Feature coincidence trees for registration of ultrasound breast images

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

Abstract—**Registration of an image, the query or reference, to a database of rotated and translated exemplars constitutes an important image retrieval and indexing application which arises in biomedical imaging, digital libraries, georegistration, and other areas. Two important issues are the specification of a class of discriminatory and generalizable image features and determination of an appropriate image-dissimilarity measure to rank the closeness of the query image with respect to images in the database. The theoretically best set of features and dissimilarity measure are those which can be implemented with the lowest misregistration error rate. In this paper we study a method based on feature discrimination using feature coincidence trees and mutual α-information measures of feature correlation. Feature coincidence trees represent the commonality between pairs of images using joint histograms of many simple features, or tags, which are organized in a data structure similar to that Amit and Geman's randomized trees for shape recognition. The mutual alpha-information measure is a ranking discriminant applied to the joint histograms which is motivated by a large deviations framework for detection error rates. We illustrate the methodology in the context of registering ultrasound scans of human breast images.**

I. INTRODUCTION

The focus application of this paper is co-registration of a pair of ultrasound images of the breast, called the reference and the secondary images, respectively. Accurate registration of 3D breast ultrasound image volumes is an essential part of whole breast imaging for detection of asymptomatic breast lesions. Such lesions are missed by community practitioners in up to 45% of women with dense breasts. The image registration problem falls under the general area of indexing databases of images $\mathcal{X} = \{X_i\}_{i=1}^{K}$ for the purposes of finding the best match to a reference image X_0. In this context the reference corresponds to a reference image and the database corresponds to a set of transformed versions of the secondary image, e.g. rotation and translation. To date the most effective methods for medical image registration have been pixel and voxel based and include: color histogram matching, texture matching using cross correlation, and, more recently, mutual information maximization on pixel coincidence histograms [11] as used in Radiology at the University of Michigan (the MIAMI-Fuse© registration algorithm). While these methods are adequate for some applications, they are overly sensitive to spurious image components such as speckle, shear/compressive tissue deformation, and shadowing which are ubiquitous to ultrasound breast imaging.

We present an approach to image registration which gets around the disadvantages of pixel based registration and indexing techniques such as the ones previously mentioned. The key

to our approach is the inclusion of highly specific image features and use of a generalized information divergence matching criterion based on the Chernoff bound of detection theory. In this approach we select the most relevant and robust features (curves, edges, textures and simple spatial relations) from a large breast image database. These features are organized in an efficient hierarchical database, which we call a feature tree. To register a target image to a reference image, coincidence histograms of the features captured by the feature tree for the reference and target images are constructed. The coincidence histograms are used to register the images through a generalized mutual information divergence measure, called the mutual α-information. This algorithm specializes to the standard mutual information (MIAMI-Fuse©) algorithm when the features are the set of single-pixel gray levels and $\alpha \rightarrow 1$. 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 captures 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 extraction of two types of low dimensional local features, called tags, to populate the feature tree; one 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 used by Geman and Koloydenko [5]. ICA is an iterative method which is closely related to the projection pursuit technique of non-linear regression and was applied to images by Olshausen, Hyvärinen and others [8], [6]. 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. For clarity of presentation we focus on registration of 2D slices. The extension to 3D voxel registration is straightforward but will be presented elsewhere.

II. 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 the secondary image 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 histogram $f(Z_i, Z_j)$ and the marginal histograms $f(Z_i)$, $f(Z_j)$. The similarity between features Z_0 and Z_i can be gauged by the dif-

ference between $f(Z_i, Z_0)$ and the product $f(Z_i)f(Z_0)$ which measures statistical dependence or mutual information (MI). The mutual α-information is defined as the α-divergence of fractional order $\alpha \in [0, 1]$ between $f(Z_i, Z_0)$ and $f(Z_i)f(Z_0)$ [2]

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

The α-divergence is equal to the Hellinger distance squared when $\alpha = 1/2$, and to the Kullback-Liebler (KL) divergence [7] when $\alpha \to 1$. The case $\alpha \to 1$ corresponds to the standard Shannon mutual information $\sum_{z_i, z_0} f(z_i, z_0) \log(f(z_i, z_0)/[f(z_i)f(z_0)])$ used by Viola and Wells [11] 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 [9], [4].

