Multi-Attribute Non-initializing Texture Reconstruction Based Active Shape Model (MANTRA)

Robert Toth[1], Jonathan Chappelow[1], Mark Rosen[2], Sona Pungavkar[3], Arjun Kalyanpur[4], and Anant Madabhushi[1]

[1] Rutgers, The State University of New Jersey, New Brunswick, NJ, USA
[2] University of Pennsylvania, Philadelphia, PA, USA
[3] Dr. Balabhai Nanavati Hospital, Mumbai, India
[4] Teleradiology Solutions, Bangalore, India

Abstract. In this paper we present MANTRA (Multi-Attribute, Non-Initializing, Texture Reconstruction Based Active Shape Model) which incorporates a number of features that improve on the the popular Active Shape Model (ASM) algorithm. MANTRA has the following advantages over the traditional ASM model. (1) It does not rely on image intensity information alone, as it incorporates multiple statistical texture features for boundary detection. (2) Unlike traditional ASMs, MANTRA finds the border by maximizing a higher dimensional version of mutual information (MI) called combined MI (CMI), which is estimated from kNN entropic graphs. The use of CMI helps to overcome limitations of the Mahalanobis distance, and allows multiple texture features to be intelligently combined. (3) MANTRA does not rely on the mean pixel intensity values to find the border; instead, it reconstructs potential image patches, and the image patch with the best reconstruction based on CMI is considered the object border. Our algorithm was quantitatively evaluated against expert ground truth on almost 230 clinical images (128 1.5 Tesla (T) T2 weighted *in vivo* prostate magnetic resonance (MR) images, 78 dynamic contrast enhanced breast MR images, and 21 3T *in vivo* T1-weighted prostate MR images) via 6 different quantitative metrics. Results from the more difficult prostate segmentation task (in which a second expert only had a 0.850 mean overlap with the first expert) show that the traditional ASM method had a mean overlap of 0.668, while the MANTRA model had a mean overlap of 0.840.

1   Introduction

The Active Shape Model (ASM) [1] and Active Appearance Model (AAM) [2] are both popular methods for segmenting known anatomical structures. The ASM algorithm involves an expert initially selecting landmarks to construct a statistical shape model using Principal Component Analysis (PCA). A set of intensity values is then sampled along the normal in each training image. During segmentation, any potential pixel on the border also has a profile of intensity values sampled. The point with the minimum Mahalanobis distance between the mean training intensities and the sampled intensities presumably lies on the object border. Finally, the shape model is updated to fit these landmark points, and the process repeats until convergence. However, there are several limitations with traditional ASMs with regard to image segmentation. (1) ASMs require

an accurate initialization and final segmentation results are sensitive to the user defined initialization. (2) The border detection requires that the distribution of intensity values in the training data is Gaussian, which need not necessarily be the case. (3) Limited training data could result a near-singular covariance matrix, causing the Mahalanobis distance to not be defined.

Alternatives and extensions to the traditional ASM algorithm have been proposed [3,4,5]. An interesting alternative classifier-based method was proposed in [3] where Taylor-series gradient features are calculated and the features that improve classification accuracy during training are used during segmentation. Then, the classifier is used on the features of the test image to determine border landmark points. The classifier approach provides an alternative to the Mahalanobis distance for finding landmark points, but requires an offline feature selection stage. The segmentation algorithm presented in [5] gave very promising results as it implemented a multi-attribute based approach and also allowed for multiple landmark points to be incorporated; however, it still relies on the Mahalanobis distance for its cost function which might not be optimal.

MANTRA differs from the traditional AAM in that AAMs employ a global texture model of the entire object, which is combined with the shape information to create a general appearance model. For several medical image tasks however, local texture near the object boundary is more relevant to obtaining an accurate segmentation instead of global object texture, and MANTRA's approach is to create a local texture model for each individual landmark point.

In this paper we present a novel segmentation algorithm: Multi-Attribute Non-Initializing Texture Reconstruction Based ASM (MANTRA). MANTRA comprises of a new border detection methodology, from which a statistical shapes model can be fitted. In the following page we briefly describe several novel aspects of MANTRA and several ways it overcomes limitations associated with the traditional approach.

(a) Local Texture Model Reconstruction: To overcome the limitations associated with using the Mahalanobis distance, MANTRA performs PCA on pixel neighborhoods surrounding the object borders of the training images to create a local texture model for each landmark point. Any potential border landmark point of the test image has a neighborhood of pixels sampled, and the PCA-based local texture model is used to reconstruct the sampled neighborhood in a manner similar to AAMs [2]. These training reconstructions are compared to the original pixels values to detect the object border, where the location with the best reconstruction is presumably the object border.

