta f_0 + (1-\beta) f_1) - [\beta H_\alpha(f_0) + (1-\beta) H_\alpha(f_1)], \quad \alpha \in (0,1), \tag{3}$$

where

$$H_\alpha(f) = \frac{1}{1-\alpha} \int f^\alpha(z) dz$$

is the α-entropy of the probability density f and $\beta \in [0,1]$ is a primary-secondary mixing parameter whose choice depends on the application, i.e. the choice of features. The α-Jensen difference was originally proposed by us [22, 21] for geo-registration applications including SAR and electro-optical radar imaging A generalization was recently proposed by He, Ben-Hamza and Krim [10] for registering an arbitrary number of image modalities. The α-Jensen difference was proposed several years before by one of our collaborators, Michel *etal* [25], for classifying images in the time-frequency plane.

3.1.2 Empirical Registration Functions

In practice the joint feature density $f_{0,i}$ is unknown so that the above ensemble registration functions are unimplementable. The statistical problem of registration can be stated as follows: based on a single realization $\{(Z_0(k), Z_i(k))\}_{k=1}^{N^2}$ of joint features, e.g. the point cloud illustrated in either of the panels of Fig. 3 corresponding to a single image pair, empirically estimate the ensemble α-information or the ensemble α-Jensen difference. This estimation problem is closely related to entropy estimation which has a long history in the statistics and information theory communities [9, 18, 1, 30, 3, 5, 26].

The approach to empirical estimation that will be taken here is motivated by our recent work in entropy estimation [16, 12, 11]. We will investigate two different techniques: density plug-in techniques and minimal graph techniques. Each of these methods has its own complementary strengths and weaknesses which justify considering both of them for registration applications.

Density plug-in methods are in widespread use for estimation of Shannon mutual-information and other entropy functions [31, 23, 3]. These methods involve non-parametric density estimation, such as the histogram binning method or kernel density estimators, to yield empirical estimates of $f_{0,i}(z_0, z_i)$ and its marginals which are subsequently substituted into an ensemble formula, e.g. (2) or (3). For example, when a non-parametric estimate of $\hat{f}_{0,i}$ is available the plug-in estimate of α-information is $D_\alpha(\hat{f}_{0,i} \parallel \hat{f}_0 \hat{f}_i)$. The main difficulties with density plug-in methods are due to the infinite dimension of the spaces in which the unconstrained densities lie. Specifically: density estimator performance is poor without stringent smoothness conditions; no unbiased density estimators generally exist; density estimators have high variance and are sensitive to spurious outliers; for high dimensional feature spaces the integration in (2) might be difficult.

A more direct estimation method which applies to continuously valued features is based on constructing minimal graphs that span the feature point realizations. For example, to estimate the α-entropy of the primary image's feature density f_0, the minimal spanning tree (MST) method constructs a graph that spans all of the feature realization points $\{Z_0(k)\}_{k=1}^{N^2}$ in such a way so as to minimize the total length of the graph's edges, examples are shown in Fig. 4. Specifically, if the MST puts an edge between $Z_0(i)$ and $Z_0(j)$ the edge has length $\|e_{ij}\| = \|Z_0(i) - Z_0(j)\|^\gamma$ where $\|\cdot\|$ is Euclidean distance and $\gamma = (1-\alpha)d$, where d is the dimension of the feature vector $Z_0(k)$. This method can be applied to estimation of the α-Jensen difference by constructing the MST over the merged sample $\{Z_0(k)\}_{k=1}^{N^2} \cup \{Z_i(k)\}_{k=1}^{N^2}$ and selecting $\beta = 1/2$ in (3). Minimal graph techniques can also be extended to estimation of α-information but for lack of space we will not discuss this here (see [11]).

