Image matching 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

Matching a reference image to a secondary image extracted from a database of transformed exemplars constitutes an important image retrieval task. Two related 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 present a general method based on matching high dimensional image features, using entropic similarity measures that can be empirically estimated using entropic graphs such as the minimal spanning tree (MST). The entropic measures we consider are generalizations of the well known Kullback-Liebler (KL) distance, the mutual information (MI) measure, and the Jensen difference. Our entropic graph approach has the advantage of being implementable for high dimensional feature spaces for which other entropy-based pattern matching methods are computationally difficult. We compare our technique to previous entropy matching methods for a variety of continuous and discrete features sets including: single pixel gray levels; tag sub-image features; and independent component analysis (ICA) features. We illustrate the methodology for multimodal face retrieval and ultrasound (US) breast image registration.

Keywords: minimal spanning tree, MST, alpha-entropy estimation, image retrieval, 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

Image retrieval and image registration fall in the general area of pattern matching problems, where the best match to a reference or query image X_0 is to be found in a database of secondary images $\mathcal{X} = \{X_i\}_{i=1}^K$. The best match is expressed as a partial re-indexing of the database in decreasing order of similarity to the reference image using a similarity measure. In the context of image registration the database corresponds to an infinite set of transformed versions of a single secondary image, e.g. rotation and translation, which are compared to the reference image to register the secondary to the reference.

There are three key ingredients to image matching which can impact matching accuracy and computational efficiency: 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 a matching criterion 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 search implementation. This paper is concerned with item 2. Specifically we propose a general class of feature similarity measures that is based on entropy, can be implemented with entropic graphs, and does not require histogram or density estimation.

There are many techniques available for image retrieval and image registration [62, 30, 35]. Some of the most widespread techniques are: histogram matching [32]; texture matching [2]; intensity cross correlation [49]; optical flow matching [40]; kernel-based classification methods [9]; boosting retrieval methods [11, 36]; information divergence minimization [72, 68, 67, 21]; and mutual information (MI) maximization [74, 19]. These last two methods can be called "entropic methods" since both use a matching criterion defined as a relative entropy of the feature distributions. The main advantage of entropic methods is that they can capture non-linear relations between features in order to improve discrimination between poor and good image matches. When combined with a highly discriminatory feature set, and reliable prior information, entropic methods are very compelling and have been shown to be virtually unbeatable for some multimodality image registration

applications [42, 51, 30]. However, due to the difficulty in estimating the relative entropy for high dimensional features, the application of entropic methods have been limited to one or two feature dimensions. The independent successes of relative entropy methods, e.g., MI image registration, and the use of high dimensional features, e.g., SVM's for handwriting recognition, suggest that an extension of entropic methods to high dimensions would be worthwhile.

This paper introduces a way to extend entropic methods of image matching to high dimensional feature spaces using several innovations. First we generalize the framework of entropic methods to include Rényi's α entropies. Second we implement a novel and simple α-entropy estimator using entropic graphs [23] and in particular the MST. The computation/storage complexity of the MST grows only linearly in feature dimension and its asymptotic bias is independent of the feature distribution [24, 22]. This is to be contrasted with histogram estimators of entropy whose complexity increases exponentially in the number of dimensions and whose asymptotic bias depends on the underlying feature distribution. We illustrate the application of our entropic method to two applications for which high dimensional features are beneficial: ultrasound (US) breast image registration and image retrieval from a multimodality face database.

The α-entropy of a multivariate distribution is a generalization of the better known Shannon entropy. Alfred Rényi introduced the α-entropy in a 1961 paper [60] and since then many important properties of α-entropy have been established [4]. From Rényi's α-entropy the Rényi α-divergence and the Rényi α-mutual information (α-MI) can be straightforwardly defined. For $\alpha = 1$ these quantities reduce to the standard (Shannon) entropy, (Kullback-Liebler) divergence, and (Shannon) MI, respectively. Another useful quantity that can be derived from the α-entropy is the α-Jensen difference, which is a generalization of the standard Jensen difference and will be a key player in our extension of entropic pattern matching methods to high feature dimension. As we will show, this generalization allows us to define an image matching algorithm that benefits from a simple estimation procedure and an extra degree of freedom (α).

Some additional comments on relevant prior work by us and others is in order. Various forms of α-entropy have been exploited by others for applications including: reconstruction and registration of interferometric synthetic aperture radar (ISAR) images [21, 17]; blind deconvolution [16]; and

time-frequency analysis [3, 75]. Again, our innovation with respect to these works is the extension to high dimensional features via entropic graph estimation methods. On the other hand, the alpha-entropy approaches described here should not be confused with entropy-alpha classification in SAR processing [8] which has no relation whatsoever to our work. A tutorial introduction to the use of entropic graphs to estimate multivariate α-entropy and other entropy quantities was published by us in a recent survey article [23]. As introduced in [27] and studied in [24, 26, 28] 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 most attractive entropic graph algorithm. This graph estimator can be viewed as a multidimensional generalization of the Vasicek Shannon entropy estimator for one dimensional features [73, 6].

The two applications presented in this paper were primarily selected to illustrate the flexibility of our method. In the ultrasound breast registration application we adopt two types of high dimensional features: local tags and local basis projection coefficients. Local tags are extracted from the image by fast and simple adaptive quantization scheme proposed by Geman and Koloydenko [18]. Local basis projection coefficients are implemented by projecting local 8 by 8 sub-images of the image onto a DCT basis for the face retrieval application and projecting onto an ICA basis for the image registration application. Local feature extraction via basis projection is a commonly used technique for image representation [64, 71, 70]. ICA (Independent Components Analysis) features are somewhat less common but have been similarly applied by Olshausen, Hyvärinen and others [43, 33]. The high dimensionality ($= 64$ for local basis projections) of these feature spaces precludes the application of standard entropic pattern matching methods and provides a good illustration of the power of our approach.

The outline of this paper is as follows. Section 2 defines the general class of information theoretic measures which is the subject of this paper. Section 3 presents the methods of extracting tag and ICA features used in our studies. The MST entropy estimator is presented in Section 4. Section 5 illustrates the MST-based entropy method for registration and retrieval. Conclusions and future directions are discussed in Section 6.

2 Entropic Feature Similarity Measures

Let Y be a q-dimensional random vector and let $f(y)$ and $g(y)$ denote two possible densities for Y. In the sequel Y will be a feature vector constructed from the query image and a secondary image in an image database and f and g will be histograms or densities. For example, information divergence methods of image retrieval [66, 13, 70] specify f as the estimated density of the query image features and g as the estimated density of the secondary image features. When the features are discrete valued the densities f and g are interpreted as probability mass functions.

The basis for entropic methods of image matching is a measure of similarity between densities f and g. A very general entropic similarity measure is the Rényi α-divergence, also called the Rényi α-relative entropy, between f and g of fractional order $\alpha \in (0,1)$ [60, 10, 4]

$$\begin{aligned} D_\alpha(f\|g) &= \frac{1}{\alpha-1}\log\int g(z)\left(\frac{f(z)}{g(z)}\right)^\alpha dz \\ &= \frac{1}{\alpha-1}\log\int f^\alpha(z)g^{1-\alpha}(z)dz. \end{aligned} \tag{1}$$