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

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

where $\beta(n)$ and $\alpha(n)$ are Type II and Type I errors, respectively, of the test of H_0 vs. H_1. Then [3]:

$$\liminf_{n \to \infty} \frac{1}{n} \log P_e(n) \qquad = \qquad (2)$$
$$- \sup_{\alpha \in [0,1]} \{(1 - \alpha)D_\alpha(f(Z_i, Z_0) \parallel f(Z_i)f(Z_0))\}$$

Thus the mutual α-information gives the asymptotically optimal rate of exponential decay of the error probability for testing H_0 vs H_1. Note that the Bayes classification error probability above is different from that defined by Vasconcelos etal [10] for optimal decision-theoretic indexing.

III. FEATURE COINCIDENCE TREES

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 1 and 2 illustrate the feature tree data structure for 4×4 subimages. The two images are each dropped down a feature tree and incidences and coincidences of features at all of the nodes of the two trees are counted. The counter is incremented for every coincidence of a particular feature pair occurring at a common position within each of the two images. This results in a 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 its arguments. These are then used in the mutual α-information formula (1) to come up with a registration score for the images.
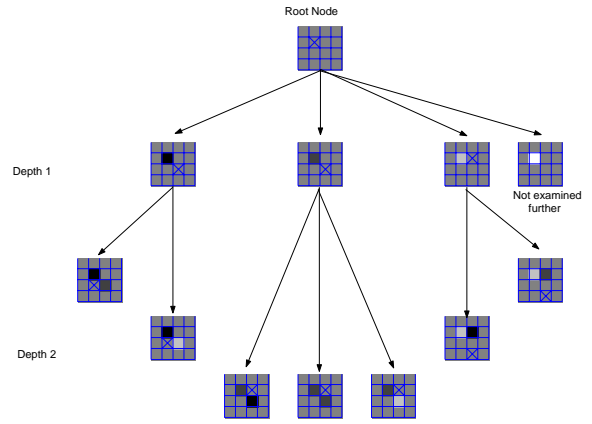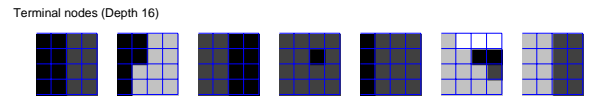

Fig. 1. *Part of feature tree data structure.*

Fig. 2. *Leaves of feature tree data structure.*

Our general feature selection and organization scheme is similar to the randomized tree classifier structures introduced by Amit and Geman [1] 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. 4×4 pixel neighborhoods, are captured by coding each pixel with labels derived from the tags. Non-local spatial features are then captured by cataloging pairs of tags which are in particular relative spatial configurations.

A. Tag Selection via Adaptive Thresholding

Adaptive thresholding is a quantization scheme described by Geman and Koloydenko [5] which was introduced to study the invariant characteristics of natural images. Let Δ be a positive granularity parameter. The quantized value assigned to a pixel within a 4×4 sub-image depends on the gray values of its neighbors. The darkest pixel(s) are assigned 0, the next brightest pixel(s) are assigned 0 if the difference is less than Δ and label 1 otherwise, the next brightest pixel(s) are assigned label 2 if the difference is less than Δ, and so on. Using this scheme on our ultrasound breast image database tags associated with the relatively uniform background areas (dark or bright) with small spatial variances are correctly classified as speckle and could be easily eliminated.

134 different tag types were identified among the samples. The number of tag types was controlled by selectively traversing the decision tree so that it was balanced, and by imposing constraints on the tag types. Tags were required to have fewer than 14 pixels of the same intensity, to avoid spurious tags arising from speckle. Also, tags were required to have at least two different intensity types within the center four pixels. Each image was block-quantized and dropped down the partition tree. The pairwise coincidences of the tag types at the leaves were

11

recorded in a histogram over 'tag space'. If an image block did not correspond to any of the tag types, then the pixel was not used in the image registration/retrieval algorithm.

