Image registration using alpha-entropy measures and entropic graphs

Huzefa Neemuchwala[*#], Alfred Hero[*†+], and Paul Carson[*#]
Dept. of Biomedical Engineering[*], Dept. of EECS[†], Dept. of Statistics[+], and Dept. of Radiology[#]
The University of Michigan Ann Arbor, MI 48109, USA

June 2002

Abstract

Registration of a reference image to a secondary image extracted from a database of transformed exemplars constitutes an important image retrieval and indexing application. Two important problems are: specification of a general class of discriminatory image features and an appropriate similarity measure to rank the closeness of the query to the database. In this paper we propose a solution using high dimensional image features, which can be either continuous or discrete valued, and a general class of feature similarity measures based on Rényi's α-entropy function. This class of measures contains the well known mutual information measure, its mutual α-information (α-MI) variants, and the α-Jensen difference. When the features are discrete valued, the α-MI can be estimated from the joint feature histogram constructed from the reference and the secondary images using a data structure called a feature coincidence tree. However, histogram estimation techniques become impractical for continuous valued features in high dimension. For such features we propose an alternative similarity measure: the α-Jensen difference which can be accurately estimated using an entropic-graph estimator such as the minimal spanning tree (MST). A low time-memory complexity MST is used to compare a variety of continuous and discrete features including single pixel gray levels, tag subimages, and independent component analysis (ICA) coefficient vectors. The methodology is illustrated for ultrasound breast image registration for which we find that the best small angle registration performance is attained by implementing minimal graph entropy estimators on a set of ICA feature vectors.

Keywords: minimal spanning tree entropy estimators, feature coincidence trees, ultrasound breast image registration

[1]Corresponding author: Prof. Alfred Hero, 4229 EECS, 1301 Beal St., Ann Arbor, MI 48109-2122 USA. Tel: 734-763-0564. FAX: 734-763-8041. email: hero@eecs.umich.edu. This work was supported in part by PHS grant P01 CA85878 and in part by a Biomedical Engineering Fellowship to the first author.

1 Introduction

The image registration problem falls in the general area of indexing and retrieval over databases of images $\mathcal{X} = \{X_i\}_{i=1}^K$ for the purposes of finding the best match to a reference image X_0. A comprehensive review of many techniques in image retrieval and indexing is presented in [49]. The best match is expressed as a partial re-indexing of the database in decreasing order of similarity to the reference image using an index function. In the context of image registration the database corresponds to a set of transformed versions of the secondary image, e.g. rotation and translation, which are compared to the reference image. There are three key ingredients to image retrieval and indexing which impact accuracy and computational efficiency [12]:

1. Selection of image features that discriminate between different image classes yet possess invariance to unimportant attributes of the images e.g. rigid translation, rotation and scale;

2. Application of an index function that quantifies feature similarity, is capable of resolving important differences between images, yet is robust to image perturbations;

3. Query processing and optimization techniques which allow fast implementation of search for best matching.

This paper is concerned with the appropriate choice of features and the selection of the feature similarity measure.

The focus application of this paper is the co-registration of a pair of ultrasound (US) images of the breast. Success in this application should lead to full 3D and 4D registration: the fourth dimension being time. Accurate registration of 3D breast US image volumes could be the breakthrough to address efficient use of whole breast imaging to detect and quantify changes that: are indicative of a breast lesion; can aid discrimination of malignant from benign lesions [4, 52]; can be used to detect multifocal secondary masses [9] and can quantify response to chemotherapy or radiation therapy [21]. Breast lesions are missed by community practitioners in up to 45% of women with dense breasts [23]. Improved image registration could also improve spatial compounding of US images. In compounding, partly correlated views of the region of interest (ROI) are generated by scanning the ROI at different transducer tilt angles and then registering pairs of separate views. Compounding can result in an improvement in the signal-to-speckle ratio of US

images and lead to better delineation of specular reflectors [24]. The field of view of high resolution US images is insufficient for full use of 3D US in detecting asymptomatic lesions in the breast and for tracking changes in response to treatment. To create an image of the entire breast or a large fraction of it, the small volume covered by a single scan can be extended by repeatedly scanning the breast in parallel, partially overlapping sweeps that can then be combined using registration techniques [25]. Finally, registration of images collected from different isonation angles can also be used in Doppler imaging where the color flow acquisitions do not detect blood flow well when its direction of motion is normal to the direction of the ultrasound beam and where accurate triangulation can measure flow velocity.

To date the most effective methods for image retrieval and image registration have been pixel and voxel based and include: color histogram matching [19], texture matching [2], intensity cross correlation [34], optical flow matching [28] and, mutual information (MI) registration [56]. MI registration maximizes the MI maximization on pixel coincidence histograms and is used in the MIAMI-Fuse© registration algorithm at the University of Michigan Medical Center [37]. A general survey of medical image registration methods is presented in [34] and voxel specific methods are the focus of [18]. Color indexing was first proposed by Swain and Ballard [53] and variations of this technique have appeared in [10, 19, 50]. A histogram intersection formula was presented which, under most conditions, reduces the influence of background pixels on the sharpness of the registration peak. This approach allows the comparison of known objects based upon the colors of interest. The pattern matching techniques described in our paper are applicable to multichannel (color) imagery. Texture-based image retrieval using statistical methods [2, 44] work well for images with uniform scale and orientation, while spectral image features like Gabor filters as used in [29, 32, 7] and DCT [54, 55] work well for natural imagery. However, US images are noisy and often show shadow and enhancement and degradation effects. Independent component analysis coefficients, used in this study, are tailored to the US image domain and are thus more robust to noise and other artifacts. Various other image registration measures, such as sum of squared differences and correlation coefficients (see [34, 18]), have been utilized for registration using single pixel features. However, pixel-specific methods and measures often succumb to various artifacts like spurious large intensity differences, non-uniform illumination, saturation, occlusion and require linear relationships between the intensities of reference and secondary images. The method of mutual information (MI) registration has largely overcome many of these difficulties. Mutual

information provides an information theoretic approach to image registration and has been found to be the most robust and accurate image registration algorithm for a wide variety of modalities [18].

Ultrasonic image registration presents special challenges due to the fact that rigid body registration techniques, developed initially for magnetic resonance (MR) and computed tomography(CT) images (e.g. of the human brain) cannot be applied directly to the high noise images from non-linearly compressible organs, such as the bladder, breast, heart, and in images that contain speckle, nonlinear (shear) distortions and refraction artifacts. A manual method which utilized corresponding points, lines and planes was investigated by Moskalik *et al* [39] in the first effort for 3D US image registration. Rohling [46], first used voxel intensity based techniques for ultrasound image registration in 3D compounding. In his technique the correlation ratio was used in conjunction with a 3D gradient measure [35]. A similar bivariate approach was implemented for cross modality US-MR registration by Roche *et al* [45]. Both cases worked reasonably well for clean US images. Meyer *et al* [36] demonstrated that their mutual information based software called "MIAMI Fuse" developed for multimodality image fusion [22, 37] is capable of registering reasonably difficult 3D US images. MIAMI Fuse was also used in 3D US image compounding [24], for registration of multimode and extended-volume US images and to track changes of tissue structure and vascularity of serial US scans obtained months apart [25].

Despite the initial success reported on US images, pixel- or voxel- based registration methods have been disappointing in application to certain clinical cases. Ultrasonic images are highly sensitive to transducer orientation during scanning, as the attenuation artifacts often present themselves differently in the secondary and reference images [36]. Further inconsistency in tissue structure and geometry might arise from coherent echoes and shadowing effects. Partial or complete shadows, caused by objects near the probe, are a hindrance to registration during compounding. Dense tissue such as ribs (in breast imaging) and malignant calcification cause shadows, which present themselves in different directions in the reference and secondary images. The problem is further complicated by intensity similarity of shadows, cysts and tumors which makes intensity-based discrimination difficult. Ultrasonic images typically show blurred boundaries around anatomical features. These boundaries have varying intensity due to changes in surface curvature and transducer orientation. This effect occurs because US images exhibit a strong angular dependency of the apparent brightness of specular reflectors. They also have a smaller signal-to-noise dynamic range than magnetic res-

onance (MR) and computed tomography (CT) images. Furthermore, a small field of view (FOV) and large motion and deformation of the compliant breast tissue makes it difficult to obtain consistent image pairs for registration under manual scanning. Due to these factors developing robust registration algorithms is extremely challenging [25]. Without modification, single-pixel based MI registration methods based are unlikely to succeed in US registration. Our experience has shown that MI suffers severe misalignment errors in those image volumes with major shadows and obtained at different beam orientations. This paper describes a method to improve upon the accuracy of single-pixel-based MI registration using higher order features and better feature matching criteria.

A recent study by Shekhar and Zagrodsky [48] delves into 3D US image registration for the human heart using MI-based registration. The study is interesting because US images of a non-linearly deformable organ are registered. They show that the MI criterion, on raw US images, produces a rippled surface for 2D translations of the image and emphasize that the ripples confound the search for the maximum corresponding to the solution, and that the removal of undesired local maxima in the MI function is key to making optimization robust and reliable. Median filtering for speckle removal leads to marginal improvement in registration accuracy, though smoothing of the mutual information function is observed. They also suggest Partial Volume interpolation [33] and image requantization for smoothing. The method proposed here is an alternative way to improve the behavior of the objective function.