When the density f is supported on a compact domain and g is uniform over this domain the α-divergence reduces to the Rényi α-entropy:

$$H_\alpha(f) = \frac{1}{1-\alpha}\log\int f^\alpha(z)dz. \tag{2}$$

When specialized to various values of α the α-divergence can be related to other well known divergence measures. Two of the most important examples are the Hellinger affinity $2\log\int\sqrt{f(z)g(z)}dz$ obtained when $\alpha = 1/2$, which is related to the Hellinger-Battacharya distance squared,

$$\begin{aligned} D_{Hellinger}(f\|g) &= \int\left(\sqrt{f(z)}-\sqrt{g(z)}\right)^2 dz \\ &= 2\left(1-\exp\left(\tfrac{1}{2}D_{\frac{1}{2}}(f\|g)\right)\right), \end{aligned}$$

and the Kullback-Liebler (KL) divergence [39], obtained when $\alpha \to 1$,

$$\lim_{\alpha\to 1} D_\alpha(f\|g) = \int g(z)\log\frac{g(z)}{f(z)}dz.$$

2.1 Mutual Information Image Matching

The MI similarity measure was introduced for gray scale image registration by Viola and Wells [74] and has since been applied to a variety of image matching problems [19, 42, 51, 59]. Let X_0 be a reference, or query, image and consider a database X_i, $i = 1, \ldots, K$ of secondary images. We assume that the images are sampled on a grid of $M \times N$ pixels. Let (z_{0k}, z_{ik}) be the pair of gray levels extracted from the k-th pixel location in the query and in the i-th image in the database, respectively. The basic assumption underlying MI image matching is that $\{(z_{0k}, z_{ik})\}_{k=1}^{MN}$ are independent identically distributed (i.i.d.) realizations of a pair (Z_0, Z_i) of random variables having joint density $f_{0i}(z_0, z_i)$. If the query and the secondary images were correlated, e.g., identical images, then Z_0 and Z_1 would be dependent random variables. On the other hand if the two images were statistically independent the joint density of Z_0 and Z_1 would factor into the product of the marginals $f_{0i}(z_0, z_i) = f_0(z_0) f_i(z_i)$. This suggests using the α-divergence $D_\alpha(f_{0i}(z_0, z_i) \| f_0(z_0) f_i(z_i))$ between $f_{0i}(z_0, z_i)$ and $f_0(z_0) f_i(z_i)$ as a similarity measure. For $\alpha \in (0, 1)$ we call this the α-mutual information (MI) between Z_0 and Z_i and it has the form

$$D_\alpha(f(Z_i, Z_0) \parallel f(Z_i) f(Z_0)) = \frac{1}{\alpha - 1} \log \int f_{0i}^{\alpha}(z_0, z_i) f_0^{1-\alpha}(z_0) f_i^{1-\alpha}(z_i) dz_0 dz_i. \tag{3}$$

When $\alpha \to 1$ the α-MI converges to the standard (Shannon) MI

$$\text{MI} = \int f_{0i}(z_0, z_i) \log \left(\frac{f_{0i}(z_0, z_i)}{f_0(z_0) f_i(z_i)} \right) dz_0 dz_i.$$

For registering two discrete $M \times N$ images, Viola and Wells [74] search over a set of transformations of the secondary image to find the one that maximizes the MI (4) between the query and the transformed secondary. The MI is defined using features $(Z_0, Z_i) \in \{z_{0k}, z_{ik}\}_{k=1}^{MN}$ equal to the discrete-valued intensity levels at common pixel locations (k, k) in the query image and the rotated secondary image. We call this the "single pixel MI" in the sequel. Viola and Wells empirically approximated the single pixel MI (4) by "histogram plug-in" estimation, which when extended to the α-MI gives the estimate

$$\hat{\text{MI}} \stackrel{\text{def}}{=} \frac{1}{\alpha - 1} \log \sum_{z_0, z_i = 0}^{255} \hat{f}_{0i}^{\alpha}(z_0, z_i) \left(\hat{f}_0(z_0) \hat{f}_i(z_i) \right)^{1-\alpha}. \tag{4}$$

In (4) we assume 8-bit gray level, $\hat{f}_{0i}$ denotes the joint intensity level "coincidence histogram"

$$\hat{f}_{0i}(z_0, z_i) = \frac{1}{MN} \sum_{k=1}^{MN} I_{z_{0k}, z_{ik}}(z_0, z_i), \tag{5}$$

and $I_{z_{0k}, z_{ik}}(z_0, z_i)$ is the indicator function equal to one when $(z_{0k}, z_{ik}) = (z_0, z_i)$ and is equal to zero otherwise.

To illustrate the general procedure, the coincidence histogram is shown in Fig. 1 for the case of registration of US breast images X_0, X_1 (Fig. 2). Fig. 1 shows two cases. At top left is the coincidence histogram when the reference and secondary images are taken from the same two-dimensional slice of the US breast volume and are in perfect alignment ($X_0 = X_1$). At bottom left is the same histogram when the secondary image is rotated by 8^o. The top right and bottom right panels in Fig. 1 are analogous except that the secondary images is extracted from a different two-dimensional slice separated from the reference (query) by 2mm. 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. In both cases the spread of the histogram is greater for the bottom panels (out of alignment) than for the top panels (in alignment) of the figure. The α-MI will take on greater values for the less spread top panels than for the more spread bottom panels.

2.1.1 Relation of α-MI to Chernoff Bound

α-MI (3) can be motivated as an appropriate registration function by large deviations theory through the Chernoff bound. Define the average probability of error $P_e(n)$ associated with a decision rule for deciding whether Z_i and Z_0 are independent (hypothesis H_0) or dependent (hypothesis H_1) random variables based on a set of i.i.d. samples $\{z_{0k}, z_{ik}\}_{k=1}^{n}$, where $n = MN$. For any decision rule, 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 the probabilities of Type II (say H_0 when H_1 true) and Type I (say H_1 when H_0 true) errors, respectively, of the decision rule and $P(H_1) = 1 - P(H_0)$ is the prior probability of H_1. When the decision rule is the optimal minimum probability of error test the

Chernoff bound implies that [12]:

$$\liminf_{n\to\infty} \frac{1}{n} \log P_e(n) = - \sup_{\alpha\in[0,1]} \{(1-\alpha) D_\alpha(f_{0i}(z_0, z_i) \| f_0(z_0) f_1(z_i)\} . \tag{6}$$

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 the number $n = MN$ of samples. In particular, this implies that the α-MI can be used to select optimal features Z_0, Z_i defined as those features that maximize the right side of (6). The appearance of the maximization over α implies the existence of an optimal parameter α ensuring the lowest possible registration error. This optimal value α is not generally equal to 1 so that the MI criterion is generally suboptimal in the sense of minimizing the asymptotic probability of error. For a more detailed discussion on the issue of optimal selection of α we refer the reader to [24].

2.2 α-Jensen Similarity Measure

An alternative entropic similarity measure between two distributions is the α-Jensen difference. This function has been independently proposed by Ma [25] and He *et al* [21] for image registration problems. It was also used by Michel *et al* in [52] for characterizing complexity of time-frequency images. For two densities f and g the α-Jensen difference is defined as [4]

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

where $\alpha \in (0, 1)$ and $\beta \in [0, 1]$. As the α-entropy $H_\alpha(f)$ is strictly concave in f Jensen's inequality implies that $\Delta H_\alpha(\beta, f, g) > 0$ when $f \neq g$ and $\Delta H_\alpha(\beta, f, g) = 0$ when $f = g$ (a.e.). Thus the α-Jensen difference is a bone fide measure of dissimilarity between f and g.

The α-Jensen difference can either be applied as a surrogate for the α-MI or the α-divergence. When applied as a surrogate for α-divergence one identifies $f = f_i(z_i)$ and $g = f_0(z_0)$ in (7). In this case an image match occurs when the α-Jensen difference is minimized over i. This is the approach taken by [21, 47] for image registration applications and discussed in more detail below.

On the other hand, the α-Jensen difference can also be used as a surrogate for the α-MI if one identifies $f = f_{0i}(z_0, z_i)$ and $g = f_0(z_0) f_i(z_i)$ in (7). In this case to find a matching image to

a query the α-Jensen difference is maximized over i. Asymptotic comparison between the α-MI and the α-Jensen difference can give useful insight [24]. It can be shown that when the features Z_0, Z_i are nearly independent than the most discriminating value of α is 1/2 for the α-MI. For the α-Jensen difference the best value of α is 1 and the best value of β is 1/2. While use of α-Jensen as a surrogate for α-MI is certainly worthy of additional study, its computational requirements and its performance appear similar to that of α=MI and therefore we do not consider it further in this paper.

3 Feature-based Matching

Scalar single pixel intensity level is only one possible feature that can be used to perform image matching. As pointed out by Leventon and Grimson [42], MI does not take into account joint spatial behavior of the coincidences and this can cause poor registration, especially in multi-modality situations. Alternative vector valued features have been investigated by many image retrieval and registration researchers. We will focus on two types of vector features which generalize pixel intensity levels: local tag features and local basis projection features.

3.1 Local Tag Features

Tag features were introduced by Amit and Geman [1] and used for shape recognition. 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. For gray scale images, the number of different tag types can be extremely large. For example, if the image intensities are quantized to an 8-bit plane then there would exist $(256)^{64}$ different 8×8 tag types. Therefore, methods for pruning the tag types are essential for practical implementation. Randomized feature selection and adaptive thresholding are methods of pruning which were described by Geman and Koloydenko [18] and which we adapted to the US image registration application described below. For more details on our implementation of tag features see [54].