B. Tag Selection via ICA

Here we adopt the ICA approach of [8] to infer an optimal basis which decomposes the ultrasound image X_i into sparse approximately statistically independent components $\{S_j\}$:

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

The basis elements $\{S_j\}$ are selected from an over-complete linearly dependent basis using the training set of ultrasound images. The coefficients a_{ij} are selected according to the minimum description length (MDL) criterion. Here ICA was implemented using Olshausen's `SPARSENET` code (available from `http://redwood.ucdavis.edu/bruno/`).
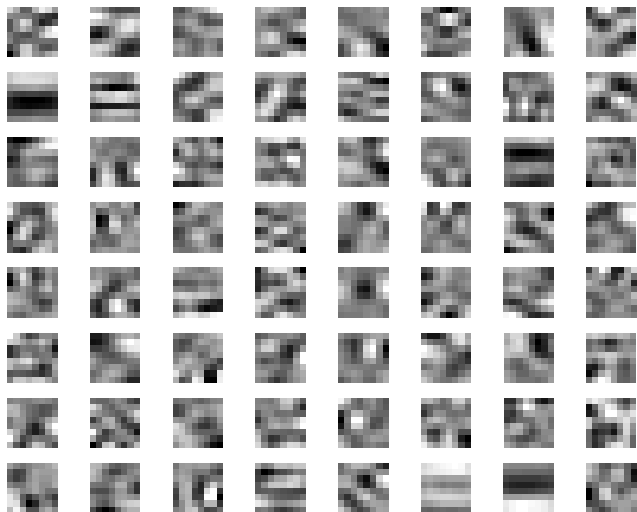

Fig. 3. *Estimated ICA basis set for ultrasound breast image database*

Using the ICA algorithm a set of 8×8 basis vectors were learned from 10 consecutive image slices extracted from a single ultrasound volume scan of the breast (Case151). 64 of the ICA basis vectors are shown in Fig. 3. Only basis elements that corresponded to distinct edge or texture information in the associated basis were retained, resulting in a total of approximately 256 different types. The tag assigned to a particular pixel is the feature type that has least Euclidean distance from the 8×8 sub-image centered there. An alternative method, not studied here, is to project sub-images of the reference and secondary images onto the ICA basis and use the residuals in the joint histogram.

C. Spatially Organized Tags

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

IV. NUMERICAL EXAMPLE

Fig. 4. Bar images with contrast 1.02, 1.07 and 1.78. Background is low variance white Gaussian while bar is uniform intensity.

We first illustrate the advantages of using feature coincidence trees for a simple example. In Fig. 4 three simple 60×75 bar images are shown at varying contrast ratios. The backgrounds of all three images are independent realizations of (truncated) Gaussian distributed (mean=128, variance =6) textures. The mutual information criterion ($\alpha \rightarrow 1$), based on both single pixel tags and on 4×4 adaptively thresholded tags, is calculated for each rotated and bilinearly interpolated pair of images over rotation angles falling between ± 5 degrees. Curves in Fig. 5 show that, as contrasted to the tag-based MI, the single pixel MI has non-monotonic objective function with peak value dependent on constrast. For the tag based MI 25 edge and corner tag types are used in this simulation which were determined by the adaptive thresholding. The plots suggest that significant improvement in MI resolution can be achieved by combining 4×4 tags and single pixel gray levels in the feature set.

V. REGISTRATION RESULTS

Three test cases were chosen from the breast database. These will be referred to as Case 151, Case 142 and Case 162, shown in Figs 6. The image slice chosen from Case 151 (image slice 40) had significant connective tissue structure. Case 142 (slice 35) had a distinct malignant tumor, while Case 162 (slice 60) was degraded due shadowing. We simulated the decorrelating effect of speckle by registering a slice to a rotated version of a proximal but different slice (approximately 2mm away along the depth of the scan). Shown in Table I are resolution optimizing α values and resultant optimal peak resolution values of the mutual α-information as measured by an estimate of curvature of $(1 - \alpha)D_\alpha(f(Z_i, Z_0)\|f(Z_i)f(Z_0))$ in th evicinity of its peak over i. Note that the resolution optimizing value of α are close to 1 only in three out of the 12 cases studied. Note also that all of
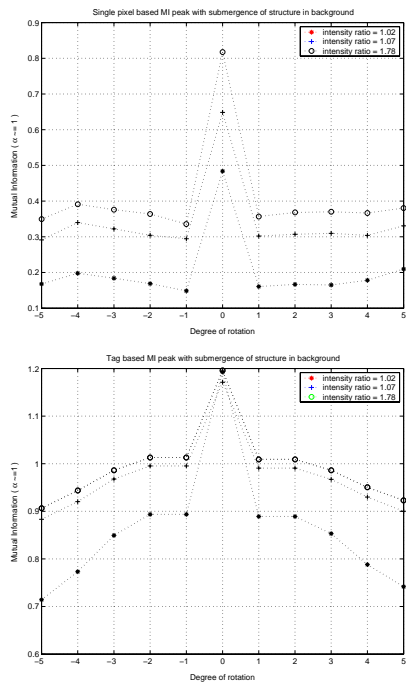
Fig. 5. Upper curves are single pixel based MI trajectories while lower curves are 4×4 tag based MI trajectories for bar images.