(b) Use of Multiple Attributes with Combined Mutual Information: Since mutual information (MI), a metric that quantifies the statistical interdependence of multiple random variables, operates without assuming any functional relationship between the variables [6], we employ it as a robust image similarity measure to compare the reconstructions to the original pixel values. In order to overcome the limitations of using image intensities to represent the object border, 1st and 2nd order statistical features [7,8] are generated from each training image. These features have been previously shown to be useful in both computer aided diagnosis systems and registration tasks [7,8,9,10]. To integrate multiple image attributes, we utilize Combined MI (CMI) because of its property to incorporate non-redundant information from multiple sources, and its previous

success in complementing similarity measures with information from multiple feature calculations [10,11,12]. Since CMI operates in higher dimensions, histogram-based estimation approaches would become too sparse when more than 2 features are used. Therefore, we implement the k nearest neighbor (kNN) entropic graph technique to estimate the CMI [13]. The values are plotted in a high dimensional graph, and the entropy is estimated from the distances to the k nearest neighbors, which is subsequently used to estimate the MI value.

(c) Non-requirement of Model Initialization: Similarly to several other segmentation schemes, MANTRA is cast within a multi-resolution framework, in which the shape is updated in an iterative fashion and across image resolutions [14]. At each resolution increase, the area of the search neighborhood decreases, allowing only fine adjustments to be made in the higher resolution. This overcomes the problem of noise near the object boundary and makes MANTRA robust to different initializations.

The experiments were performed on nearly 230 images comprising 3 MR protocols and 2 body regions. Three different 2D models were tested: MANTRA, the traditional ASM, and ASM+MI (a hybrid with aspects of both MANTRA and ASM). Quantitative evaluation was performed against expert delineated ground truth via 6 metrics.

2   Brief Overview of MANTRA

MANTRA comprises of a distinct training and segmentation step (Figure 1).

Training
1. Select Landmark Points of object border on each training image.
2. Generate Shape Model Using PCA as in traditional ASMs [1].
3. Generate Texture Features: K statistical texture feature scenes are generated for each of the N training images, which include gradient and second order co-occurrence features [7,8]. Then, a neighborhood surrounding each landmark point is sampled from each of the K feature scenes for all N training images.
4. Generate Texture Model Using PCA: Each landmark point has K texture models generated by performing PCA on all N neighborhood vectors for each given feature.

Segmentation
5. Overlay Mean Shape on test image to anchor the initial landmark points.
6. Generate Texture Features: The same texture features used for training (gradient and second order co-occurrence [7,8]) are generated from the test image.
7. Reconstruct Patches Using Texture Model: A neighborhood is searched near each landmark point, and the search area size is inversely related to the resolution, so that only fine adjustments are made at the highest resolution. For any potential border landmark point, its surrounding values are reconstructed from the training PCA models.
8. Use kNN Entropic Graphs to Maximize CMI: kNN entropic graphs [13] are used to estimate entropy, and then CMI. The location with the highest CMI value between its reconstructed values and its original values is the new landmark point.
9. Fit Shape Model To New Landmark Points: Once a set of new landmarks points have been found, the current shape is updated to best fit these landmark points [1], and
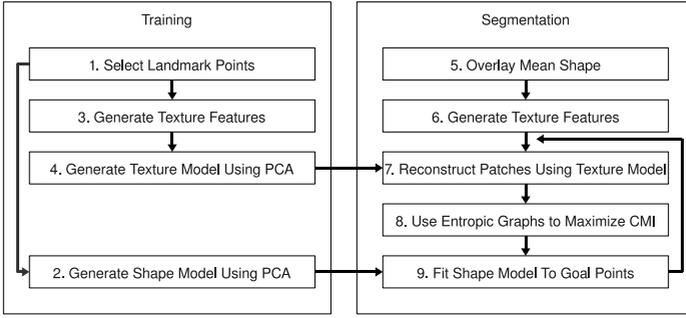
Fig. 1. The modules and pathways comprising MANTRA, with the training module on the left and the testing module on the right

constrained to +/- 2.5 standard deviations from the mean shape. The resolution is then doubled at each iteration until convergence is obtained.