To illustrate we show in Fig. 3 tag features at a given pixel location for two US breast images in the same 2D slice but at two rotation angles. Coincidences of tag types are calculated by counting joint occurrences of feature types at identical spatial locations in the two images. The amplified tag pattern in the image on the left captures the edge of the tumor. A similar tag type will be observed in the secondary image on the right if it nearly aligned. These tags capture the local intensity pattern in the neighborhood of the pixel. The advantage of tags for matching US breast images is that they can more easily discriminate between speckle and tissue echos than can single pixel intensity levels.

3.2 Local Basis Projection Features

Basis projection features are extracted from an image by projecting local sub-images onto a basis of linearly independent sub-images of the same size. Such an approach is widely adopted in image retrieval applications, in particular with DCT or more general 2D wavelet bases [71, 70, 13, 64, 41, 48, 14]. Others have extracted a basis set adapted to image database using principal components (PCA) or independent components analysis (ICA) [43, 33].

The ICA basis is especially well suited for our purposes since it aims to obtain vector features which have statistically independent elements that can facilitate estimation of α-MI and other entropic measures. Specifically, in ICA an optimal basis is found which decomposes the image X_i into a small number of approximately statistically independent components (sub-images) $\{S_j\}$:

$$X_i = \sum_{j=1}^{p} a_{ij} S_j. \tag{8}$$

In the sequel we select basis elements $\{S_j\}$ from an over-complete linearly dependent basis using randomized selection over the database. For image i the feature vectors Z_i are defined as the coefficients $\{a_{ij}\}$ in 8 obtained by projecting the image onto the basis.

In Fig. 4 we illustrate the ICA basis selected for the US breast image database. ICA was implemented using Hyvarinen and Oja's [33] `FastICA` code (available from [34]) which uses a fixed-point algorithm to perform maximum likelihood estimation of the basis elements in the ICA data model (8). Figure 4 shows a set of 64 8×8 basis vectors which were learned from over 5000

8×8 training sub-images randomly selected from 10 consecutive image slices of a single ultrasound volume scan of the breast (Case 151 in Fig. 5). Given this ICA basis and a pair of to-be-registered $M \times N$ images, coefficient vectors are extracted by projecting each 8×8 neighborhood in the images onto the basis set. For the 64 dimensional ICA basis shown in Fig. 4 this yields a set of MN vectors in a 64 dimensional vector space which will be used to define features.

3.3 Discrete vs. Continuous Features

While adaptive thresholding yields tag features that are discrete valued, ICA and other basis projection features are continuous valued. The potentially high dimension of the basis projection feature space makes estimation of the α-MI and α-Jensen similarity measures problematic. A brute force method would be to discretize the vector of projection coefficients, e.g. using vector quantization [44], and generate histograms over the Voronoi cells. These histograms could then be used in the formula for α-MI or α-Jensen difference to yield plug-in estimators of these quantities. This presents difficulties for image matching applications since the reference and secondary images must all use the same cell partition in order to maintain consistency. For high dimensional feature space this brute force method also suffers from large bias unless one uses an impractically large number of cells. An alternative that can be applied to directly estimating the α-Jensen difference is explored in the next section.

4 Minimum Spanning Trees for Entropy Estimation

Implementation of the α-Jensen registration criterion can be accomplished by plugging in the feature histogram estimates to (7) analogously to the α-MI plug-in estimator (4). However, when the number of feature dimensions increases histogram methods become impractical due to the curse of dimensionality: for fixed resolution per coordinate dimension the number of histogram bins increases geometrically in feature vector dimension. For example, for a 32 dimensional feature space, only 10 cells per dimension would require keeping track of 10^{32} bins in the histogram, an unworkable and impractically large burden for any current computer. As high dimensional feature

spaces can be more discriminatory this creates a barrier to performing histogram-based entropic 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 [23, 24].

As explained in [23] the most computationally attractive entropic graph method known today is the minimal spanning tree (MST). Given a set $\mathcal{Z}_n = \{z_1, z_2, ..., z_n\}$ of n i.i.d vectors in a d-dimensional feature space $\mathbf{R^d}$ a spanning tree is a connected acyclic graph which passes through all n points in $\mathcal{Z}_n$. The MST spans all n points and connects them with $n-1$ edges denoted $\{e_i\}$. More specifically, for a given edge weight exponent $\gamma \in (0, d)$ the MST is defined as the spanning tree which minimizes the (total) length:

$$L(\mathcal{Z}_n) = \min_{e \in T} \sum_e \|e\|^\gamma, \tag{9}$$

where $\|e\|$ denotes the Euclidean (L2) length of the edge. See Fig 6 for an illustration for points in the plane. In the sequel we adopt $\gamma = 1$ for all experiments.

The MST length $L_n = L(\mathcal{Z}_n)$ is plotted as a function of n in Fig. 7 for the case of an i.i.d. uniform sample (right panel) and non-uniform sample (left panel) of $n = 100$ points in the plane. It is intuitive that the length of the MST spanning the more concentrated non-uniform set of points increases at a slower rate in n than does the MST spanning the uniformly distributed points. This observation has motivated the MST as a way to test for randomness in the plane [31]. More precisely, with $\alpha \stackrel{\text{def}}{=} (d-\gamma)/d$ the Beardwood, Halton, Hammersley theorem [5] and its extensions [77, 65] imply that the log of the length function normalized by n^α converges (a.s.) within a constant factor to the *alpha*-entropy. Thus we can identify the difference between the asymptotes shown on the left Fig. 7 as the difference between the α-entropies of the uniform and non-uniform densities ($\alpha = 1/2$). Thus, if f is the underlying density of $\mathcal{Z}_n$, the α-entropy estimator

$$\hat{H}_\alpha(\mathcal{Z}_n) = 1/(1-\alpha) \left[\log L(\mathcal{Z}_n)/n^\alpha - \log \beta_{d,\gamma}\right], \tag{10}$$

is an asymptotically unbiased and almost surely consistent estimator of the α-entropy of f where $\beta_{d,\gamma}$ is a constant which does not depend on the density f.

Using the above fact, the MST approach to estimating the α-Jensen difference between the feature densities of two images is implemented as follows. Assume two sets of feature vectors

$\mathcal{Z}_0 = \{z_0^{(i)}\}_{i=1}^{n_0}$ and $\mathcal{Z}_1 = \{z_1^{(i)}\}_{i=1}^{n_1}$ are extracted from images X_0 and X_1 and are i.i.d. realizations from multivariate densities f_0 and f_1, respectively. In the applications explored in this paper $n_0 = n_1$ but it is worthwhile to maintain this level of generality. Define the set union $\mathcal{Z} = \mathcal{Z}_0 \cup \mathcal{Z}_1$ containing $n = n_0 + n_1$ unordered feature vectors. If n_0, n_1 increase at constant rate as a function of n then any consistent entropy estimator constructed from the vectors $\{Z^{(i)}\}_{i=1}^{n_0+n_1}$ 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 finite sample entropic graph estimator of α-Jensen difference

$$\Delta \hat{H}_\alpha(\beta, f_0, f_1) = \hat{H}_\alpha(\mathcal{Z}_0 \cup \mathcal{Z}_1) - [\beta \hat{H}_\alpha(\mathcal{Z}_0) + (1-\beta)\hat{H}_\alpha(\mathcal{Z}_1)], \tag{11}$$

where $\beta = n_0/n$, $\hat{H}_\alpha(\mathcal{Z}_0 \cup \mathcal{Z}_1)$ is the MST entropy estimator constructed on the n point union of both sets of feature vectors and the marginal entropies $\hat{H}_\alpha(\mathcal{Z}_0)$, $\hat{H}_\alpha(\mathcal{Z}_1)$ are constructed on the individual sets of n_0 and n_1 feature vectors, respectively. We can similarly define a density-based estimator of α-Jensen difference. Observe that for affine image registration problems the marginal entropies $\{H_\alpha(f_i)\}_{i=1}^K$ over the set of image transformations will be identical, obviating the need to compute estimates of the marginal α-entropies.