In this paper we extend and improve upon MI registration technology. The key to our approach is the inclusion of highly specific image features and use of a generalized information divergence matching criterion related to the Chernoff bound of detection theory. An innovation is the introduction of a general pattern matching criterion which can be used to register high dimensional image features such as curves, edges, textures and spatial relations. The criterion can be adapted to both discrete valued and continuous valued feature vectors selected using a representative training set of images. In the case of discrete features we can compute the joint coincidence histogram of features of a given type occurring at the same spatial location in both the reference and the transformed exemplar (secondary) images. This histogram is constructed using a efficient hierarchical database, which we call a feature coincidence tree, and its spread, as measured by the joint entropy, is an indicator of the degree of misregistration.

As in standard MI registration, to register the secondary to the reference a sequence of transformations is

applied to the secondary until the spread of the coincidence histogram is minimized. However, unlike standard MI registration methods, we replace the MI similarity function with the more general class of α-entropy functions, where α is a parameter in the interval $[0, 1]$. This class includes various information divergence measures such as: the mutual α-information (α-MI), the α divergence, and the α-Jensen difference. The combination of higher order features and the α-entropy similarity functions specifies a large and flexible set of registration algorithms. As the purpose of this paper is to introduce this powerful feature-similarity combination we do not consider the important problem of practical implementation of the sequence of registration transformations which will be the subject of a future paper.

Our approach specializes to the standard mutual information algorithm (e.g. MIAMI-Fuse©) when the features are the set of single-pixel gray levels and $\alpha \to 1$. The advantages of our approach are: 1) use of the generalized mutual α-information can lead to a more stable objective function; and 2) use of higher order features can capture non-local spatial information which is ignored in the standard single pixel MIAMI-Fuse© algorithm and can lead to more accurate and robust image registration.

We investigate two types of discrete features, called tags, to populate the feature coincidence tree. One of these is based on gray-scale adaptive thresholding and the other on independent component analysis (ICA). Gray scale thresholding is a fast and simple adaptive quantization scheme proposed by Geman and Koloydenko [8]. ICA is an iterative method which is closely related to the projection pursuit technique of non-linear regression and has been applied to image analysis by Olshausen, Hyvärinen and others [30, 20]. The continuous valued ICA coefficients, obtained by projecting the image onto its ICA bases, are discretized by vector quantization, implemented by partitioning the coefficient vector into a finite number of (Voronoi) cells.

When the number of dimensions of the feature space increases beyond 10 or so histogram methods become impractical due to the curse of dimensionality: for fixed resolution per dimension the number of histogram bins increases geometrically in dimension. As high dimensional feature spaces can be more discriminatory this creates a barrier to performing histogram-based registration. We circumvent this barrier by applying a novel technique for estimating the α-entropy using entropic graphs whose vertices are the locations of the feature vectors in feature space. As introduced in [16] and studied in [12, 13, 15, 17] an entropic graph is any graph whose normalized total weight (sum of the edge lengths) is a consistent

estimator of α-entropy. An example of an entropic graph is the minimal spanning tree (MST) and due to its low computational complexity it is the focus of this image registration application. The MST is constructed over the set of joint feature vectors from the secondary and reference image to yield an estimate of Jensen's entropy difference. We compare and contrast this approach with histogram methods for the US registration application.

All of the comparisons are evaluated through a combination of simulations and real data acquired from clinical studies. The database used in this paper was a set of 3D ultrasound scans of the left or right breast of 21 female subjects, aged 21-49 years, going to biopsy for possible breast cancer. The lower age range was chosen to provide a sample of more complex breasts, which are also somewhat more difficult to diagnose than typical breasts of older women. Fig 1 shows slices of breast ultrasound image volumes representative of those found in clinical practice. The women were imaged on their backs with the transducer imaging through the breast toward the chest wall. Three test cases chosen from the breast database and referred to as Case 151, Case 142 and Case 162 are presented. The image slice chosen from Case 151 exhibits significant connective tissue structure as the bright thin lines or edges. Case 142 was diagnosed as a malignant tumor in echogenic fibroglandular tissues. The tumor characteristically shows discontinuous edges with a darker center and shadows below the borders. The area of enhancement below the tumor is not common. Case 162 shows an uncommon degree of degradation due to shadowing. The bottom two-thirds of the image include the chest wall and the dark shadow and reverberations behind the acoustically impenetrable boundary between the lung and chest wall. Some edge information is evident, however shadowy streaks are observed due to dense tissue absorbing the sound beam, refraction and phase correlation at oblique boundary or poor acoustic impedance match (air bubbles) between the transducer and the skin. For clarity of presentation we focus on registration of 2D slices. The extension to 3D voxel registration is straightforward but will be presented elsewhere.

The outline of this paper is as follows. Section 2 introduces α-mutual information as a similarity measure. Section 3 presents feature coincidence trees for discrete features. Methods of tag selection using adaptive thresholding are presented in Section 4, while discrete and continuous ICA features are discussed in Section 5. In Section 6 the α-Jensen difference is discussed and in Section 7 the MST entropy estimator is presented along with methods of acceleration. A simple numerical example is presented in Section 8

and in Section 9 presents registration results for a ultrasonic breast image database. Conclusions and future directions are described in Section 10.

2 Indexing with Mutual α-Information as a Similarity Measure

Let X_0 be a the reference image and consider a database X_i, $i = 1, \ldots, K$ of images generated from a database of secondary images to be indexed relative to the reference. Let Z_i, Z_j be feature vectors extracted from X_i, X_j and define the joint density $f(Z_i, Z_j)$ and the marginal densities $f(Z_i)$, $f(Z_j)$. The similarity between features Z_0 and Z_i can be gauged by the difference between $f(Z_i, Z_0)$ and the product $f(Z_i)f(Z_0)$ which can be measured by the mutual information (MI).

When the features are discrete the joint probability distributions $f(Z_i, Z_0)$ of reference and secondary images can be plotted on a two-dimensional histogram of intensity values (Fig 2). Fig. 3 shows four joint histogram scatter plots of a pair of reference (X_i) and secondary (X_j) ultrasound images extracted from a breast US volume. It can be seen that when the reference and secondary images are the same ($X_j = X_i$) except for rotation, the joint histogram shows perfect correlation of intensity values in pixel coordinates [a]. When the reference and secondary images are mis-aligned by 8 degrees, the joint histogram is dispersed [b]. However, consider the case where the reference and secondary images are extracted from two slices in the same breast US volume where the slices are separated by 2mm away from each other. At this separation distance along the depth of the scan, the speckle in the images is decorrelated, but the anatomy in the images remains largely unchanged. Now, the histogram scatter plot [c] shows a dispersion similar to the one seen in [b], in spite of perfect alignment. In case of perfectly correlated image slices a simple registration objective function which might be applied is the spread of the two dimensional histogram in (b) about a fixed reference, such as the diagonal shown in [c]. However, ultrasound images obtained at different scan angles show different speckle patterns. Hence plots [c] and [d] are observed more often in practice and such a simple reference-based objective function is not effective. A more effective objective function might be the entropy of the joint histograms which measures the spread of the joint distribution independent of any fixed reference. The mutual information is a generalization of this entropy measure which measures the

spread relative to the maximally spread product of the marginal histograms:

$$\sum_{z_i,z_0} f(z_i,z_0)\log(f(z_i,z_0)/[f(z_i)f(z_0)]) \tag{1}$$

which is used by Viola and Wells [56] and others for image registration. Note that this is a different usage from the KL divergence between $f(Z_i)$ and $f(Z_0)$ which has been proposed as an indexing measure by several authors [51, 6, 54].

The mutual α-information, a generalization of (1), is defined as the α-divergence of fractional order $\alpha \in [0,1]$ between $f(Z_i,Z_0)$ and $f(Z_i)f(Z_0)$ [3]. For discrete valued features the α-MI is:

$$D_\alpha(f(Z_i,Z_0) \parallel f(Z_i)f(Z_0)) = \frac{1}{\alpha-1}\log\sum_{z_i,z_0} f^\alpha(z_i,z_0)f^{1-\alpha}(z_i)f^{1-\alpha}(z_0) \tag{2}$$

where the summation is over all values that Z_0 and Z_i can take on. For continuous valued features the α-MI is:

$$D_\alpha(f(Z_i,Z_0) \parallel f(Z_i)f(Z_0)) = \frac{1}{\alpha-1}\log\int_{z_i,z_0} f^\alpha(z_i,z_0)f^{1-\alpha}(z_i)f^{1-\alpha}(z_0) \tag{3}$$

The α-divergence is equal to the Hellinger distance squared when $\alpha = 1/2$, and to the Kullback-Liebler (KL) divergence [27] when $\alpha \to 1$. The case $\alpha \to 1$ corresponds to the standard Shannon mutual information (1).