3 Methodology

This section is focused on Steps 4, 6-9 of the MANTRA scheme, as Steps 1-3, 5 are identical to corresponding steps in [1].

3.1 Generating Texture Models

We define the set of N training images as $S_{tr} = \{\mathcal{C}^\alpha \mid \alpha \in \{1, \ldots, N\}\}$, where $\mathcal{C}^\alpha = (C, f^\alpha)$ is an image scene where $C \in \Re^2$ represents a set of 2D spatial locations and $f^\alpha(c)$ represents a function that returns the intensity value at any $c \in C$. For $\forall \mathcal{C}^\alpha \in S_{tr}, X^\alpha \subset C$ is a set of M landmark points manually delineated by an expert, where $X^\alpha = \{c_m^\alpha \mid m \in \{1, \ldots, M\}\}$. For $\forall \mathcal{C}^\alpha \in S_{tr}, K$ features scenes $\mathcal{F}^{\alpha,k} = (C, f^{\alpha,k}), k \in \{1, \ldots, K\}$ are then generated. For our implementation, we used the gradient magnitude, Haralick inverse difference moment, and Haralick entropy texture features [7,8]. For each training image $\mathcal{C}^\alpha$, and each landmark point c_m^α, a κ-neighborhood $\mathcal{N}_\kappa(c_m^\alpha)$ (where for $\forall d \in \mathcal{N}_\kappa(c_m^\alpha), \| d - c_m^\alpha \|_2 \leq \kappa, c_m^\alpha \notin \mathcal{N}_\kappa(c_m^\alpha))$ is sampled on each feature scene $\mathcal{F}^{\alpha,k}$ and normalized. For each landmark point m and each feature k, the normalized feature values for $\forall d \in \mathcal{N}_\kappa(c_m^\alpha)$ are denoted as the vector $\mathbf{g}_m^{\alpha,k} = \left[f^{\alpha,k}(d) / \sum_d f^{\alpha,k}(d) \mid d \in \mathcal{N}_\kappa(c_m^\alpha) \right]$. The mean vector for each landmark point m and each feature k is given as $\bar{\mathbf{g}}_m^k = \left[\frac{1}{N} \sum_\alpha f^{\alpha,k}(d) \mid \alpha \in \{1, \ldots, N\}, d \in \mathcal{N}_\kappa(c_m^\alpha) \right]$ and the covariance matrix of $\mathbf{g}_m^{\alpha,k}$ over $\forall \alpha \in \{1 \ldots N\}$ is denoted as φ_m^k. Then, PCA is performed by calculating the Eigenvectors of φ_m^k and retaining the Eigenvectors that account for most ($\sim 98\%$) of the variation in the training data , denoted as $\mathbf{\Phi}_m^k$.

3.2 Reconstructing Local Image Texture

We define a test image as the scene $\mathcal{C}_{te}$, where $\mathcal{C}_{te} \notin S_{tr}$, and its corresponding K feature scenes as $\mathcal{F}^k, k \in \{1, \ldots, K\}$. The M landmark points for the current iteration

j are denoted as the set $X_{te} = \{c_m \mid m \in \{1, \ldots, M\}\}$. A γ-neighborhood $\mathcal{N}_\gamma$ (where $\gamma \neq \kappa$) is searched near each current landmark point c_m to identify a landmark point $\tilde{c}_m$ which is in close proximity to the object border. For $j = 1, c_m$ denotes the initialized landmark point, and for $j \neq 1, c_m$ denotes the result of deforming to $\tilde{c}_m$ from iteration $(j - 1)$ using the statistical shape model [1]. For $\forall e \in \mathcal{N}_\gamma(c_m)$, we sample a κ-neighborhood $\mathcal{N}_\kappa(e)$ on each feature scene $\mathcal{F}^k$ and normalize, denoted as the vector $\mathbf{g}_e^k = \{f^k(d)/\sum_d f^k(d) \mid d \in \mathcal{N}_\kappa(e)\}$. Then, for each e (which is a potential location for $\tilde{c}_m$), the K vectors $\mathbf{g}_e^k, k \in \{1, \ldots, K\}$ are reconstructed from the training PCA models, where the vector of reconstructed pixel values for feature k is given as

$$\mathcal{R}_e^k = \bar{\mathbf{g}}_m^k + \boldsymbol{\Phi}_m^k \cdot (\boldsymbol{\Phi}_m^k)^T \cdot (\mathbf{g}_e^k - \bar{\mathbf{g}}_m^k). \tag{1}$$