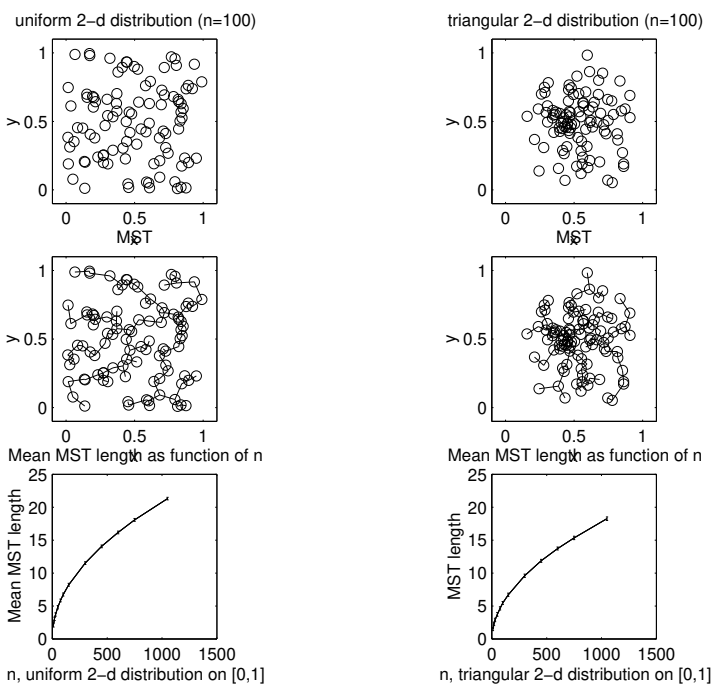

Figure 4: *Illustration of MST estimator of entropy for two dimensional features ($d = 2$) and edge weight exponent $\gamma = 1$. Left column: MST results for uniform feature density f_0 on the unit square. Right column: MST results for pyramidal feature density f_0 on the unit square. First row: $n = 100$ realizations of each feature density; Second row: MST for each case. Third row: total length of MST increases with rate $n^{\frac{d-\gamma}{d}} = \sqrt{n}$ and rate constants equal to the α-entropy of f_0, $\alpha = \frac{d-\gamma}{d} = 1/2$.*

General minimal graph techniques for entropy estimation were introduced by Hero and Michel [14, 15, 16]. We showed the remarkable fact that the length of a minimal spanning graph, such as a minimal spanning tree (MST) or k-nearest neighbors graph (k-NNG), forms a strongly consistent estimator of

the α-entropy. Furthermore, recently obtained theoretical results of Hero and Ma [12] assert that the minimal graph estimator of entropy can be significantly more accurate (lower mean-squared error) than plug-in estimators. Here α is related to the dimension of the feature space and can be varied over 0 to 1 by using different length weighting exponents γ.

We have shown [16, 12] that minimal graph estimators have several advantages over standard plug-in estimators of entropy and mutual information: 1) they do not require specification of difficult auxiliary parameters such as bin size and kernel width; and 2) as shown by Hero and Ma [13], their estimation error can converge to zero much faster than that of density estimation methods; 3) fast implementations exist using MST and k-NNG constructions; 4) optimal pruning methods can be applied to these graphs to render the α-entropy estimates robust to anomalous feature points, e.g. points due to speckle or tumors not present in both images. We believe that such pruning methods will be effective for registering two medical images where one of the images has small lesions. These minimal graph techniques have been applied with success to geo-registration problems by us [22]. A representative example is shown in Figs. 4 which shows a minimal spanning tree and its pruned version using a combinatorial optimization technique known as the k-point MST (see publications [22, 11] for explanation).

3.2 Higher Order Feature Selection

As an alternative to single-pixel features we will investigate higher order features which capture spatial dependencies such as edges, ridges, organ outlines, and other non-local structure. When added to single-pixels such features can produce much more robust features for registration. Two different types of features will be investigated: randomized feature trees which organize simple local image features into a hierarchy of subimages of increasing complexity, and randomized databases of local features selected by independent component analysis (ICA). These feature structures will allow us to sidestep the issue of rigidly defining non-local image configurations and clusters such as, image gradients, boundary detectors, distinct points, discriminating structures or other statistical parameters. For each of these methods we will investigate methods of bagging, boosting and randomization for feature selection from a representative image database used for training.