The mutual α-information can be justified as an appropriate registration function by large deviations theory through the Chernoff bound. Define the average probability of error $P_e(n)$ associated with deciding whether Z_i and Z_0 are dependent or independent random variables. i.e. deciding between dependent features, $H_1 : Z_i, Z_0 \sim f(z_i,z_0)$ vs. dependent features $H_0 : Z_i, Z_0 \sim f(z_i)f(z_0)$, based on a set of i.i.d. samples $Z_j^{(1)},\ldots,Z_j^{(n)}$, $j = 0, i$. This error probability has the representation:

$$P_e(n) = \beta(n)P(H_1) + \alpha(n)P(H_0),$$

where $\beta_(n)$ and $\alpha(n)$ are Type II (miss) and Type I (false alarm) errors, respectively, of the test of H_0 vs. H_1. Then the Chernoff bound implies that [5]:

$$\liminf_{n\to\infty} \frac{1}{n} \log P_e(n) = - \sup_{\alpha\in[0,1]} \{(1-\alpha) D_\alpha(f(Z_i, Z_0) \parallel f(Z_i) f(Z_0))\} \tag{4}$$

Thus the mutual α-information gives the asymptotically optimal rate of exponential decay of the error probability for testing H_0 vs H_1 as a function of n. While we do not investigate this issue in this paper, the appearance of the maximum over α suggests the existence of an optimal parameter α ensuring low error registration.

3 Feature Coincidence Trees for Discrete Features

To perform MI registration with discrete features we must populate the bins of the joint histogram $\{f(z_i, z_0)\}$ for each image in the database. First, a universal set of features is selected according to certain criteria discussed below. These features are organized into bins on a tree-structured database for which the complexity of the feature indexed at each node increases as tree depth increases. Figures 4 and 5 illustrate the feature tree data structure when the features are defines as 4×4 sub-images. The two images are each dropped down a feature tree and incidences and coincidences of features over all pixel locations and 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. See Fig 6 for illustration. This results in a histogram called the feature coincidence histogram denoted $\hat{f}(Z_1, Z_2)$. The histogram marginals $\hat{f}(Z_1)$ and $\hat{f}(Z_2)$ of the coincidence histogram are extracted by summing over one of the arguments of $f(Z_1, Z_2)$. These are then used in the mutual α-information formula (2) to come up with a registration score for the images.

Our general feature selection and organization scheme is similar to the randomized tree classifier structures introduced by Amit and Geman [1] and used for shape recognition from binary transcriptions of handwriting. 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. 8×8 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.

This allows 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 gray scale images, the number of different tag types can be extremely large. For example, if the image contains 256 gray values then there would exist $(256)^6 4$ different 8×8 tag types. These would be useless as feature discriminants since their individual occurrences would be extremely low. Methods for pruning the tag types are described below.

4 Tag Selection via Adaptive Thresholding

Adaptive thresholding is a quantization scheme 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 8×8 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 breast 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. Figure 6 shows tag coincidences in the reference and secondary images. Coincidences of tag types are calculated using the feature tree. The exaggerated tag pattern is meant to capture the edge of the tumor. A similar tag type will be observed in the secondary image also. The tags capture the local intensity pattern in the neighborhood of the pixel.

As mentioned earlier, the tag feature space is potentially large and requires pruning. We select relevant tag types by a process of randomized training. From within the US breast image database a large number of pixel neighborhoods are selected and quantized. After discarding repetitions, tag types that correspond to spurious speckle pattern are identified and eliminated. Speckle, after quantization using the above scheme, can easily be identified. since it exhibits only one or two non-zero pixels amongst the 8×8 pixels in the tag. 300 different tag types are identified among the samples. The number of tag types is controlled by selectively traversing the decision tree so that it is balanced, and by imposing constraints on the tag types. Also, tags are required to have at least two different intensity types within the center pixels so that

processing is centered near the edges in the image. Each image is then block-quantized and dropped down the partition tree. The pairwise coincidences of the tag types at the leaves are recorded in a histogram over 'tag space'. If an image block does not correspond to any of the tag types, then the pixel is not used in the image registration algorithm. While we do not explore it here, these tag features can be extended to account for spatial dependencies between pairs of tag types [1].

5 Continuous and Discrete Feature Selection via ICA

5.1 Continuous ICA Features

Images can be described in terms of their projection onto a basis, represented as a linear superposition of bases functions. Such a projection feature set was investigated by Vasconcelos [55, 54] for general image indexing problems using a Gabor wavelet basis. Here we take a different approach adopting a basis extracted from an independent components analysis (ICA) of the image database. The ICA basis can be adapted so as to best account for the image structure in terms of a collection of statistically independent components. This is an alternative to adaptive thresholding and is called the ICA approach in [30]. In ICA, an optimal basis is found which decomposes the ultrasound image X_i into a small number of approximately statistically independent components $\{S_j\}$:

$$X_i = \sum_{j=1}^{p} a_{ij} S_j \tag{5}$$

The basis elements $\{S_j\}$ are selected from an over-complete linearly dependent basis using randomized selection over the training set of ultrasound images in a representative database. The number of basis elements are selected according to the minimum description length (MDL) criterion. For image i the feature vectors Z_i are defined as the coefficients $\{a_{ij}\}$ in 5 obtained by projecting the image onto the basis. Here ICA was implemented using Hyvarinen and Oja's [20] `FastICA` code (available from `http://www.cis.hut.fi/projects/ica/fastica/`) which uses a fixed-point algorithm to perform maximum likelihood estimation of the basis elements in the ICA data model (5). Figure 7 shows a set of 40 8×8 basis vectors which were learned from over 5000 8×8 training subimages randomly selected from 10 consecutive image slices of a single ultrasound volume scan of the breast (Case 151 in Fig. 1).

Given this ICA basis and a pair of to-be-registered $N \times M$ images, coefficient vectors are extracted by projecting each 8×8 neighborhood in the images onto the basis set. For the 40 dimensional ICA basis shown in Fig. 7 this yields a set of NM vectors in a 40 dimensional vector space which will be used to define features.

5.2 Discrete ICA Features

Recall that in the feature tree method we generate histogram in higher dimensional space by labeling pixels with tag types. To extend this labeling technique to ICA feature vectors, discretization of the ICA coefficient vectors is essential. We do this by vector quantization: we partition the space of the multi-dimensional ICA coefficients into a finite number of Voronoi cells. The Voronoi partitions are created using the K-Means clustering algorithm on features extracted from the image training set. This algorithms creates a partition which minimizes the mean squared error of the Euclidian norm between the centroid of the Voronoi cell and the feature vectors in the cell. If n Voronoi cells are created then, each pixel p_i is given a label $0, 1, \ldots, n$ based upon the ICA coefficient at that pixel neighborhood. 512 cells were created using 8 and 16 dimensional ICA bases. The same partitioning is used for the reference and the secondary images to maintain consistency in pixel labels.

6 α-Jensen Difference Function for Continuous Features

With continuous features the α-MI can be estimated from continuous densities, e.g., through histogram or kernel density estimators. However the performance of such "plug-in" estimates becomes very poor as the feature dimension increases [12]. An alternative, discussed here and in the next section, is a direct estimator of another entropy-based index function: the α-Jensen difference. This function has been independently proposed by Ma [14] and He *et al* [11] for image registration problems. Let f_0 and f_1 be two densities and let $\beta \in$ [0, 1] be a mixture parameter. It was also used by Michel *et al* in [38] for time-frequency images. The α-Jensen difference was defined in [3] is the difference between α-entropy of the mixture:

$$f = \beta f_0 + (1 - \beta) f_1 \tag{6}$$

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

$$\Delta H_\alpha(\beta, f_0, f_1) = H_\alpha(\beta f_0 + (1-\beta) f_1) - [\beta H_\alpha(f_0) + (1-\beta) H_\alpha(f_1)] \tag{7}$$

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 Jensen's 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 feature vectors $Z_0 = \{Z_0^{(i)}\}_{i=1,...,n_0}$ and $Z_1 = \{Z_1^{(i)}\}_{i=1,...,n_1}$ are extracted from images X_0 and X_1, respectively. Assume that each of these sets consist of independent realizations from densities f_0 andf_1 respectively. Define the union $Z = Z_0 \cup Z_1$ containing $n = n_0 + n_1$ unordered feature vectors. Any consistent entropy estimator constructed on the $Z^{(i)}$'s will converge to $H_\alpha(\beta f_0 + (1-\beta) f_1)$ as $n \to \infty$ where $\beta = \lim_{n\to\infty} n_0/n$. This motivates the following consistent minimal-graph estimator of Jensen difference for $\beta = n_0/n$:

$$\Delta \hat{H}_\alpha(\beta, f_0, f_1) = \hat{H}_\alpha(Z_0 \cup Z_1) - [\beta \hat{H}_\alpha(Z_0) + (1-\beta) \hat{H}_\alpha(Z_1)] \tag{8}$$

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 rigid registration 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 $\{\Delta H_\alpha(\beta, f_0, f_i)\}_{i=1}^K$

7 Minimum Spanning Tree and Rényi Entropy

Implementation of the α-Jensen registration criterion can be accomplished by plugging in the density estimates to (7) but it is better to use direct methods based on entropic graphs such as the Minimal Spanning Tree (MST). The MST method 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 $Z_n = \{z_1, z_2, ..., z_n\}$ 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:

$$L(Z_n) = \min_{e \in T} \sum_{e} \|e\|^{\gamma} \tag{9}$$

where $\|e\|$ denotes Euclidean (L2) norm of the edge. See Fig 8 for illustration. The overall length of the MST can be used to construct a strongly consistent estimator of entropy.