As contrasted with histogram or density plug-in estimator of entropy or Jensen difference, the MST-based estimator enjoys the following properties [24, 22, 29]: it can easily be implemented in high dimensions; it completely bypasses the complication of choosing and fine tuning parameters such as histogram bin size, density kernel width, complexity, and adaptation speed; as the topology of the MST does not depend on the edge weight parameter γ, the MST α-entropy estimator can be generated for the entire range $\alpha \in (0,1)$ once the MST for any given α is computed; the MST can be naturally robustified to outliers by methods of graph pruning. On the other hand the need for combinatorial optimization may be a bottleneck for a large number of feature samples for which accelerated MST algorithms are necessary.

4.1 Computational Acceleration of the MST

Two principal algorithms exits for computing the MST, the Prim algorithm [58] and the Kruskal algorithm [38]. For sparse graphs the Kruskal algorithm is the fastest general purpose MST com-

putation algorithm. Kruskal's algorithm maintains a list of edges sorted by their weights and grows the tree one 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 Kruskal algorithm is of order $O(E \log E)$ and the the memory requirement is $O(E)$, where E is the initial number of edges in the graph.

The most simple-minded construction of the MST is to include all the possible edges within the feature set. This results in N^2 edges for N points; a time requirement of $O(N^2)$ and a memory requirement of $O(N^2 \log N)$. 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 $M \times N$ pixels, the total number of points in the graph is $2 \times M \times N \approx 150{,}000$ for images of size 256×256. This is beyond the capabilities of even the fastest available desktop processors.

Significant acceleration can be obtained by sparsification of the initial graph before tree construction. We have implemented a method for sparsification that allows us to construct MST's for several hundred thousand points in seconds of desktop computing time. This implementation uses a disc windowing method for constructing the edge list. Specifically, we center discs at each point under consideration (See Fig. 8 for illustration). We also use a list intersection approach similar to [56] to prune unnecessary edges within the disk. The two methods greatly reduce the number of edges that must be sorted for each point. We have empirically found that for approximately uniform distributions, a constant disc radius is optimal. For non uniform distributions, the disc radius is better selected as the distance to the k^{th}-nearest neighbor (kNN). Fig. 9 shows the bias of modified MST algorithm as a function of the radius parameter and the number of nearest neighbors for a uniform density on the plane.

It is straightforward to prove that, if the radius is suitably specified, our MST construction yields a minimum spanning tree. Recall that the Kruskal algorithm ensures construction of the exact MST [38].

(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 total edge weight greater than disc radius or have led to a cyclic path. Expanding the disc radius would then provide the path which is lowest in weight and non-cyclic.

For further information and variants of the MST acceleration technique described above we refer the reader to [54].

5 Applications

We herein illustrate the entropic graph approach to image matching for three different problems. The first is a toy problem involving registration of a simulated image and is meant to illustrate the superiority of entropic methods over correlation methods. The second application is registration of US breast images and the third application is image retrieval over a face database.

5.1 Toy Image Registration Problem

In this simple example we compare the MST-based estimate of α-Jensen difference using basis projection features against the standard correlation coefficient method [49]. Fig. 10 shows two images, containing a square at constant intensity, immersed in a background of uncorrelated white noise. The square in the secondary image is translated along the diagonal from the original central position in the primary image. The objective is to register the square in the two images by finding the diagonal translation of the secondary image which best matches the first image. For each candidate translation the derived features were merged into one feature set as described in Section 4 to compute the MST-based α-Jensen difference. We adopted a 4 dimensional feature set consisting of the coefficients obtained from projecting the 8×8 sub-image at each pixel location i onto a 2D basis of centered vertical and horizontal lines plus the two Cartesian coordinates of i.

Figure 10 is a plot of the MST-based α-Jensen criterion as a function of diagonal translation

of the secondary image. Superimposed on this plot is plotted the estimated correlation coefficient. Since the α-Jensen difference trajectory displays a deep minimum at the correct translation (0^o offset in the plot), the MST-based entropic registration method clearly outperforms the correlation method in aligning the two images.

5.2 Ultrasonic Breast Image Registration

Ultrasound breast images are notoriously difficult to register due to speckle noise, elastic deformations, and shadows. Here we compare several entropic registration methods including MI and α-MI to the MST-based α-Jensen method. The database used for this application 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. Each volume scan acquired at 1cm depth resolution yields 90 cross-sectional images at 0.4cm horizontal resolution. 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 5 shows slices of breast ultrasound image volumes representative of those found in clinical practice. The women were imaged on their backs with the transducer placed so as to image 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 uncommon. 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 of our methods to fully 3D voxel registration is straightforward but will not be presented here.The value of α used for all simulations

is 0.5.

Figure 11 shows representative profiles of the registration objective function for registering a slice of Case 142 US breast image volume 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 methods we restricted our investigation to rotation transformations over a small range (-15^o to $+15^o$) of angles. The panel on far left of Fig. 11 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 having a unique global minimum at the correct 0^o rotation of the reference image. 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. An ICA basis of 8×8 sub-images was generated by randomized feature selection on the image volumes. Two cases were investigated: a reduced dimension feature set consisting of only the first 8 of the 64 feature dimensions and the full 64 dimensional features. Here we assume that the dimensions of space that are not spanned by the independent components are filled by Gaussian noise. We see that computing the nongaussian projection pursuit directions, we effectively estimate the independent components ([33]). When all nongaussian directions have been found, all independent components have been estimated. For 8 dimensional ICA features we observe from Fig. 11 that the profile of the histogram plug-in estimate of α-Jensen is degraded (middle panel) exhibiting several local minima. This is expected since histogram estimation becomes unstable in such a high (8) dimensional feature space. In the full 64 dimensional ICA feature space the MST-based Jensen difference criterion maintains a smooth profile (right panel) with a single global minimum at the correct location. In the 64 dimensional feature space the histogram plug-in estimates of α-MI or α-Jensen difference are 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 criteria under increasing noise conditions. Figures 12 and 13 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^o relative rotation) by an amount depending on the SNR, the registration features, and entropy/MI estimation methods. The lack of smoothness in the plots is likely due to the image-specific nature of the simulation - averaging results over a much larger database of breast images would undoubtedly produce smoother graphs.

Figure 12 shows a comparison of histogram-based estimates derived from pixel tag features. We see that tag features have lower MSE than the standard single pixel features. Also higher order ICA features 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 13 shows a comparison of the α-MI ($\alpha = 0.5$) and α-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. The performance of higher order continuous ICA features estimated from the MST is seen to be better than those of single pixel features for which the α-MI is estimated using histograms. More extensive experiments are necessary but these preliminary results indicate that the MST method of entropy estimation have significantly greater robustness to additive noise than histogram plug-in methods.

5.3 Multimodal Face Retrieval

Face detection and retrieval from 2D images is a very difficult problem due to the high variability of facial expressions, poses, and illuminations. Many different approaches to this problem have been proposed [46, 7, 76]. Our objective in this paper is not to compete with these many fine tuned approaches. Rather we simply wish to illustrate the flexibility of the MST entropic matching method that we have presented in earlier sections.

The widespread availability of thermal cameras, such as those used at airports for detecting fever patients in the recent SARS outbreak, provides an opportunity to couple information from visible-light and infrared sources for face identification. The Equinox visible/IR face database [15]

consists of 7GB worth of images of persons photographed with differing illumination conditions, poses, and facial expressions using a joint co-registered visible longwave infrared (V/LWIR) camera. Figure 14 shows a sampling of faces in this database. Given a V/LWIR image pair for a person the multimodal face retrieval problem is to extract a corresponding pair of images of the person from the database. Multimode retrieval using visible and thermal imagery is difficult due to the prominence of contour information as opposed to the textural details available in visible imagery. The complete lack of textural information typically leads investigators to use facial landmarks for indexing images. However, facial landmarks change positions and aspects with expressions and movement making them unreliable. This makes entropic methods of retrieval highly compelling for this problem due their ability to capture complex relations using high dimensional features and requiring no user intervention.

We implemented MST-based entropic retrieval as follows. We pulled queries at random from the database and used this query to test the image matching method between the query and the remaining images in the database. Two sample queries are illustrated in Figure 15. The remainder of the database was compared to the query image and rank-ordered with respect to their similarity with the query image. A perfect match was declared if the image with highest rank, as measured by estimated α-Jensen difference, matched the person in the query image. Rather than implement ICA, which has been reported to have deficiencies for face recognition [45], we used a 8x8-DCT basis set to extract the features. Matching is done in a 66-dimensional space, using 64 of the DCT coefficients dimensions plus 2 dimensions corresponding to spatial coordinates. The α-Jensen difference is computed by building the MST over this high-dimensional space for each image pair. The measured correct retrieval rate was a respectable 95.5% even using this relatively simple feature set and the simple MST α-Jensen image matching method. over a 300 random queries, corresponding to different facial expressions and illumination conditions.