3.2.1 Randomized Feature Trees

Randomized feature trees were introduced by Amit and Geman [2] for shape recognition from binary transcriptions of handwriting and other computer vision applications. A set of primitive local features, called tags, are selected which provide a coarse description of the topography of the intensity surface in the vicinity of a pixel. Local image configurations, e.g. subimages associated with 4×4 pixel neighborhoods, are captured by coding each pixel with labels derived from the tags. Non-local spatial features are then captured by cataloging pairs of tags which are in particular relative spatial configurations.

The tags will be selected by the adaptive thresholding technique as described by Geman and Koloydenko [8] which was introduced to study the invariant characteristics of natural images. Let Δ be a positive granularity parameter. The quantized value assigned to a pixel within a 4×4 sub-image depends on the gray values of its neighbors. The darkest pixel(s) are assigned 0, the next brightest pixel(s) are assigned 0 if the difference is less than Δ and label 1 otherwise, the next brightest pixel(s) are assigned label 2 if the difference is less than Δ, and so on. Using this scheme on our ultrasound image database tags associated with the relatively uniform background areas (dark or bright) with small spatial variances are correctly classified as speckle and can be easily eliminated. Such tags are irrelevant to the image registration and our preliminary studies indicate that their elimination results in a reduction by almost 75% of the total number of possible tag types. The remaining 25% tag types were observed to distinguish

straight and curved edges in the images.

Once selected, the tag features are efficiently organized on a tree structured database, called the feature tree, as illustrated in Figures 5 and 6 for 4×4 tag subimages. For the purposes of image registration, the primary and deformed secondary images X_0 and X_i will each be dropped down the feature tree and incidences and coincidences of features at all of the nodes of the two trees are counted. The counter is incremented for every coincidence of a particular feature pair occurring at a common position within each of the two images. This results in a feature coincidence histogram $\hat{f}(z_0, z_i)$. The histogram marginals $\hat{f}(z_0)$ and $\hat{f}(z_i)$ of the coincidence histogram are extracted by summing over one of its arguments. These will be then used in the mutual α-information formula (2) to come up with a registration score for the images.