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

TABLE I

NUMERATOR =OPTIMAL VALUES OF α AND DENOMINATOR = MAXIMUM
RESOLUTION OF MUTUAL α-INFORMATION FOR REGISTERING VARIOUS
IMAGES (CASES 151, 142, 162) USING VARIOUS FEATURES (PIXEL, TAG,
SPATIAL-TAG, ICA). 151/8, 151/16, 151/32 CORRESPOND TO ICA
ALGORITHM WITH 8, 16 AND 32 BASIS ELEMENTS RUN ON CASE 151.

the feature based methods have significantly higher resolution than the standard single pixel (column labeled "pixel" in Table I) registration method. Systematic methods for determining the optimal value of α is an area for future work as is the issue of optimally combining the gray level and tag features studied here.

REFERENCES

[1] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Computation*, vol. 9, pp. 1545–1588, 1997.
[2] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, pp. 349–369, 1989.
[3] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer-Verlag, NY, 1998.
[4] M. N. Do and M. Vetterli, "Texture similarity measurement using kullback-liebler distance on wavelet subbands," in *IEEE Int. Conf. on Image Processing*, pp. 367–370, Vancouver, BC, 2000.
[5] D. Geman and K. A, "Invariant statistics and coding of natural microimages," in *IEEE Workshop on Statist. Computat. Theories of Vision*, Fort Collins, CO, June 1999.
[6] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 1999.
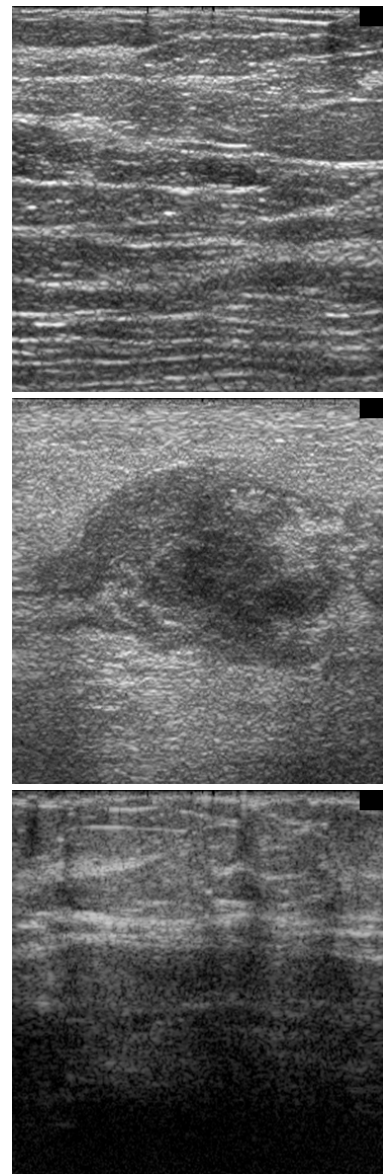[7] S. Kullback and R. Liebler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

Fig. 6. Three ultrasound breast scans. From top to bottom are: case 151, case 142 and case 162.

[8] M. Lewicki and B. Olshausen, "Probabilistic framework for the adaptation and comparison of image codes," *J. Opt. Soc. Am.*, vol. 16, no. 7, pp. 1587–1601, 1999.
[9] R. Stoica, J. Zerubia, and J. M. Francos, "Image retrieval and indexing: A hierarchical approach in computing the distance between textured images," in *IEEE Int. Conf. on Image Processing*, Chicago, 1998.
[10] N. Vasconcelos and A. Lippman, "Bayesian representations and learning mechanisms for content based image retrieval," in *SPIE Storage and Retrieval for Media Databases 2000*, San Jose, CA, 2000. http://nuno.www.media.mit.edu/people/nuno/.
[11] P. Viola and W. Wells, "Alignment by maximization of mutual information," in *Proc. of 5th Int. Conf. on Computer Vision, MIT*, volume 1, pp. 16–23, 1995.