6 Conclusions

In this paper we presented a simple entropic graph approach to image matching. The method is capable of capturing entropy and divergence between features in high dimensional spaces. Comparisons to other entropy based techniques such as MI with histogram plug-in estimates showed that the technique could better exploit differences in multivariate feature distributions without suffering the curse of dimensionality in high dimensions. We presented a technique to accelerate MST computation for large number of feature vectors and reduce its memory complexity to the point where we could rapidly compute the MST length for data sets containing up to 1 million points in 8 dimensional space. For a US breast image registration application, we showed that the MST-based algorithm gave lower misregistration errors than single pixel MI methods.

It is known that color information offers distinct performance advantages for image retrieval and in particular US image registration [69, 20, 32, 63]. Application of our MST–based image matching techniques to color images, such as US Doppler blood flow color images with other color flow or gray scale images is a worthwhile 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 US images. Finally, specification of the best value of α in the α-entropy similarity criteria is an open issue that needs further investigation.

7 Acknowledgments

We thank 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 semiautomatic methods for image annotation and retrieval in QBIC", *Proc. SPIE Storage and Retrieval for Image and Video Databases III*, pp. 24-35, 1995.

[3] Baraniuk R, Flandrin P, Jensen AJEM, and Michel O, "Measuring time frequency information content using the Rényi entropies," *IEEE Trans. on Inform. Theory*, vol. IT-47, no. 4, , April 2001.

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

[5] Beardwood J, Halton JH, and Hammersley JM, "The shortest path through many points," *Proc. Cambridge Philosophical Society*, vol. 55, pp. 299–327, 1959.

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

[7] Chellappa R, Wilson CL, and Sirohey S, "Human and machine recognition of faces: A survey," *IEEE Proceedings*, vol. 83, no. 5, pp. 705–740, 1995.

[8] Cloude SR and Pottier E, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 75, pp. pp. 68–78, 1997.

[9] Cristiani N and Shaw-Taylor J, *Suport Vector Machines and other kernel-based learning methods*, Cambridge U. Press, 2000.

[10] Csiszár I, "Information-type measures of divergence of probability distributions and indirect observations," *Studia Sci. Math. Hung.*, vol. 2, pp. 299–318, 1967.

[11] de Bonet JS and Viola P, "Structure driven image database retrieval," in *Advances in neural information processing*, volume 10, 1997.

[12] Dembo A and Zeitouni O,"Large deviations techniques and applications", *Springer-Verlag NY*, 1998.

[13] 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.

[14] 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.

[15] Equinox, Inc. *visible/LWIR face database*, 2003. `http://www.equinoxsensors.com/products/HID.html`.

[16] Erdogmus V, Prncipe J, and Vielva L, "Blind deconvolution with minimum rényi's entropy," in *EUSIPCO*, Toulouse, France, 2002.

[17] Frieden B and Bajkova AT, "Reconstruction of complex signals using minimum rényi information," in *Proc. of Meeting of Intl. Soc. for Optical Engin. (SPIE)*, volume 2298, 1994.

[18] 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.

[19] Gilles S, "Description and experimentation of image matching using mutual information," Technical report, Oxford University, 1996. `http://www-rocq.inria.fr/~gilles/IMMMI/mutual_info.ps.gz`.

[20] 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.

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

[22] Hero AO, Costa J, and Ma B, "Asymptotic relations between minimal graphs and alpha entropy," Technical Report 334, Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, Mar, 2003. `http://www.eecs.umich.edu/~hero/det_est.html`.

[23] Hero AO, Ma B, Michel O, and Gorman J, "Applications of entropic spanning graphs," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 85–95, Sept. 2002. `http://www.eecs.umich.edu/~hero/imag_proc.html`.

[24] 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 `http://www.eecs.umich.edu/~hero/det_est.html`.

[25] 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.

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

[27] 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.

[28] 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.

[29] Hero AO and Michel O, "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.

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

[31] Hoffman R and Jain AK, "A test of randomness based on the minimal spanning tree," *Pattern Recognition Letters*, vol. 1, pp. 175–180, 1983.

[32] 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.

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

[34] Hyvärinen A, *Fast ICA Code*. `http://www.cis.hut.fi/projects/ica/fastica/`.

[35] Jenkinson M, Bannister P, Brady M, and Smith S, "Improved methods for the registration and motion correction of brain images," Technical report, Oxford University, 2002.

[36] Kieu T and Viola P, "Boosting image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[37] 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.

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

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

[40] 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.

[41] 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.

[42] Leventon ME and Grimson WEL, "Multi-modal volume registration using joint intensity distributions," Technical report, MIT AI Laboratory, 1998. `http://www.ai.mit.edu/projects/vision-surgery`.

[43] 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.

[44] Linde Y, Buzo A and Gray RM, "An algorithm for vector quantization design", *IEEE Transactions on Communications*, vol. 28, pp. 84-95, January 1980.

[45] Liu C and Wechsler H, "Comparative assessment of independent component analysis," in *Proc. the 2nd International Conference on Audioand Video-based Biometric Person Authentication*, pp. 22–24, Washington D. C., March 1999.

[46] Liu Y, Collins RT, and Rothfus WE, "Robust midsagittal plane extraction from coarse, pathological 3d images," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132–137, 2000.

[47] 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.

[48] 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.

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

[50] 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.

[51] 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.

[52] 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.

[53] Neemuchwala HF, Hero AO, and Carson PL, "Image registration using entropic spanning graphs", *Proc.of 36th Asilomar Conf Signals, Systems and Computers, Pacific Grove, CA*, Nov 2002.

[54] Neemuchwala HF, Hero AO, and Carson PL, "Fast algorithms for Minimum spanning tree construction", *Technical Report CSPL Communications and Signal Processing Laboratory*, The University of Michigan, 48109-2122, 2002

[55] 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.

[56] 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

[57] 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.

[58] Prim RC, "Shortest connection networks and some generalizations," *Bell Syst. Tech. Journ.*, vol. 36, pp. 1389–1401, 1957.

[59] Rangarajan A, Hsiao IT, and Gindi G, "Integrating anatomical priors in ect reconstruction via joint mixtures and mutual information," in *IEEE Medical Imaging Conference and Symposium on Nuclear Science*, volume III, Oct. 1998.

[60] Rényi A, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.*, volume 1, pp. 547–561, 1961.

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

[62] 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.

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

[64] Srivastava A, Lee AB, Simoncelli EP, and Zhu SC, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, , Jan. 2003.

[65] Steele JM, *Probability theory and combinatorial optimization*, volume 69 of *CBMF-NSF regional conferences in applied mathematics*, Society for Industrial and Applied Mathematics (SIAM), 1997.

[66] 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.

[67] Stoica R, Zerubia J, and Francos JM, "Image retrieval and indexing: A hierarchical approach in computing the distance between textured images," in *IEEE Int. Conf. on Image Processing*, Chicago, Oct. 1998.

[68] Stoica R, Zerubia J, and Francos JM, "The two-dimensional wold decomposition for segmentation and indexing in image libraries," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, Seattle, May 1998.

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

[70] 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.

[71] 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.

[72] Vasconcelos N and Lippman A, "A Bayesian framework for content-based indexing and retrieval," in *IEEE Data Compression Conference*, Snowbird, Utah, 1998. `http://nuno.www.media.mit.edu/people/nuno/`.

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

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

[75] Williams WJ, Brown ML, and Hero AO, "Uncertainty, information, and time-frequency distributions," in *Proc. of Meeting of Intl. Soc. for Optical Engin. (SPIE)*, volume 1566, pp. 144–156, 1991.

[76] Yang MH, Kriegman DJ, and Ahuja N, "Detecting faces in images: A survey," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 24, no. 1, pp. 24–58, Jan 2002.

[77] Yukich JE, "robability theory of classical Euclidean optimization", volume 1675 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1998.