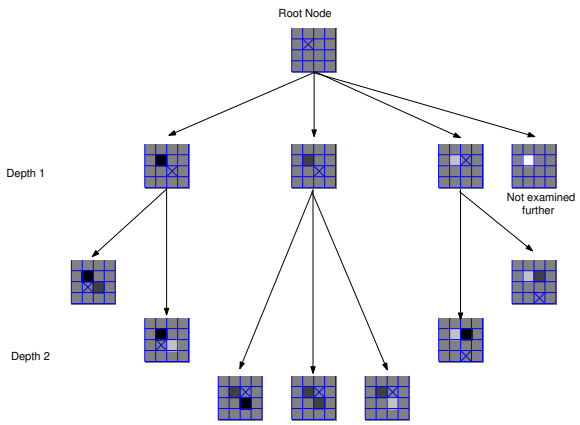

Figure 5: *Part of feature tree data structure.*

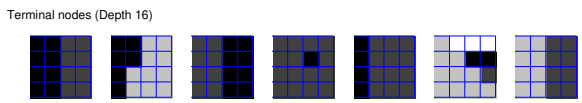

Figure 6: *Leaves of feature tree data structure.*

Feature trees have had the most success in pattern recognition when a large number of features, tags, and spatial relationships are used. The resultant high computation load is reduced by exploiting the recursive partitioning property of the tree hierarchy: for binary trees the search complexity is reduced from K to $\log 2(K)$ and for quad trees further search complexity reductions are obtained, where K is the number of features considered. Additional reductions in average search complexity are obtained by putting the most common and discriminating feature tags near the top of the tree and pruning a large number of the remaining tags. Discovery of the most discriminating spatial relationships and the most discriminatory local features is essential for finding the most convenient feature organization in the tree. We plan to investigate various methods for organizing features and pruning including tree balancing techniques, Huffman coding methods, and inductive learning using randomization [2], bagging [4] and boosting [7]. These latter methods were developed for classification and are based on a forest of different feature trees. Each of these feature trees is constructed from randomly sampling from the database of images and the classifier is obtained by aggregation of decisions of the trees, e.g. using majority voting or averaging of the posterior distributions, to produce the most stable and informative features. We will extend these methods to image registration in two ways: 1) substitution of more appropriate feature selection criteria than the standard miss-classification error; and 2) aggregating trees by averaging the computed α-information or α-Jensen values. Examples of feature selection criteria to be investigated are: the curvature of the empirical registration function and the number of coincident features for co-registered images compared to the average number of coincident features for misregistered images.

3.2.2 ICA Features

Independent components analysis (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 [20, 27, 17]. ICA is an extension of the better known singular value decomposition (SVD) methods and have been shown to yield sparse low dimensional feature spaces for natural images. We apply the method of ICA to the ultrasound registration problem as follows. Based on a random sample of a large number of $M \times M$ subimages from the representative training database of ultrasound images we construct a basis of a small number p of $M \times M$ subimages which decomposes the database of subimages into

$$X_l = \sum_{k=1}^{p} a_{lk} S_k$$

where $\{S_k\}$ are statistically independent components ($M \times M$ images) and a_{lk} are coefficients which approximate each subimage X_l in the database. The selection of of independent components is performed using an independence criterion such as kurtosis, likelihood, or minimum description length (MDL).

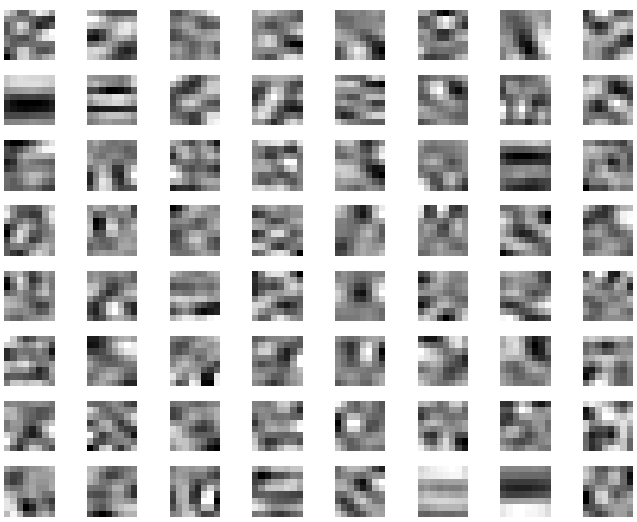

Figure 7: *Estimated ICA basis set for ultrasound breast image database*

As an example, using the ICA algorithm a set of 8×8 basis vectors were learned from 10 consecutive image slices extracted from a single ultrasound volume scan of the breast (Case151) shown in Fig. 1. Here ICA was implemented using Olshausen's `SPARSENET` code (available from `http://redwood.ucdavis.edu/bruno/`). 64 of the ICA basis vectors are shown in Fig. 7. Only basis elements that corresponded to distinct edge or texture information in the associated basis were retained, resulting in a total of approximately 256 different types. Two different methods of constructing the feature vectors Z_0 and Z_i will be investigated: 1) a tag assigned to a particular pixel is the feature type that has least Euclidean distance from the 8×8 sub-image centered there; 2) sub-images of the reference and secondary images are projected onto the ICA basis and either the coefficients or residuals are used to construct the joint histogram. In method 1 standard plug-in estimates of the α-information or α-Jensen difference will be used. In the latter case the joint histogram is the realization of a continuous density $f_{0,i}$ and the minimal graph methods of Hero *etal* will give more accurate estimates of the α-information or α-Jensen difference.