The MST length $L_n = L(Z_n)$ is plotted as a function of n in Fig. 9 for the case of i.i.d uniformly and non-uniformly distributed points in the plane for γ=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's spanning the uniformly distributed points. This fact motivates the application of MST to test randomness of a set of points. It can be shown [12] the length function when normalized by $\sqrt{n}$ produces sequences that converge within a constant factor to the alpha entropies with α=1/2, as illustrated in Fig. 9 . More generally, for i.i.d. points in $\Re^d$, by changing the value of γ in (9), one can change the convergent limit to the α-entropy where $\alpha = (d - \gamma/d)$. The MST is called an entropic spanning graph as its normalized log-length converges (a.s.) within a constant to an alpha-entropy. Specifically, the Rényi entropy estimator

$$\hat{H}_{\alpha}(Z_n) = 1/(1-\alpha)[lnL(Z_n)/n^{\alpha} - ln\beta_{L,\gamma}] \tag{10}$$

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 [12, 17]. In image registration, when two images are properly matched under a sequence of transformations, corresponding regions of interest should overlap and the resulting joint probability distribution will be highly concentrated. Thus the Rényi entropy of the overlapped images should achieve the minimum value over all of the transformations. This would be reflected as the smallest length of the MST [31]. A transformation that minimizes Rényi entropy can be calculated, since misregistration would increase the dispersion of the joint probability distribution.

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 [13, 15]; 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. Furthermore, for large dimensions the MST can be implemented when histograms cannot, due to the "curse of dimensionality". For example, for a 32 dimensional space, only 10 cells per dimension gives 10^{32} bins in the histogram, an unworkable and impractically large burden for any computer. On the other hand the need for combinatorial optimization may be a bottleneck for a large number of feature samples and accelerated MST algorithms are necessary.

7.1 Computational Acceleration of the Kruskal MST Algorithm

The Kruskal Algorithm [26] is widely believed to be the fastest general purpose algorithm to solve the MST problem for sparse graphs. A set of given edges, sorted by their weights, is maintained in a list and Kruskal's 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 $O(ElogE)$ where E is the initial number of edges in the graph. The memory requirement is $O(E)$.

In the present application, the most simple-minded initial estimate of the MST includes 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 $O(N^2logN)$. 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 $\times$ 256. Desktop processors cannot fulfill memory requirements of the standard Kruskal algorithm in this case. Even with larger machines, the algorithm has a foreboding 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. 11, 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. 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. 11). 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. 10 shows bias of the modified MST algorithm as a function of the radius parameter.

It is straightforward to prove that, if the radius is suitably specified, the disc based tree construction described above is a minimum spanning tree. Recall that the Kruskal algorithm ensures construction of the exact MST [26].

(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 led 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 be completed without first adding more edges to 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 varies with the dimensionality and the underlying density 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 $O(ElogE)$, where E is a fraction of N^2 for large N. Figure 11 compares the performance of the standard Kruskal algorithm with our modified algorithm. The time and memory requirements are tremendously reduced. To obtain the optimal trade-off between bias of the MST length and time-memory complexity,

the disc radius should be selected at the knee of the curve seen in Fig 10. For uniform distributions, the distance of the k^th-NN along the dimension of ordering, remains roughly the same. The real advantage of the kNN technique lies in adapting the disc radius on a point-by point basis for non-uniform distributions (Fig 12). However, one of the problems underlying non-uniform distribution can be seen there. The MST length estimate from the disc based algorithm converges slowly to the true length, requiring examination of a large number of neighbors in the process. The number of nearest neighbors requiring examination grows in higher dimensions and the search now approaches a full search, thus losing some of its scalability (Fig 12 b). To address this issue, we utilize a kNN approach to determine the radii of the disks, following a list intersection approach similar to [43]. Variants of the technique described above have been examined and are described in [41]; where we also present results and methods to deal with dimensionality issues for non-uniform distributions.

8 Numerical Illustration

In this simple example we compare the MST-based estimate of α-Jensen difference with ICA features against the standard histogram-based estimate of MI using single pixel features. The backgrounds in the images in Fig 13 have been constructed using a random mixture of 15 ICA bases elements. The aim is to register the dominant pattern seen in the foreground as it translates across the image on a fixed background. The problem is similar to tracking prominent features as they move in the image. The mutual α-information based on single pixels, for α = 0.5 and the α-Jensen divergence based on the 15 ICA features are calculated for each integer translation for the pair of images. Using the ICA technique we can systematically eliminate features arising from the background and compute the MST-based entropy estimate solely on the distribution of the prominent feature of interest. This is not possible with the single pixel method. Profiles in Fig. 13 show that as contrasted to the feature based αJensen divergence, the αMI has a lower discrimination ability evident from the curvature of the peak around perfect alignment (0 pixel translation). The example suggests that significant improvement in resolution can be achieved through the use of image specific features and entropic graph estimators.

9 Ultrasonic Breast Image Registration

Figure 14 show the representative profiles of the registration objective function for registering a slice of US breast image volume, labeled Case 142 in Fig. (1), to a slice 2mm deeper in the same image volume. At this separation distance, the speckle noise decorrelates. However the underlying anatomy remains approximately unchanged. As the aim of this study is to quantitatively compare different feature selection and registration similarity indices we restricted our investigation to rotation transformations over a small range (-8^o to $+8^o$) of angles. The panel on far left of Fig. 14 indicates that, for single pixel features, entropic-graph (MST) estimates of α-Jensen difference and histogram plug-in estimates of α-MI give similarity functions with virtually identical profiles with a unique global minimum at 0^o rotation. The profile of the histogram plug-in estimate of the α-Jensen difference for single pixels (not shown) is very similar to the α-MI profile. For 8 dimensional ICA feature vectors, chosen by training on Case 151, we observe that the profile of the histogram plug-in estimate of α-Jensen degrades, with the appearance of several local minima (middle panel). This is expected since histogram estimation becomes unstable in high dimensional feature space. In a 64 dimensional ICA feature space, again chosen by training on Case 151, we observe that the entropic-graph maintains a smooth profile with a single global minimum (right panel) while the histogram estimate of α-MI is not implementable.

We next investigated the effect of additive noise on small-angle registration performance. We tested registration accuracy for single-pixels, tags, and discrete and continuous ICA features using the α-MI, entropic-graph, and α-Jensen discriminating criteria under increasing noise conditions. Figures (15 and 16) show plots of registration root mean square (rms) error versus increasing levels of additive (truncated) Gaussian noise in the images. Shown on the plots are standard error bars. 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/MI estimation techniques implemented. The lack of smoothness in the plots is likely due to the image-specific nature of the simulation - averaging results from registering different types of breast images would undoubtedly produce smoother graphs.

Figure 15 shows a comparison of the histogram and feature coincidence tree methods for tag features. We see that tag features have lower MSE than the standard single pixel features. Also higher order ICA fea-

tures perform better than single pixels. Notice also that increasing the dimensionality of the ICA coefficient makes the registration more robust, by further lowering misregistration error. Figure 16 shows a comparison of the α-MI and $alpha$-Jensen discrimination criteria. The α-Jensen difference function is calculated for single pixels and 8D ICA coefficients using higher order histograms and the MST entropic-graph estimate. It also shows the single pixel MI criterion under a range of SNR conditions similar to the ones explained above. The performance of higher order continuous ICA features estimated from the MST is seen to be better than those of single pixel features estimated from histograms. More extensive experiments are necessary but thus far our results indicate that the MST method of entropy estimation has significantly greater robustness to additive noise than histogram based methods.

10 Conclusions

In this paper we demonstrated that higher order features offer distinct advantages over standard single pixels for entropy-based US image registration. They can account for local spatial information between image features which is ignored by single-pixel features. We presented two techniques for generating image dependent higher order feature vectors. Discrete features such as tags and discrete ICA coefficients were used for image registration using a generalized α-MI matching criteria. Feature coincidence trees were presented to generate histograms in higher dimensional feature space. We concluded that tags and ICA features are more robust to noise artifacts and give lower misregistration errors than single-pixel based MI methods.

The α-Jensen difference function can be used to obtain a consistent estimate of entropy of continuous feature vectors using minimal graphs such as the MST. Computing the histogram estimate of α-MI is exponentially complex in higher dimensions, since the bias of the estimate increases in the number of dimensions of the feature space. The direct graph based MST estimate of α-Jensen difference avoids other problems that plague plug-in histogram estimates, which include ideal bin size and density kernel width selection and it also has a faster asymptotic convergence rate for non-smooth densities. Combinatorial optimization is a bottleneck in MST computation in large feature sets in high dimensions. We have presented a technique to accelerate MST computation and reduce its memory complexity so as to be able to use desktop processing for large datasets. Using the disc-based algorithm it was possible to rapidly compute the MST length for

data sets up to 1 million points in 8 dimensional space. The MST-based estimate of α-Jensen gave lower misregistration errors than discrete feature vectors computed using α-MI, and both techniques outperformed single-pixel MI methods.

Applying these retrieval and registration techniques to color images, such as registration of US Doppler blood flow color images with other color flow or gray scale images is a direction of future study. Another natural extension of this work is to incorporate spatial relations amongst pairs of spatially separated tags to eliminate the effect of shadows and other nonlinear artifacts which pose problems during compounding registration of ultrasonic images. We are actively pursuing other techniques to further accelerate the MST algorithm using parallel machines and list intersection. Optimal disc-radius selection has been found critical to complexity reduction in the disc-based MST construction algorithm. We are investigating other techniques to judge the optimal disc radius for MST construction. To increase the classification capabilities of the feature trees, different methods of randomized trees are being studied. Finally, the best value of α in the α-entropy similarity criteria is an open issue that needs further investigation.

11 Acknowledgments

To Sun Chung PhD, Computer Science Department, St. Thomas University for the valuable suggestions on acceleration of the MST algorithm.