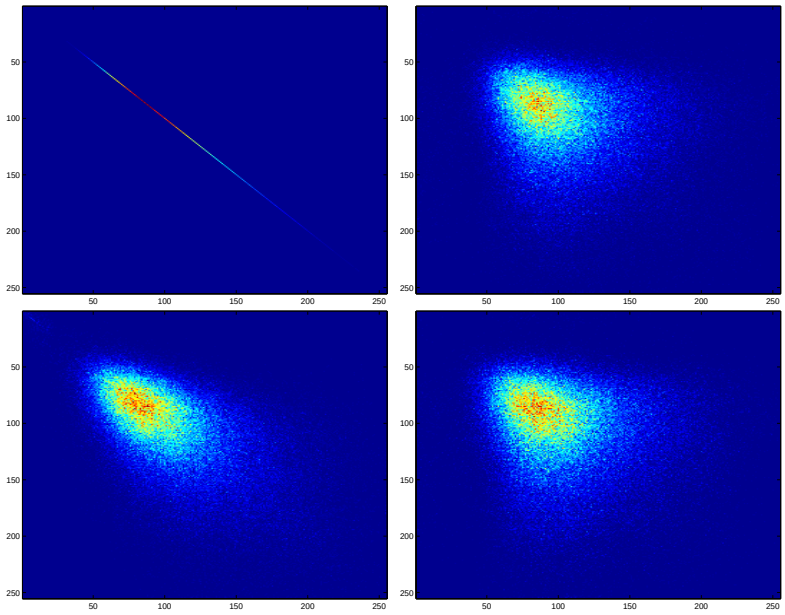

Figure 1: Joint coincidence histograms for single-pixel gray level features. Both horizontal and vertical axes of each panel are indexed over the gray level range of 0 to 255. Top left: joint histogram scatter plot for the case that reference image (X_i) and secondary image (X_j) are the same slice of the US image volume (Case 142) at perfect 0^o alignment ($X_j = X_i$). Bottom Left: same as top left except that reference and secondary are misaligned by 8^o relative rotation as in Fig. 2. Top right: same as top left except that the reference and secondary images are from adjacent (2mm separation) slices of the image volume. Bottom right: same as bottom left except that images are misaligned by 8^o relative rotation.

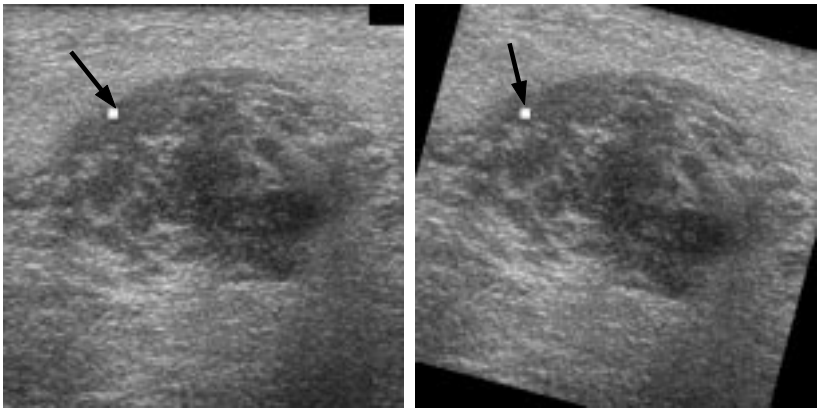

Figure 2: Single-pixel gray level coincidences are recorded by counting number of co-occurences of a pair of gray level in the reference (left) and in the secondary (right) images at a pair of homologous pixel locations. Here the secondary image (right) is rotated by 15^o relative to the reference image (left).

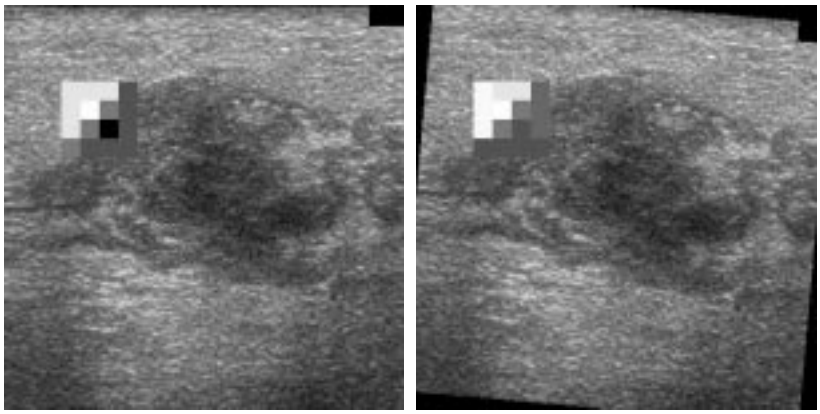

Figure 3: Local tags features applied to image registration. Each pixel is labeled by a tag type. Occurrences and coincidences of tag labels can be mapped to a coincidence histogram like Fig. 1.

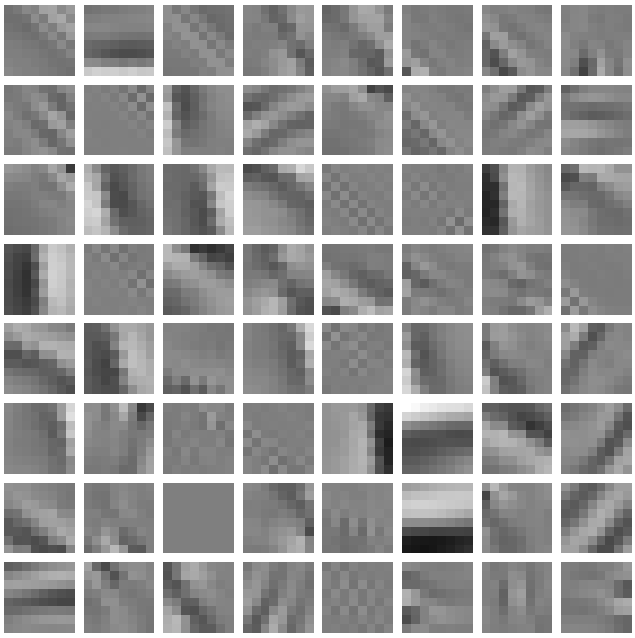

Figure 4: 8×8 *ICA basis set obtained from training on randomly selected* 8×8 *blocks in 10 ultrasound breast images. Features extracted from an image are the 64-dimensional vectors obtained by projecting* 8×8 *sub-images of the image on the ICA basis.*

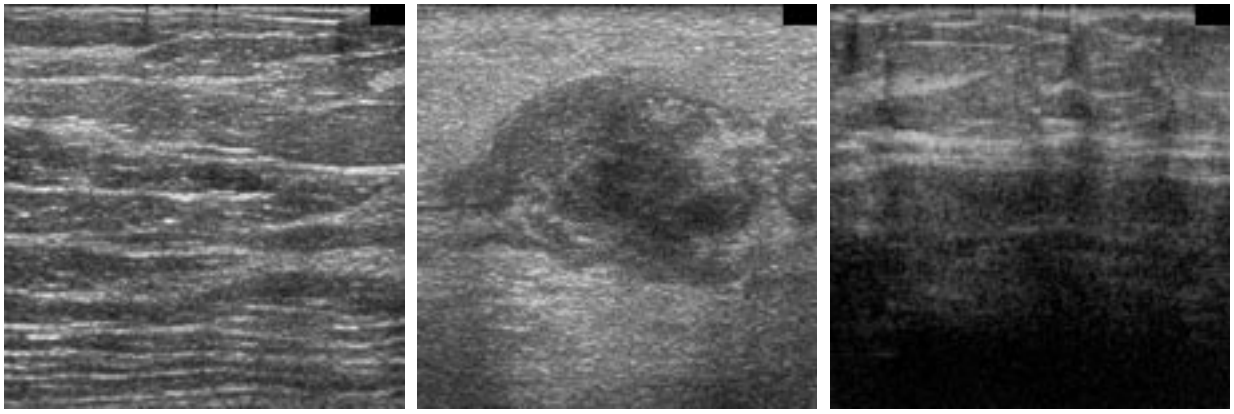

Figure 5: Three ultrasound (US) breast scans. From left to right are: Case 151, Case 142 and Case 162.