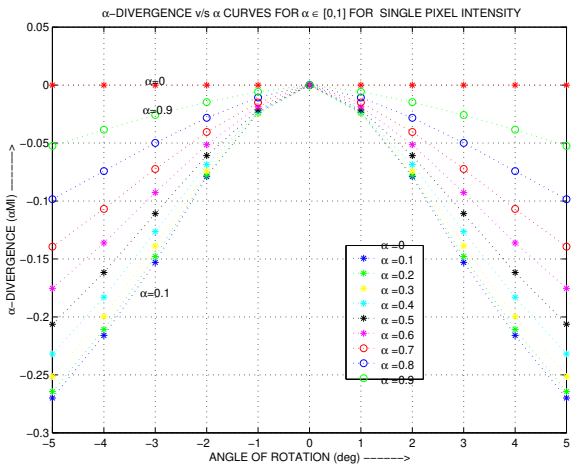

Figure 8: α-Divergence as function of angle for ultra sound image registration of case 151

3.2.3 Spatially Organized Features

By including selected spatial relationships between local tag and ICA features as part of the feature set, one can enhance registration performance. We capture these relationships by superimposing a disk over each pixel and detecting the simultaneous occurrence in the disk of a tag or feature of given type at the center of the sector and another tag of given type in some outlying spatial sector, e.g. in sectors NW, NE, SE, SW. This additional spatial information can be used to discriminate and eliminate undesired regions, e.g. shadow regions, from each image and also to introduce useful local geometric invariances into the registration process. Local invariance is an attractive property in non-rigid image registration, since it allows pixel neighborhoods (tags) to undergo relative displacement without affecting the overall registration. For example local invariance can be built into the features to allow the registration algorithm to be insensitive to small relative offsets between boundaries due to small compression deformations in one of the two images.

4 Preliminary Results

For illustration we used the database of ultrasound breast images represented by the three images in Figure 1. We simulated the decorrelating effect of speckle in two different scans of the same slice of breast by registering a slice to a rotated version of a proximal slice approximately 2mm (up to four correlation lengths) away along the depth of the scan. Feature tree and ICA feature selection were performed without randomization, bagging or boosting. Randomization would be expected to provide even better results than those obtained here. The α-information was estimated from the features using plug-in histogram binning methods. For each of the feature sets investigated we searched over the choice of α in α-information registration criterion which maximized the curvature, i.e. yielded the most highly resolved peak.

Shown in Figure 8 are typical trajectories of the α-information registration function over rotation angles applied to the secondary image slice for the standard single-pixel features and various values of α. Note that the highest peak curvature does not occur for $\alpha = 1$ meaning that that the standard Shannon MI criterion is sub-optimal. Shown in Table I are resolution-optimizing α values and resultant optimal peak curvature values for each of the three breast image cases and different sets of features. A high curvature value is better than a low value since it implies higher sensitivity to small misregistration angle errors. Note that the resolution optimizing value of α are close to 1 only in three out of the 12 cases

	151	142	162	151/8	151/16	151/32
pixel	0.3/0.9	0.6/0.3	0.6/0.3			
tag	0.5/3.6	0.5/3.8	0.4/1.4			
spatial-tag	0.99/14.6	0.99/8.4	0.6/8.3			
ICA				0.7/4.1	0.7/3.9	0.99/7.7

Table 1: Numerator =optimal values of α and Denominator = maximum resolution of mutual α-information for registering various images (Cases 151, 142, 162) using various features (pixel, tag, spatial-tag, ICA). 151/8, 151/16, 151/32 correspond to ICA algorithm with 8, 16 and 32 basis elements run on case 151.

studied. Note also that all of the proposed feature based methods have significantly higher resolution than the standard single-pixel (column labeled "pixel" in Table I) registration method. Systematic methods for determining the optimal value of α is one of the aims of future work as is the issue of optimally combining the single-pixel, ICA and tag features.