References

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

[2] Ashley J, Barber R, Flickner M, Hafner JL, Lee D, Niblack W and Petkovic D, "Automatic and semi-automatic methods for image annotation and retrieval in QBIC", *Proc. SPIE Storage and Retrieval for Image and Video Databases III*, pp. 24-35, 1995.

[3] Basseville M,"Distance measures for signal processing and pattern recognition", *Signal Processing* , Vol. 18, pp. 349-369, 1989.

[4] Bhatti PT, LeCarpentier GL, Roubidoux MA, Fowlkes JB, Helvie MA, Carson P.L., “Discrimination of Sonographic Breast Lesions Using Frequency Shift Color Doppler Imaging, Age and Gray Scale Criteria”, *J. Ultrasound Med.*, vol. 20, no.4, pp. 343-350, 2001.

[5] Dembo A and Zeitouni O,“Large deviations techniques and applications”, *Springer-Verlag NY*, 1998.

[6] Do MN and Vetterli M,“Texture similarity measurement using Kullback-Liebler distance on wavelet subbands”, *Proc. IEEE International Conference on Image Processing, Vancouver, BC*, pp. 367-370, 2000.

[7] Dunn D, Higgins WE and Wakeley J, ”Texture segmentation using 2D Gabor elementary functions”, *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol 16, no. 2, pp. 130-149, 1994.

[8] Geman D and Koloydenko A,“Invariant statistics and coding of natural microimages”, *IEEE Workshop on Statis. Computat. Theories of Vision, Fort Collins, CO*, June 1999.

[9] Gordon PB and Goldenberg SL, “Malignant breast masses detected only by ultrasound”, *Cancer*,vol. 76, pp 626-630, 1995

[10] Hafner J, Sawhney HS, Equitz W, Flickner M and Niblack W, “Efficient color histogram indexing for quadratic form distance function”, *IEEE Trans Pattern Anal. Mach. Intelligence*, vol. 17, no.7, July 1995.

[11] He Y, Ben-Hamza A and Krim H, “An information divergence measure for ISAR image registration”, *Signal Processing*, submitted, 2001.

[12] Hero AO, Ma B, Michel O and Gorman JD, ”Applications of entropic spanning graphs,” to appear in *IEEE Signal Proc. Magazine (Special Issue on Mathematical Imaging)* Oct. 2002.

[13] Hero AO, Ma B, Michel O and Gorman JD, “Alpha-Divergence for Classification, Indexing and Retrieval”, *Technical Report CSPL-328 Communications and Signal Processing Laboratory*, The University of Michigan, 48109-2122, May 2001

[14] Hero AO, Ma B and Michel O, “Imaging applications of stochastic minimal graphs”, *Proc. of IEEE Int. Conf. on Image Proc., Thesaloniki, Greece*, Oct 2001.

[15] Hero AO, Costa J and Ma B, “Convergence rates of minimal graphs with random vertices”, (submitted to) *IEEE Trans. Information Theory*, 2001.

[16] Hero AO and Michel O, “Asymptotic theory of greedy approximations to minimal K-point random graphs”, *IEEE Trans. Information Theory*, Vol. IT-45, No. 6, pp. 1921-1939, Sept. 1999.

[17] Hero AO and Michel O, “Estimation of Rényi Information Divergence via Pruned Minimal Spanning Trees”,*1999 IEEE Workshop on Higher Order Statistics, Caesaria Israel*, 1999.

[18] Hill DLG, Batchelor PG, Holden M and Hawkes DJ, “Medical image registration”, *Phys. Med. Biol.*, vol 26, pp R1-R45, 2001.

[19] Huang J, Kumar SR, Mitra M, Zhu W, “Spatial Color Indexing and Applications”, *IEEE Int’l Conf. Computer Vision ICCV ‘98, Bombay, India*, pp 602-608, Jan. 1998.

[20] Hyvarinen A and Oja E, “Independent component analysis: algorithms and applications”, *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 1999.

[21] Kedar RP, Cosgrove DO, Smith IE, Mansi JL and Bamber JC, “Breast Carcinoma: Measurement of tumor response to primary medical therapy with color Doppler flow imaging”, *Radiology*, vol. 190, pp. 825-830, 1994.

[22] Kim B, Boes JL, Frey KA, Meyer CR, “Mutual information for automated unwrapping of rat brain autoradiographs”, *Neuroimage*, vol. 5, pp. 31-40, 1997.

[23] Kolb TM, Lichy J and Newhouse JH, “Occult cancer in women with dense breasts: Detection with screening ultrasound: Diagnostic yield and tumor characteristics”, *Radiology*, vol. 207, pp. 191-198, 1998

[24] Krücker JF, Meyer CR, LeCarpentier GL, Fowlkes JB, Carson PL, “3D Spatial Compounding of Ultrasound Images Using Image-Based, Nonrigid Registration”, *Ultrasound Med Biol*, vol. 26, no.9, pp. 1475-1488, 2000.

[25] Krücker JF, LeCarpentier GL, Meyer CR, Fowlkes JB, Roubidoux MA, Carson PL, “3D image registration for multimode, extended field of view, and sequential ultrasound imaging”, RSNA ej, 1999.

[26] Kruskal JB, “On the shortest subtree of a graph and the traveling salesman problem”, *Proc. American Math. Societ*, vol. 7, 48-50, 1956.

[27] Kullback S and Liebler R,“On information and sufficiency”, *Ann. Math. Statist.*, Vol. 22, pp. 79-86, 1951.

[28] Lefébure M and Cohen LD, “ Image Registration, optical flow and local rigidity”, *J. Mathematical Imaging and Vision*, vol. 14, no. 2, pp. 131-147, March 2001.

[29] Leow W, Lai S, “Scale and orientation-invariant texture matching for image retrieval”, in *Pietikainen MK ed., Texture Analysis in Machine Vision*, World Scientific, 2000.

[30] Lewicki M and Olshausen B, “Probabilistic framework for the adaptation and comparison of image codes”, *J. Opt. Soc. Am.*, vol. 16, no. 7, pp.. 1587-1601, 1999.

[31] Ma B, Hero AO, Gorman J and Michel O, “Image registration with minimal spanning tree algorithm”, *IEEE International Conf. on Image Processing, Vancouver*, Oct. 2000.

[32] Ma WY and Manjunath BS,“NETRA: A toolbox for navigating large image databases”, *Proc. IEEE International Conference on Image Processing, Santa Barbara, California*, Vol. I, pp. 568-571, Oct 1997.

[33] Maes F, Colignon A, Vandermeulen D, Marchal G and Suetens P, “Multimodality image registration by maximization of mutual information”, *IEEE Trans Medical Imaging*, vol. 16, pp 187-198, 1997.

[34] Maintz JBA and Viergever MA,“A survey of medical image registration”, *Medical Image Analysis*, vol. 2, no. 1, pp 1-36, 1998.

[35] Maintz JBA, van den Elsen PA, Viergever MA, “Comparison of edge-based and ridge based registration of CT and MR brain images”, *Med. Image Analysis*, vol. 1, pp 151-161, 1996.

[36] Meyer CR, Boes JL, Kim B, Bland PH, LeCarpentier GL, Fowlkes JB, Roubidoux MA, Carson PL, “Semiautomatic Registration of Volumetric Ultrasound Scans”, *Ultrasound Med. Biol.*, vol. 25, no.3, pp 339-347, 1999.

[37] Meyer CR, Boes JL, Kim B, et al,“Demonstration of accuracy and clinicla versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations”, *Med Image Analysis*, vol. 1, pp. 195-206, 1996/97.

[38] Michel O, Baraniuk R, and Flandrin P, “Time-frequency based distance and divergence measures”, *IEEE International Time-Frequency and Time-Scale Analysis Symposium*, pp. 64-67, Oct 1994.

[39] Moskalik A, Carson PL, Meyer CR, Fowlkes JB, Rubin JB, Roubidoux MA, “Registration of three-dimensional Compound Ultrasound Scans of the Breast for Refraction and Motion Correction”, *Ultrasound Med. Biol.*, vol. 21, no.6, pp 769-778, 1995.

[40] Neemuchwala HF, Hero AO, and Carson PL, “Image registration using entropic spanning graphs”, (to appear in) *Proc.of 36th Asilomar Conf Signals, Systems and Computers, Pacific Grove, CA*, Nov 2002.

[41] Neemuchwala HF, Hero AO, and Carson PL, “Fast algorithms for Minimum spanning tree construction”, (to apear in) *Technical Report CSPL Communications and Signal Processing Laboratory*, The University of Michigan, 48109-2122, 2002

[42] Neemuchwala HF, Hero AO, and Carson PL, “Feature coincidence trees for registration of ultrasound breast images”, *Proc.of IEEE Int. Conf. on Image Proc., Thesaloniki, Greece*, Oct 2001.

[43] Nene SA and Nayar SK, “A Simple Algorithm for Nearest Neighbor Search in High Dimensions”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19, 1997

[44] Pentland A, Picard W, Sclaroff S.,“Photobook: Content-Based Manipulation of Image Databases”, *Proc SPIE Storage and Retrieval for Image and Video Databases II*, no. 2185, 1994.

[45] Roche A, Xavier P, Malandain G, “Generalized correlation ratio for rigid registration of 3D ultraosund with MR images”, *IEEE Tran. Medical Imaging*, vol 20, no. 10, pp. 25-31, Oct 2001.