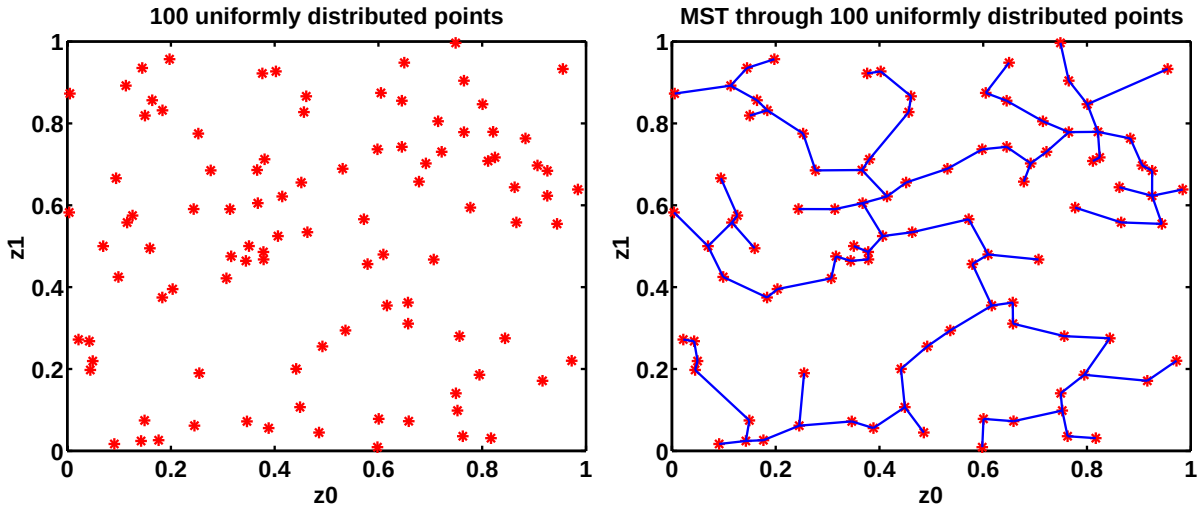

Figure 6: A set of $n = 100$ points $\{Z_i\}$ in the plane (left) and the corresponding Minimal Spanning Tree (MST) (right).

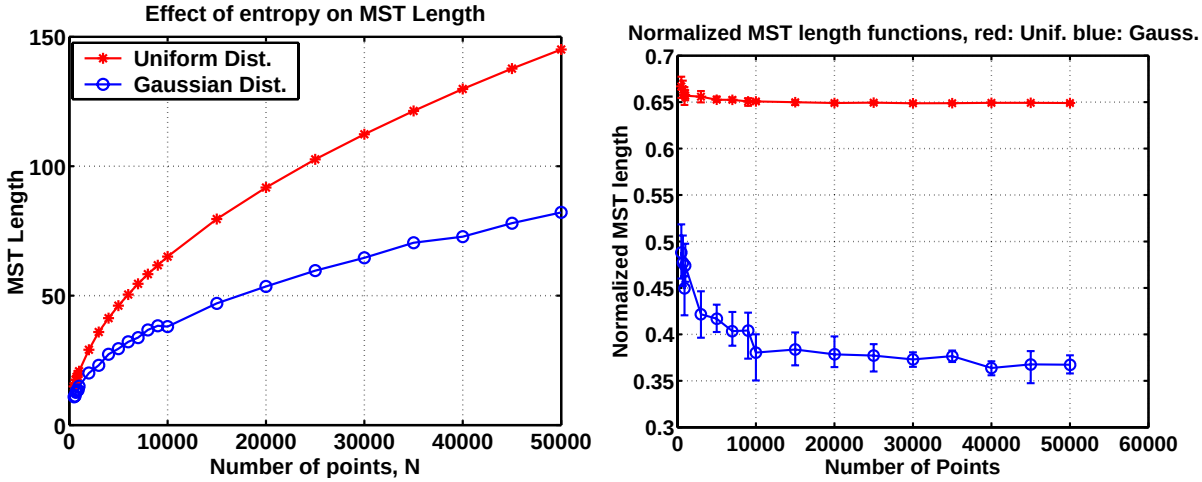

Figure 7: Length functions L_n of MST implemented with $\gamma = 1$ (left) and $L_n/\sqrt{n}$ (right) as a function of n for the uniform and normal distributed points in Fig. 6.

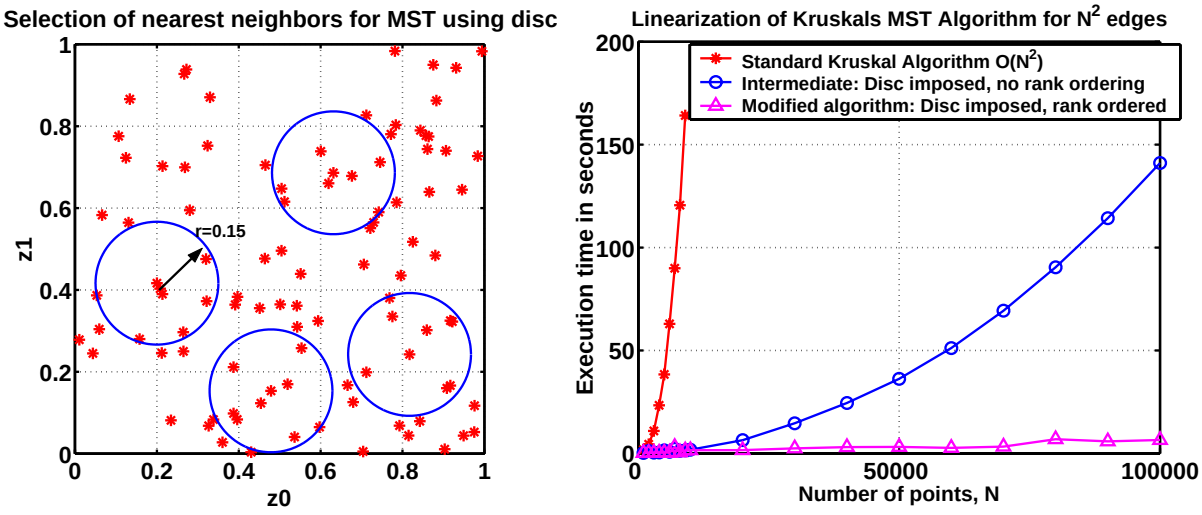

Figure 8: Disc-based acceleration of Kruskals MST algorithm from $n^2 \log n$ to $n \log n$ (left) and comparison of computation time for Kruskal's standard MST algorithm with respect to our accelerated algorithm (right)

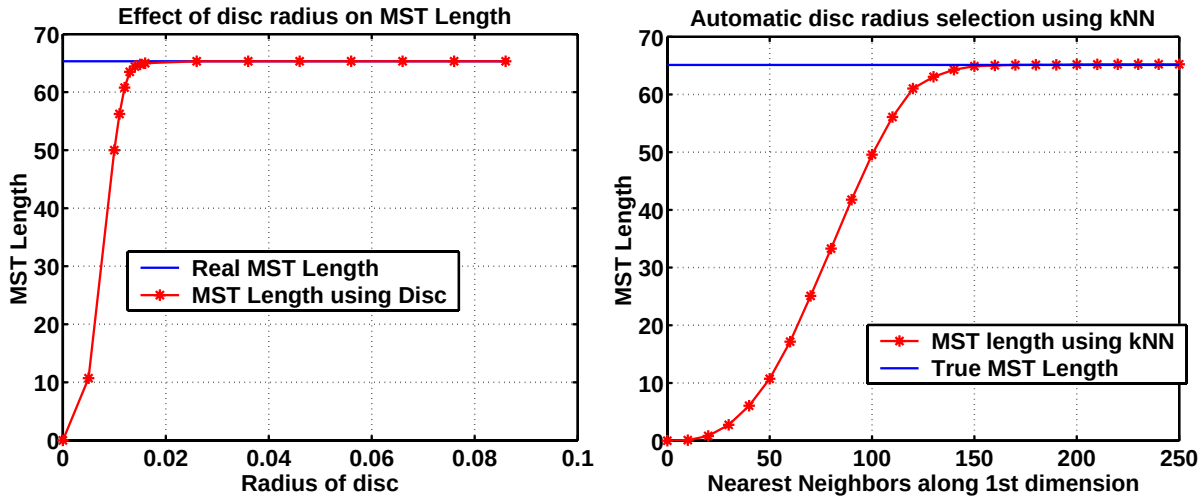

Figure 9: Bias of the $n \log n$ MST algorithm as a function of radius parameter (left) and as a function of the number of nearest neighbors (right) for uniform points in the unit square.