These preliminary results are promising and there remain several important research issues that we will investigate during this grant period. These issues include: selection of features, feature organization in the feature tree, selection of alpha parameter, numerical optimization of the mutual alpha-information, validation and comparison by simulation and experiment.

5 Research Plan

We will develop and test our new approaches to registration to increase accuracy and robustness for 3D ultrasound image registration. Our research plan will be comprised of several components: 1) investigation of different types of features for enhancing the capability of the feature tree to discriminate misregistered images in the presence of speckle noise, shadowing and distortions; 2) investigation of optimum choices of alpha in the alpha-information objective function to sharpen the registration peak and facilitate numerical optimization; 3) exploration and compensation for failure modes of the registration algorithm by controlled studies of robustness of the algorithm to increasing amounts of distortion; 4) development of fast minimal graph implementations using hierarchies of minimal spanning trees; 5) validation of results with a database of real medical ultrasound images. All of these tasks will be guided by a database of unannotated ultrasound breast images that have been previously acquired from 80 subjects. We will also consider validating these techniques on muscolo-skeletal scans, e.g. achilles tendon, if similar 3D volume data becomes available.

Paul: need your help here and in valdiation section!

1. Feature Selection

We will need to extend the feature selection algorithms to 3D volumes using ICA and feature tree tags in order to eliminate out of plane distortion effects inherent to 2D registration methods. We will also extend these techniques to include 3D color power mode Doppler image information as well as gray scale imaging. Spatial relationships between tags will be explored that can detect anomolous regions or shadows, and these will be included into the feature tree for additional robustness. Combinations of single-pixels, tags, and ICA features will also be investigated. We will implement randomization, bagging and boosting methods using a subset of 24 of the 80 subjects and evaluate improvements in registration performance on 24 image volumes randomly selected from the rest of the 80 subjects. The 24 subjects chosen for training set will be representative of the eight different classes of breast tissue complexity. Initially, to gauge pre-screened registration performance of our new method we will generate and apply separate sets of features for each class. However, we will also investigate universal sets of features which

can automatically screen out false breast classes during the registration process.

2. Choice of alpha parameter

We have shown [11] that there exists a theoretically optimal value of alpha for detecting the peak of the mutual alpha-information trajectory over the set of transformations T_i. This theory also suggests that the optimal value of α is small when $f_{0,i}$ is close to $f_0 f_i$, i.e. when the registration is very coarse, while α tends towards 1 for nearly perfectly registered images. Thus we will investigate the idea varying α over the course of registration: a coarse registration regime for which α is small in the initial iterations and increases towards a fine regime in the final iterations. In between these extremes α might increase continuously or step through only a few values.

The choice of alpha also plays a role in controlling the number of maxima in the objective function. Maintaining a small number of local maxima will be important as exhaustive search is impractical for fully 3D registration having many degrees of freedom in the transformation sequence $\{T_i\}$ applied to the secondary image. With a local optimization the presence of many local maxima makes it necessary to reinitialize the mutual information maximization several times to make sure that convergence to a global maximum has occurred. The selection of α will be investigated through a combination of analysis, simulation, and experiment with real scans to try and identify what schedules of α yield the fastest convergence of the registration algorithm. We will use the Carver-Mead simplex algorithm and the thin plate spline deformation sequence used in MIAMI Fuse©.

3. Fast Minimal Spanning Tree Approximations

The plain vanilla minimal spanning tree has polynomial runtime complexity in the number of feature points n, e.g. it has $O(n^2)$ complexity in the plane. Recently approximate MST algorithms with nearly $O(n \log n)$ complexity have been implemented [6]. Even faster algorithms can be implemented by partitioning heuristics, like the heuristic we recently introduced in [12]. Initial results are quite promising on a hierarchical scheme where the MST is only run on the centroids of a k-means algorithm, yielding average complexity of $O(k \log k) + O(n \log(n/k))$. We will investigate optimal choice of k to minimize runtime yet maintain adequate registration performance. In general it is unclear to what extent approximation of the MST will affect registration accuracy. This important tradeoff between runtime and approximation error will be studied both experimentally and via analysis.