[46] Rohling R, Gee A, Berman L, “Automatic registration of 3D ultrasound images”, *Ultrasound Med. Biol.*, vol. 24, pp 769-778, 1995.

[47] Rudin LI and Yu P, “Improved forensic photogrammetric measurements with global geometrical constraints”, *Proc. SPIE*, vol. 4709, 2001.

[48] Shekhar R and Zagrodsky V, “Mutual information-based rigid and nonrigid registration of ultrasound volumes”, *IEEE Trans Medical Imaging*, vol. 21. no.1, pp 9-22, 2002.

[49] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R, “Content-based image retrieval at the end of the early days”, *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, 2000.

[50] Smith JR and Chang SF, “Automated image retrieval using color and texture”, *Columbia University Technical Report TR 414-95-20*, July 1995.

[51] Stoica R, Zerubia J and Francos JM,“Image retrieval and indexing: A hierarchical approach in computing distance between textured images”, *Proc. IEEE International Conference on Image Processing, Chicago*, 1998.

[52] Stavros AT, Thickman DI, Rapp CL, Dennis MA, Parker SH, Sisney GA, “Solid breast nodules: use of sonography to distinguish between benign and malignant lesions”, *Radiology*, vol.196, pp. 123-134, 1995.

[53] Swain MJ and Ballard DH, ”Color indexing”, *Int’l J. Computer Vision*, vol.7, no.1, pp 11-32, 1991.

[54] Vasconcelos N and Lippman A,“Bayesian representations and learning mechanisms for content based image retrieval”, *SPIE Storage and Retrieval for Media Databases, San Jose, CA*, 2000.

[55] Vasconcelos N and Lippman A,“Embedded mixture modeling for efficient probabilistic content-based indexing and retrieval”, *SPIE Multimedia Storage and Archiving Systems, Boston, MA*, 1998.

[56] Viola P and Wells WM, “Alignment by Maximization of Mutual Information”, *Fifth Int’l Conf. Computer Vision, Cambridge, MA*, pp 16-23, IEEE, 1995

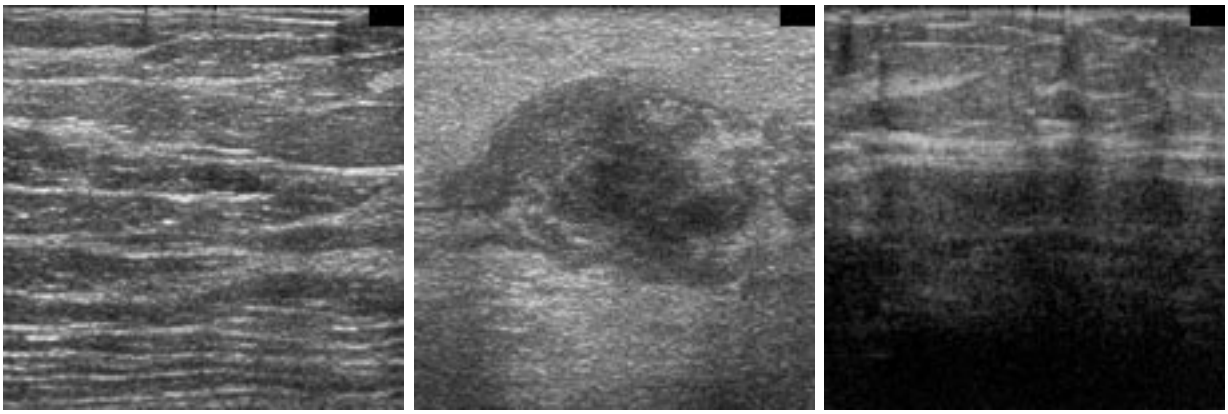

Figure 1: Three ultrasound breast scans. From left to right are: Case 151, Case 142 and Case 152.

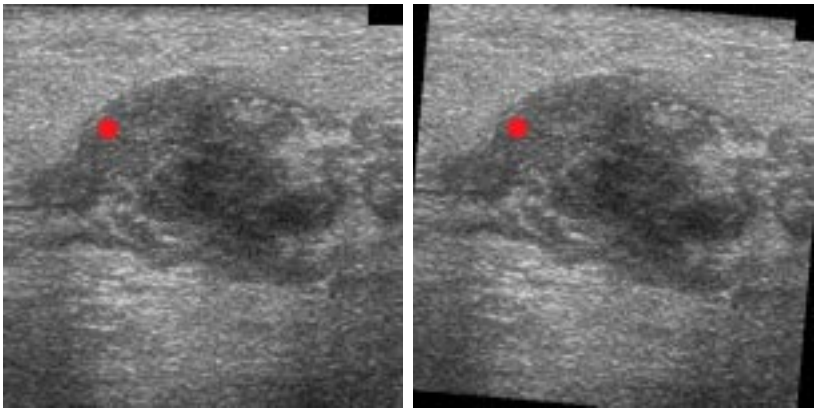

Figure 2: Single-pixel gray level coincidences are recorded by counting number of co-occurences of a pair of gray level across the reference (left) and secondary (right) images at a pair of homologous pixel locations (u, v). Here the secondary image (right) is rotated by 8^o relative to the reference image (left).

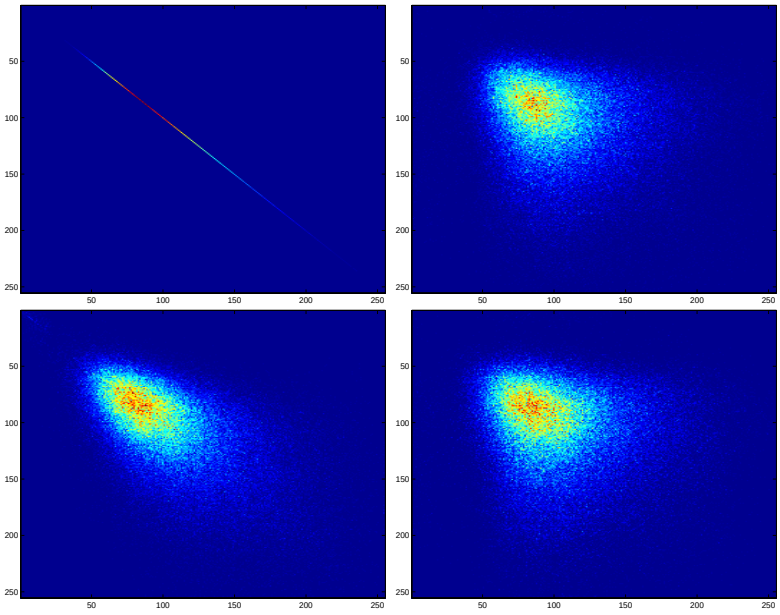

Figure 3: Joint coincidence histograms for single-pixel gray level features. Both horizontal and vertical axes of each panel are indexed over the gray level range of 0 to 255. Top left: joint histogram scatter plot for the case that reference image (X_i) and secondary image (X_j) are the same slice of the US image volume (Case 142) at perfect 0^o alignment ($X_j = X_i$). Top right: same as top left except that reference and secondary are misaligned by 8^o relative rotation as in Fig. 2. Bottom left: same as top right except that the reference and secondary images are from adjacent (2mm separation) slices of the image volume. Bottom right: same as bottom right except that images are misaligned by 8^o relative rotation.

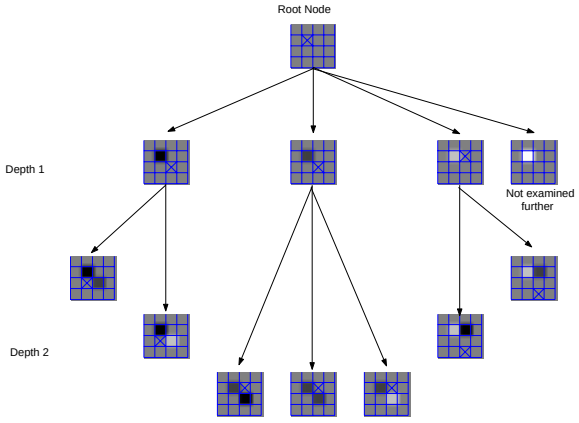

Figure 4: *Part of feature tree data structure.*

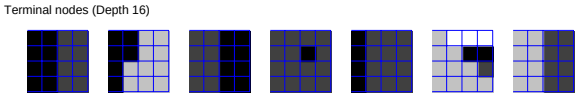

Figure 5: *Leaves of feature tree data structure.*

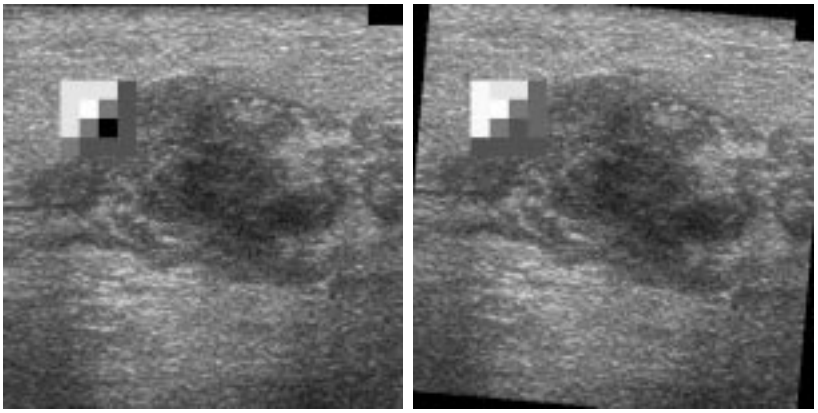

Figure 6: Tags: Each pixel is labeled by a tag type. Occurrences and coincidences of tag labels are then plotted on a joint histogram of tag features.