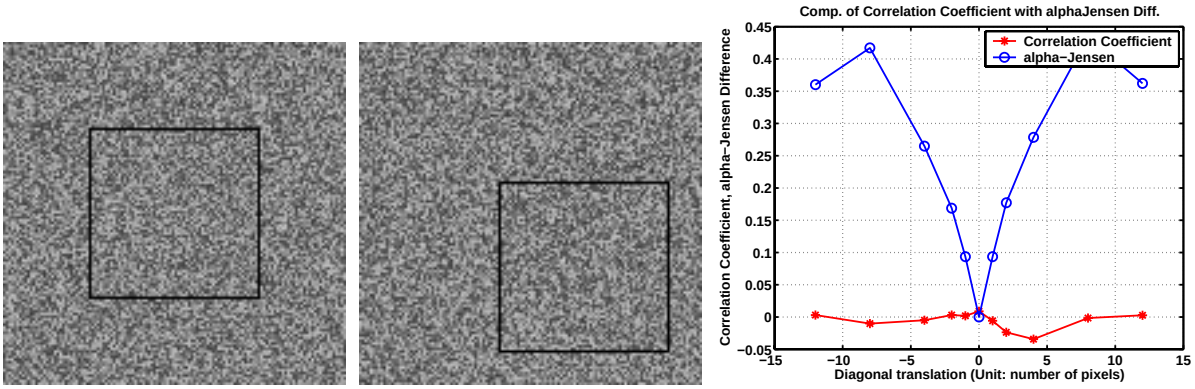

Figure 10: Square immerssed in homogeneous background noise (primary image on left). Translated square in noise (secondary image in center). Plot of αJensen difference and correlation coefficient as a function of translation of secondary image along diagonal (right).

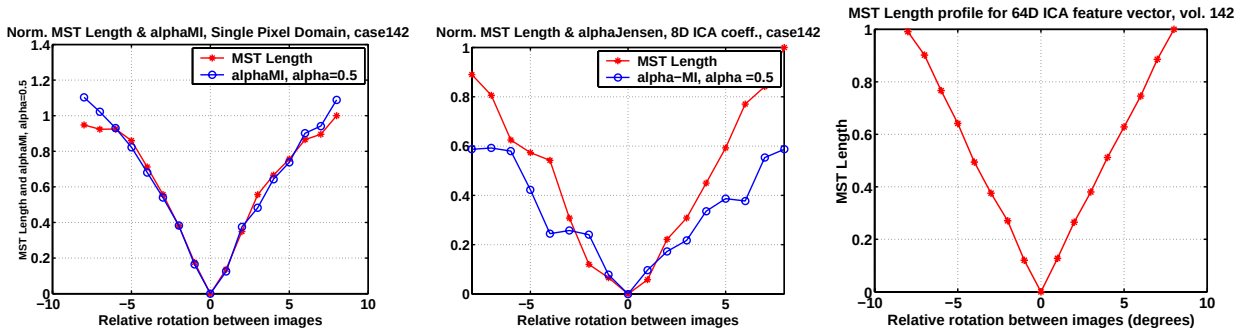

Figure 11: Profiles of image mathcing criteria for registration of US breast images taken from two slices of the image volume for Case 142: MST-based α-Jensen and histogram-based α-MI for single pixel features (left); MST-based α-Jensen and histogram-based α-MI for 8D ICA features (center); and MST-based α-Jensen for 64D ICA feature vectors (right)

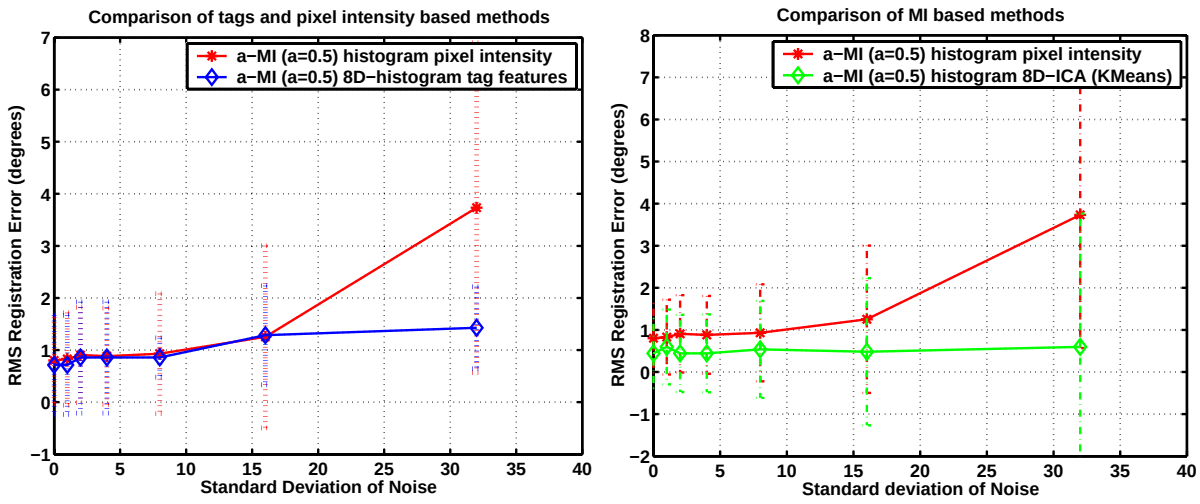

Figure 12: Effect of additive Gaussian noise on the rms of the peak position of the α-MI objective function estimated using histograms on single-pixel and feature coincidence trees of 8×8 tag features (left) and feature coincidence thistograms on discrete ICA (8D) features (right). These plots are based on 250 repeated experiments from within the breast US database. The two slices studied are separated by 2mm. Search was restricted to a maximum rotation angle of plus or minus 15^o.

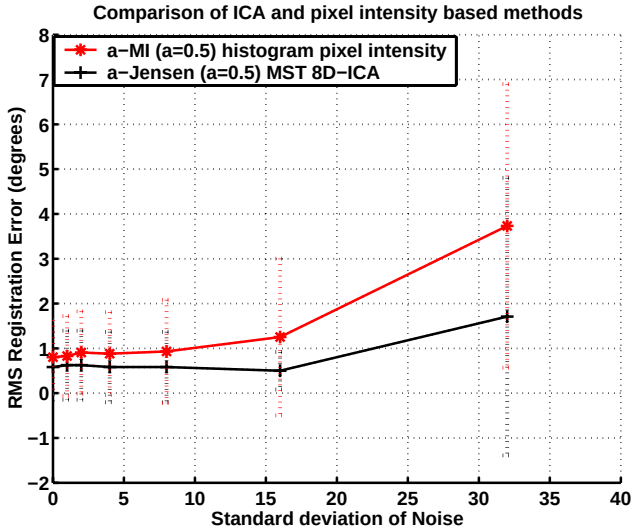

Figure 13: Effect of additive noise on the peak of the α- MI objective function estimated using histograms on single pixels vs. α-Jensen function estimated using histograms on single-pixels and 8D discrete and continuous ICA features.

Figure 14: Sampling of faces in the Equinox V/LWIR face database [15]. The database consists of 100 individual faces at various illumination, pose and facial expression configurations. Each visible-light image is co-registered to infrared counterpart by the camera.

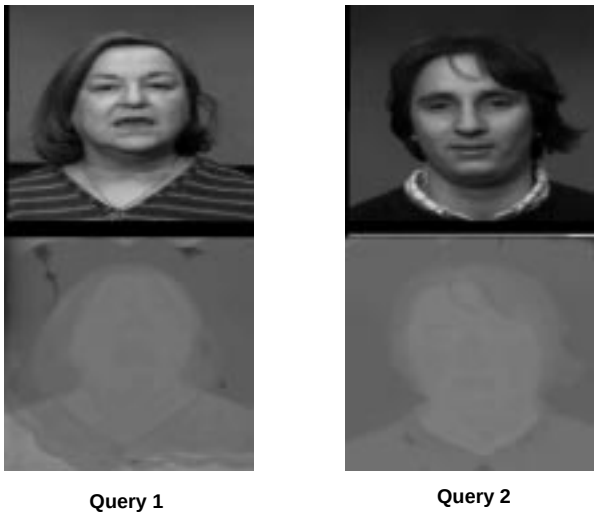

Figure 15: Two examples of queries taken from the Equinox face database.