4. Sensitivity Studies and Robust Algorithms

We will investigate robust density estimation and k-point MST for rejection of spurrious features that degrade registration performance. We will also investigate feature types that can indicate the presence and location of anomalies which can be subsequently pruned from the feature tree to enhance registration accuracy. This study will entail identifying the failure modes of the registration algorithm when presented with images having varying degrees of anatomical differences, e,g, due to tumors. We will insert such anomalous regions into speckle contaminated replicates of a single image volume. Performance will be evaluated by both registration accuracy, i.e. distance of registration function peak from true rotation/translation/defoemation, and by post-registration constrast of the anomolous region determined from the difference between the registered primary ad secondary images.

5. Validation and comparison by simulation and experiment

We will develop techniques to modify a copy of an image enough to make registration nontrivial while allowing rigorous truth data on registration success. The developed tags, feature sets and tree structures will be evaluated in comparison with the existing single-pixel techniques for accuracy and robustness on 2D data from 24 subjects not used for training of the registration methods. This comparison will be done with the new methods alone and in combination with the single pixel techniques. Use of the best combination of features plus single voxel techniques at the optimized value of α will be compared

with single voxel alone at $\alpha = 1$ to establish superiority of our methods on a real database. After integration of the new methods into a module of MIAMI Fuse© and development of 3D tags, features and tree structures, this comparative validation on the 24 subjects will be repeated with registration of nonlinearly-warped 3D gray scale and color flow volumes. Initially, successful registration will be declared if the algorithm produces a difference image with mean-squared-error less than 1% of maximum constrast. It is too early to estimate the variances to calculate the statistical power of this test but we will evaluate experimental p-values to establish statistical significance. We will then modify the registration criterion by accounting for spatially localized displacement errors which can be calculated for thousands of points in each scan volume. In the event that high computational complexity prevented whole volume registration we will reduce the study to registering half of the total image volume or less, either through truncation or downsampling of the original whole volume image. Final registration displacement error as a function of the eight classes of breast tissue complexity will be plotted as a function of breast class and we will evaluate the slope of the fit by confidence limits on that slope.

References

[1] I. Ahmad and P.-E. Lin, "A nonparametric estimation of the entropy for absolutely continuous distributions," *IEEE Trans. on Inform. Theory*, vol. IT-22, pp. 664–668, 1976.

[2] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Computation*, vol. 9, pp. 1545–1588, 1997.

[3] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. van der Meulen, "Nonparametric entropy estimation: an overview," *Intern. J. Math. Stat. Sci.*, vol. 6, no. 1, pp. 17–39, june 1997.

[4] L. Brieman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.

[5] D. L. Donoho, "One-sided inference about functionals of a density," *Annals of Statistics*, vol. 16, pp. 1390–1420, 1988.

[6] D. Eppstein, "Faster geometric k-point MST approximation," Technical Report 95-13, Dept. of Information and Computer Science, University of California, Irvine, March 1995.

[7] Y. Freund and R. Schapire, "A decision theoretic generalization of online learning and an application to boosting," *Journ. of Computer and System Sciences*, vol. 55, pp. 119–139, 1997.

[8] D. Geman and K. A, "Invariant statistics and coding of natural microimages," in *IEEE Workshop on Statist. Computat. Theories of Vision*, Fort Collins, CO, June 1999.

[9] P. Hall and S. C. Morton, "On the estimation of entropy," *Ann. Inst. Statist. Math.*, vol. 45, pp. 69–88, 1993.

[10] Y. He, A. Ben-Hamza, and H. Krim, "An information divergence measure for ISAR image registration," *Signal Processing*, Submitted, 2001.

[11] A. O. Hero, B. Ma, and O. Michel, "Alpha-divergence for classification, indexing and retrieval," Technical Report 328, Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, May, 2001. `http://www.eecs.umich.edu/~hero/det_est.html`.