3.3 Identifying New Landmarks in 3 Models: ASM, ASM+MI, and MANTRA

We wish to compare three different methods for finding new landmark points. The first is the traditional ASM method, which minimizes the Mahalanobis distance. The remaining 2 methods utilize the Combined Mutual Information (CMI) metric to find landmark points. The MI between 2 vectors is a measure of how predictive they are of each other, based on their entropies. CMI is an extension of MI, where 2 sets of vectors can be compared intelligently by taking into account the redundancy between the sets [10]. For 2 sets of vectors $\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ and $\{\mathbf{B}_1, \ldots, \mathbf{B}_n\}$, where each $\mathbf{A}$ and $\mathbf{B}$ is a vector of the same dimensionality, the MI between them is given as $I(\mathbf{A}_1 \cdots \mathbf{A}_n, \mathbf{B}_1 \cdots \mathbf{B}_n) = H(\mathbf{A}_1 \cdots \mathbf{A}_n) + H(\mathbf{B}_1 \cdots \mathbf{B}_n) - H(\mathbf{A}_n \cdots \mathbf{A}_n \mathbf{B}_1 \cdots \mathbf{B}_n)$ where H denotes the joint entropy [10,12]. To estimate this joint entropy, we utilize k-nearest-neighbor (kNN) entropic graphs, where H is estimated from average kNN distance, the details of which can be found in [13].

1. ASM: To use the Mahalanobis distance with features, we averaged the Mahalanobis distance for each feature, which yields the mth landmark point of the ASM method as

$$\tilde{c}_m = \operatorname*{argmin}_{e \in \mathcal{N}_\gamma(c_m)} \frac{1}{K} \sum_{k=1}^{K} \left[(\mathbf{g}_e^k - \bar{\mathbf{g}}_m^k)^T \cdot (\varphi_m^k)^{-1} \cdot (\mathbf{g}_e^k - \bar{\mathbf{g}}_m^k) \right]. \tag{2}$$

2. MANTRA: The MANTRA method maximizes the CMI between the reconstructions and original vectors to find landmark points, so that the mth landmark point is given as

$$\tilde{c}_m = \operatorname*{argmax}_{e \in \mathcal{N}_\gamma(c_m)} I(\mathcal{R}_e^1 \ldots \mathcal{R}_e^K, \mathbf{g}_e^1 \ldots \mathbf{g}_e^K). \tag{3}$$

3. ASM+MI: Finally, to evaluate the effectiveness of using the reconstructions, the ASM+MI method [4] maximizes the CMI between $\mathbf{g}_e$ and $\bar{\mathbf{g}}_m$ instead of between $\mathbf{g}_e$ and $\mathcal{R}_e$, so that the mth landmark point is defined as

$$\tilde{c}_m = \operatorname*{argmax}_{e \in \mathcal{N}_\gamma(c_m)} I(\bar{\mathbf{g}}_m^1 \ldots \bar{\mathbf{g}}_m^K, \mathbf{g}_e^1 \ldots \mathbf{g}_e^K). \tag{4}$$

Table 1. Quantitative results for all test performed (ASM, ASM+MI, MANTRA) as mean $\pm$ standard deviation

Object	Method	Overlap	Sensitivity	Specificity	PPV	MAD	Hausdorff
Prostate with Intensities	MANTRA	.752±.118	.880±.115	.765±.131	.849±.113	4.3±2.1	11.6±5.0
	ASM+MI	.731±.128	.831±.130	.813±.151	.879±.139	4.5±2.2	12.3±5.8
	ASM	.668±.165	.737±.187	.855±.149	.903±.134	5.6±3.1	13.7±6.8
Prostate with Features	MANTRA	.840±.096	.958±.041	.784±.098	.873±.106	2.6±1.1	8.1±3.3
	ASM+MI	.818±.094	.925±.055	.796±.113	.881±.111	2.9±1.2	8.7±3.4
	ASM	.766±.144	.814±.163	.888±.087	.933±.099	3.6±1.9	10.0±3.8
Prostate	Expert 2	.858±.101	.961±.089	.778±.119	.886±.083	2.4±1.7	7.7±5.1
Breast	MANTRA	.925±.102	.952±.102	.935±.044	.970±.022	4.9±6.7	16.3±11.6
	ASM+MI	.925±.098	.954±.098	.930±.042	.968±.021	4.9±6.6	16.6±11.3
	ASM	.924±.104	.952±.104	.934±.041	.970±.020	5.0±7.1	16.9±12.3

4   Results

Our data consisted of 128 1.5 Tesla (T), T2-weighted *in vivo* prostate MR slices, 21 3T T1-weighted DCE *in vivo* prostate MR slices, and 78 1.5T T1-weighted DCE MR breast images. To evaluate our methods, a 10-fold cross validation was performed on each of the datasets for the MANTRA, ASM+MI, and ASM methods, in which 90% of the images were used for training, and 10% were used for testing, which was repeated until all images had been tested.