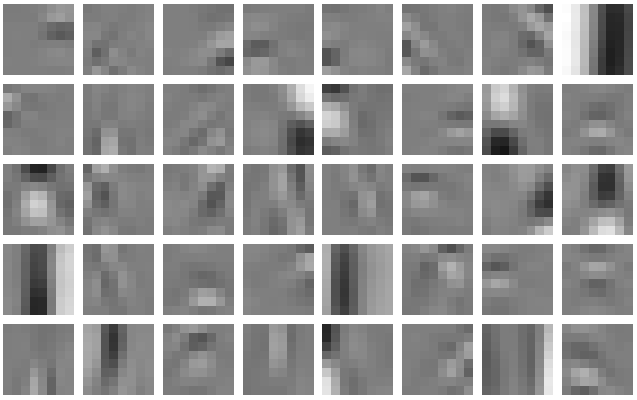

Figure 7: *ICA Basis: estimated 40-dimensional ICA basis set obtained from training on randomly selected* 8×8 *blocks in 10 ultrasound breast images.*

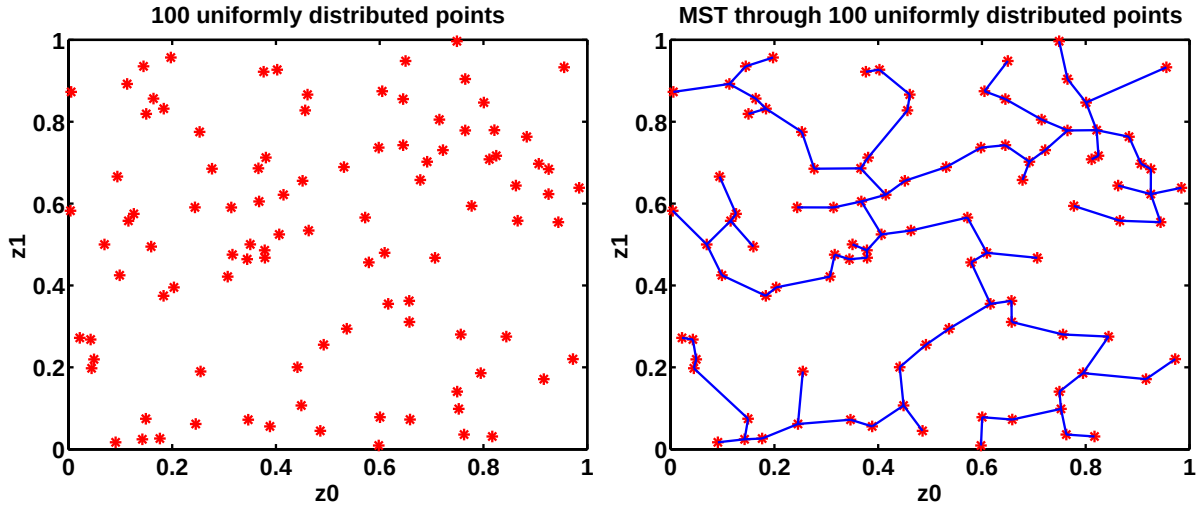

Figure 8: A set of n points $\{Z_i\}$ in the plane (left) and the corresponding Minimal Spanning Tree (MST) (right).

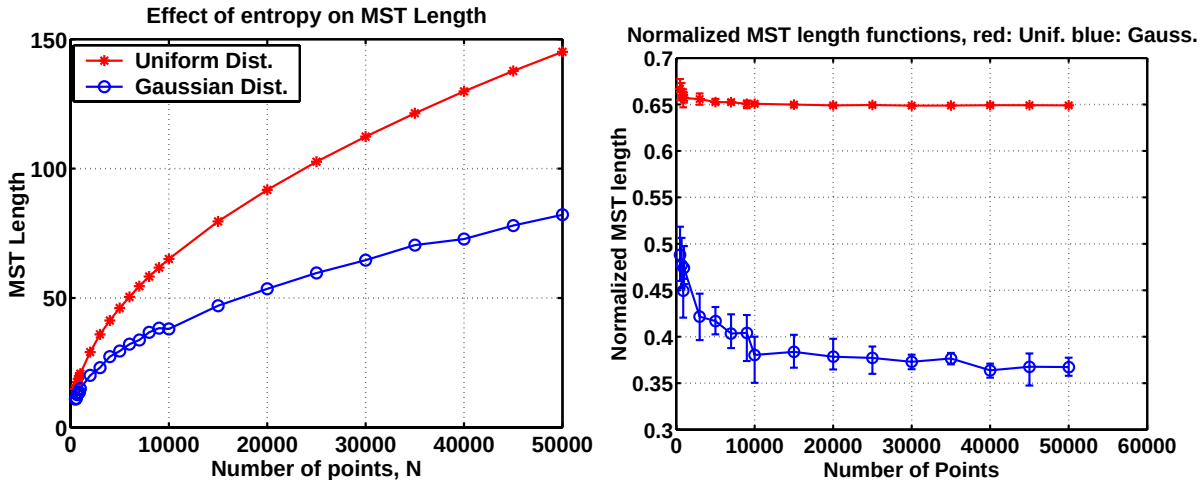

Figure 9: Length functions L_n of MST (left) and $L_n/\sqrt{n}$ (right) as a function of n for the uniform and normal distributed points in figure.

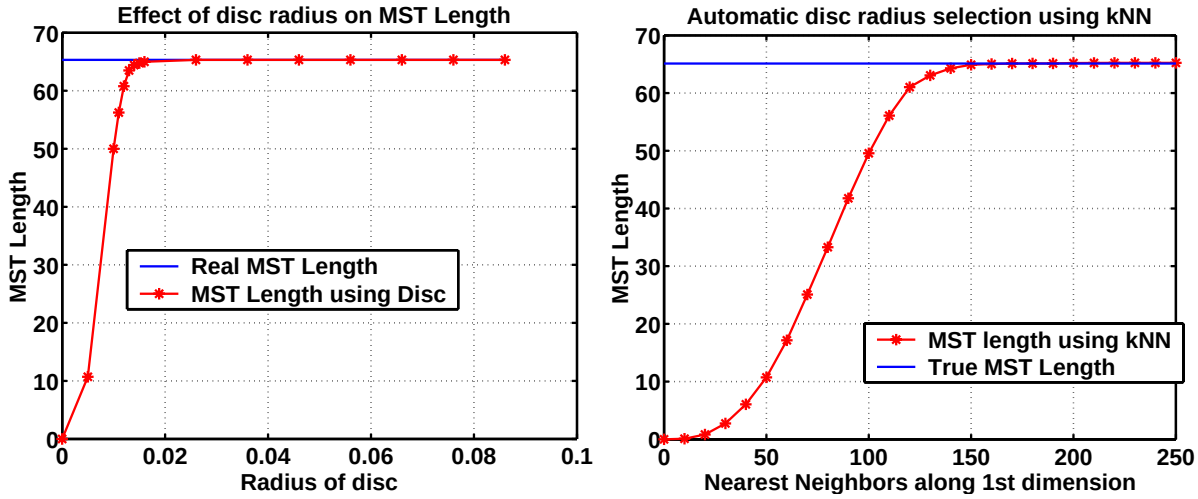

Figure 10: Bias of the $n\log n$ MST algorithm as a function of radius parameter (left) and as a function of the number of nearest neighbors (right)

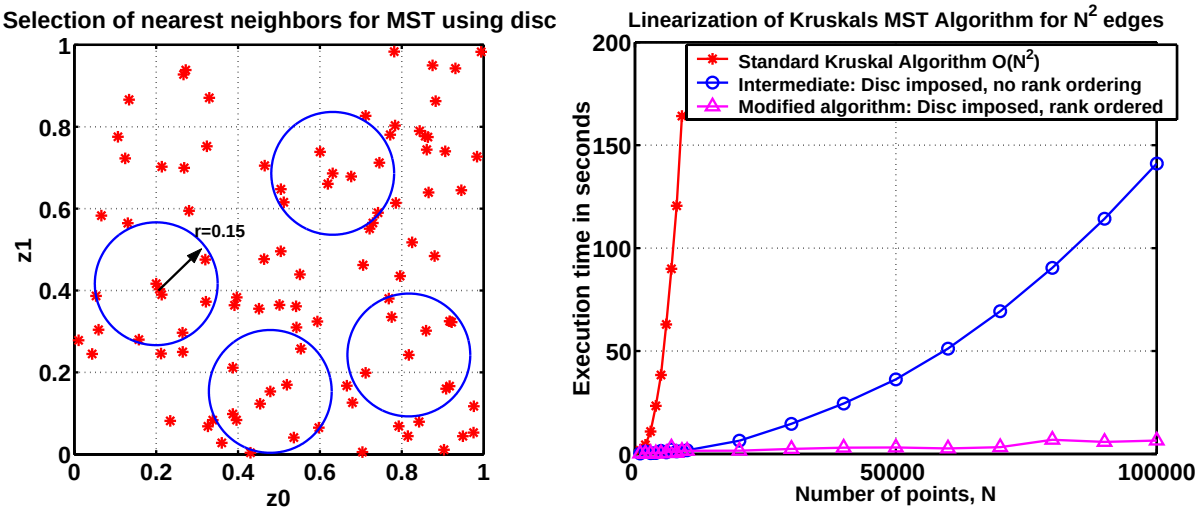

Figure 11: Acceleration of Kruskals MST algorithm from $n^2\log n$to $n\log n$ (left) and Comparison of Kruskal's MST to our $n\log n$ MST algorithm (right)