[12] A. Hero and B. Ma, "Convergence rates of minimal graphs with random vertices," *IEEE Trans. on Inform. Theory*, vol. submitted, , 2001. `http://www.eecs.umich.edu/~hero/det_est.html`.

[13] A. Hero, B. Ma, and O. Michel, "Convergence rates of random minimal graphs," *IEEE Trans. on Inform. Theory*, Submitted Aug. 2001.

[14] A. Hero and O. Michel, "Robust estimation of point process intensity features using K-minimal spanning trees," in *Proc. of IEEE Symposium on Information Theory*, Ulm, Germany, July 1997.

[15] A. Hero and O. Michel, "Robust entropy estimation strategies based on edge weighted random graphs," in *Proc. of Meeting of Intl. Soc. for Optical Engin. (SPIE)*, San Diego, CA, July 1998.

[16] A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal k-point random graphs," *IEEE Trans. on Inform. Theory*, vol. IT-45, no. 6, pp. 1921–1939, Sept. 1999.

[17] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 1999.

[18] H. Joe, "On the estimation of entropy and other functionals of a multivariate density," *Ann. Inst. Statist. Math.*, vol. 41, pp. 683–697, 1989.

[19] Krücker, J.F., C. Meyer, G. LeCarpentier, J. Fowlkes, and P. Carson, "3d spatial compounding of ultrasound images using image-based nonrigid registration," *Ultrasound in Medicine and Biology*, vol. 26, no. 9, pp. 1475–1488, 2000.

[20] 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.

[21] B. Ma, *Parametric and non-parametric approaches for multisensor data fusion*, PhD thesis, University of Michigan, Ann Arbor, MI 48109-2122, 2001.

[22] B. Ma, A. O. Hero, J. Gorman, and O. Michel, "Image registration with minimal spanning tree algorithm," in *IEEE Int. Conf. on Image Processing*, Vancouver, CA, October 2000.

[23] C. Meyer, J. Boes, B. Kim, P. Bland, G. LeCarpentier, J. Fowlkes, M. Roubidoux, and P. Carson, "Semiautomatic registration of volumetric ultrasound scans," *Ultrasound in Medicine and Biology*, vol. 25, no. 3, pp. 339–347, 1999.

[24] C. Meyer, J. Boes, B. Kim, P. Bland, K. Zasadny, P. Kison, K. Koral, K. Frey, and R. Wahl, "Demonstration of accuracy and clinical versatility of mutual information for automatic multmodality image fusion using affine and thin-plate spline warped geometric deformations," *Medical Image Analysis*, vol. 1, pp. 195–206, 1996.

[25] O. Michel, R. Baraniuk, and P. Flandrin, "Time-frequency based distance and divergence measures," in *IEEE International time-frequency and Time-Scale Analysis Symposium*, pp. 64–67, Oct 1994.

[26] A. Mokkadem, "Estimation of the entropy and information of absolutely continuous random variables," *IEEE Trans. on Inform. Theory*, vol. IT-35, no. 1, pp. 193–196, 1989.

[27] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: a strategy employed by vi?," *Vision Research*, vol. 37, pp. 3311–3325, 1997.

[28] J. P. Pluim, J. B. A. Maintz, and M. A. Viergever, "f-information measures in medical image registration," in *Proc. of Meeting of Intl. Soc. for Optical Engin. (SPIE)*, M. Sonka and K. M. Hanson, editors, volume 4322, pp. 579–587, 2001.

[29] R. Rosenberg, W. Hunt, M. Williamson, F. Gillil, , P. Wiest, C. Kelsey, C. Key, and M. Linver, "Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: Review of 183,134 screening mammograms in albuquerque," *Radiology*, vol. 209, no. 2, pp. 511–518, 1998.

[30] O. Vasicek, "A test for normality based on sample entropy," *J. Royal Statistical Society, Ser. B*, vol. 38, pp. 54–59, 1976.

[31] 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.