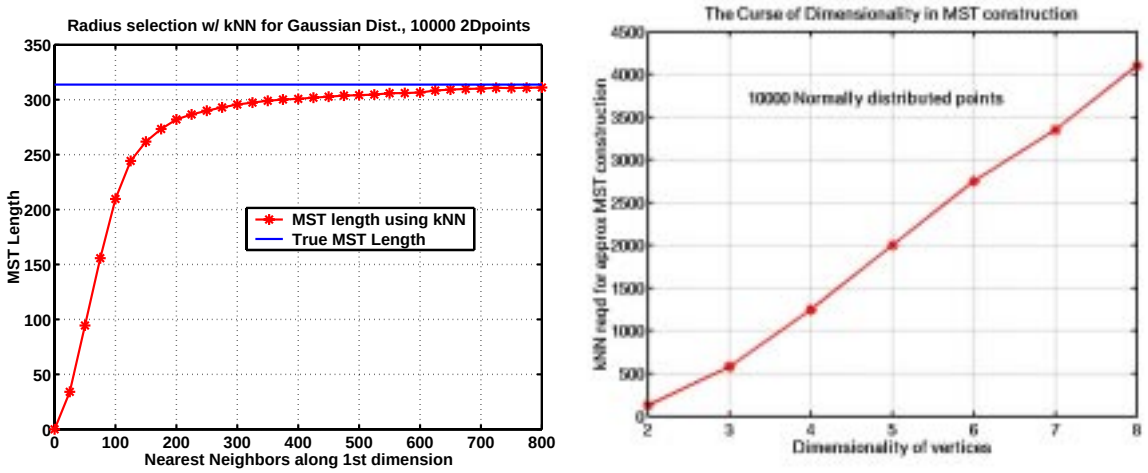

Figure 12: [a] Constructing the MST based on kNN based disc radius estimate could be a problem ion non-uniform distributions due to the slow convergence of the length function (left). [b] As the dimensionality of the dataset increases this problem aggravates and the partial search now approaches the full search (right).

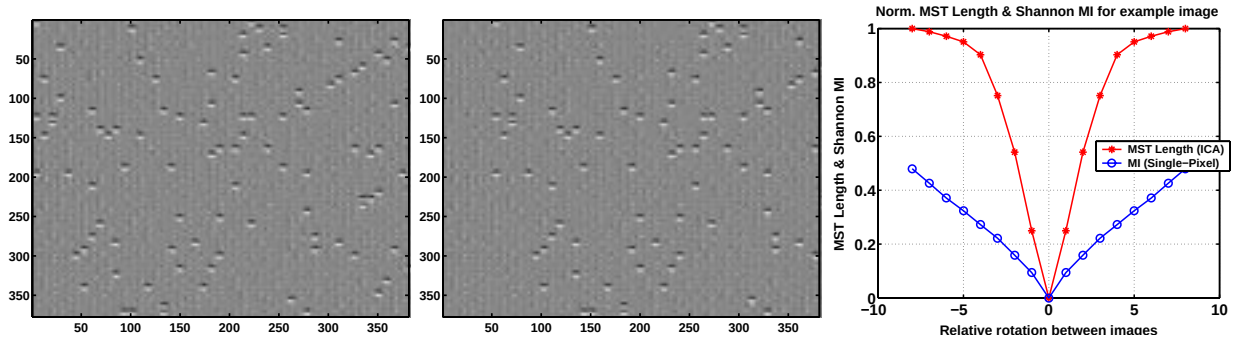

Figure 13: Image backgrounds have been synthesized using15 ICA bases elements each. The prominent feature seen in the images (right) is translated across the image (center) (exaggerated for effect). The background remains unchanged. Curves for single pixel based αMI and ICA based αJensen (right)

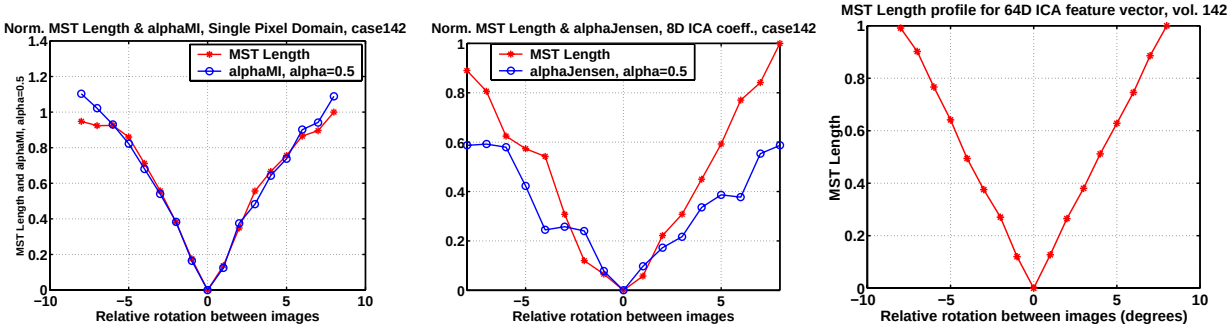

Figure 14: For registration of two slices taken from Case 142: MST and histogram based entropy estimation for single pixel features (left). MST and histogram based entropy estimation for 8D ICA features (center) and 64D ICA feature vectors (right)

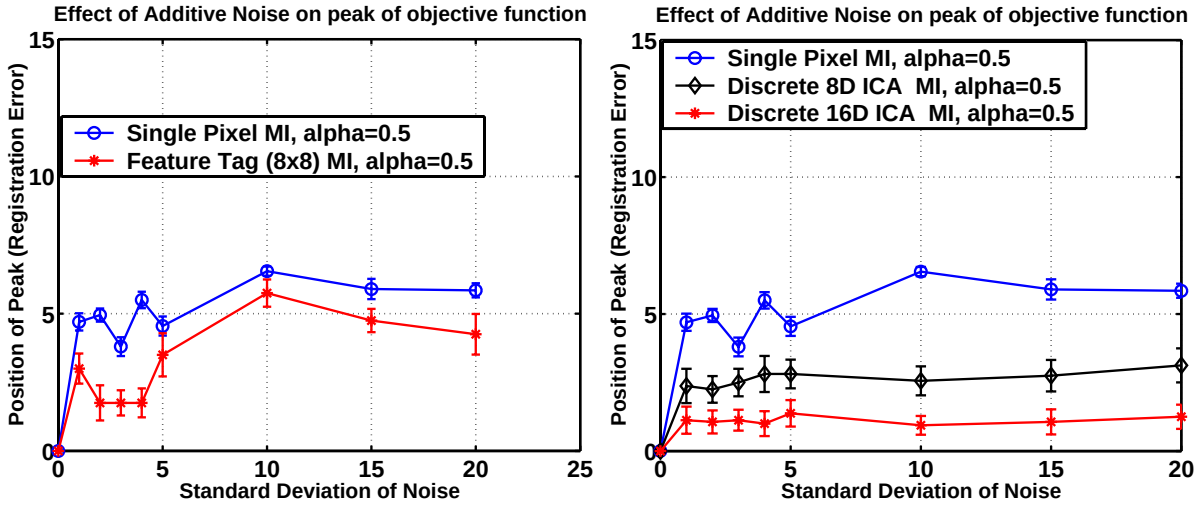

Figure 15: Effect of additive Gaussian noise on the rms of the peak position of the $alpha$- MI objective function estimated using histograms on single-pixel and feature coincidence trees of 8×8 tag features (left) and feature coincidence trees on discrete ICA (8D and 16D) features (right). These plots are based on 20 repeated experiments for the Case 142 image and its rotated cousin (maximum rotation angle was 16^o).

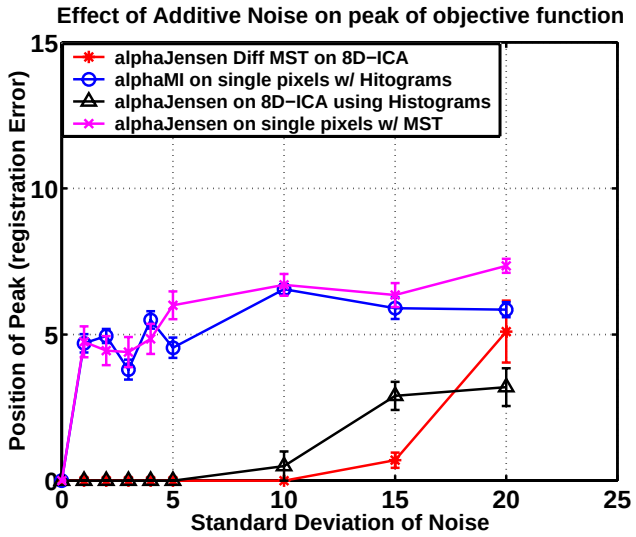

Figure 16: Effect of additive noise on the peak of the α- MI objective function estimated using histograms on single pixels, α- Jensen function estimated using histograms on single-pixels and 8D discrete and continuous ICA features.