4.1   Quantitative Results

For nearly 230 clinical images, MANTRA, ASM, and ASM+MI were compared against expert delineated segmentations (Expert 1) in terms of 6 error metrics [7,15], where PPV and MAD stand for Positive Predictive Value and Mean Absolute Distance respectively. The segmentations of an experienced radiologist (Expert 1) were used as the gold standard for evaluation. Also shown in Table 1 is the segmentation performance of a radiologist resident (Expert 2) compared to Expert 1. Note that MANTRA performs comparably to Expert 2, and in 78% of the 18 scenarios (6 metrics, 3 tests), MANTRA performs better than ASM and ASM+MI. The scenarios in which it failed (specificity and PPV of the prostate) did not take into account false negative area. Using the proposed ASM+MI algorithm performed better than the ASM method but worse than the MANTRA method, suggesting that MI is a more effective metric than the Mahalanobis distance for border detection, but also justifying the use of the reconstructions in MANTRA. In addition, using statistical texture features improved the performance of all results, showing the effectiveness of the multi-attribute approach. For breast segmentation task, all 3 methods performed equivalently, indicating that our new method is as robust as the traditional ASM method in segmenting a variety of medical images.

4.2   Qualitative Results

In Figure 2 are shown the results of qualitatively comparing the ground truth in the first column (Figures 2 (a), (e), (i), and (m)), MANTRA in the second column (Figures 2 (b),
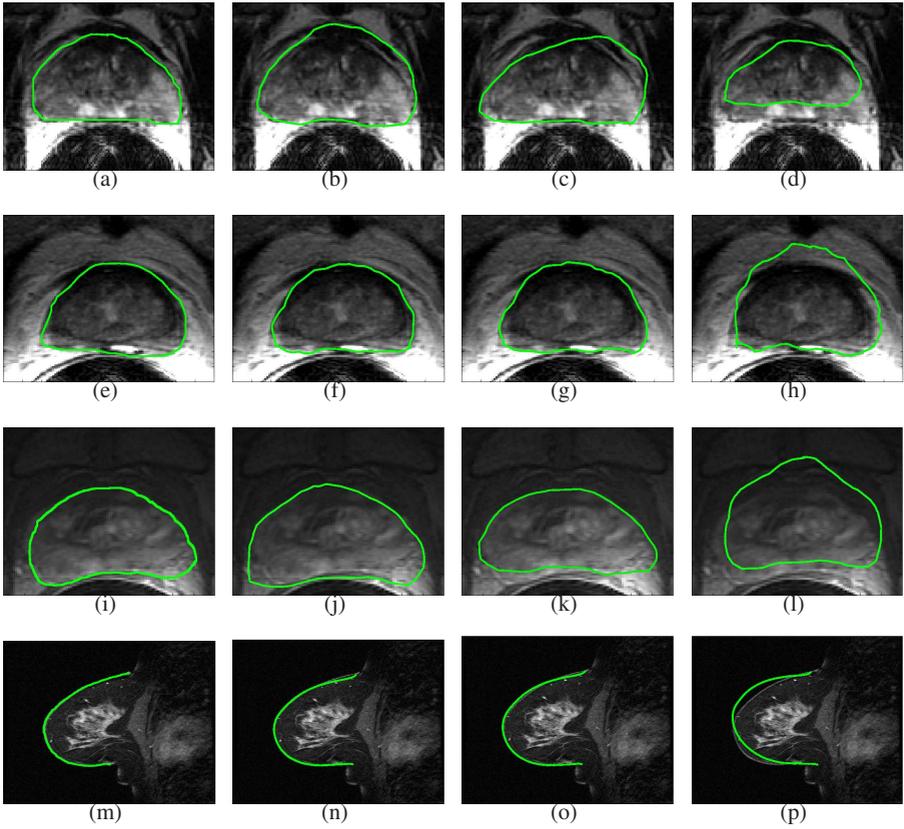
Fig. 2. The ground truth is shown in (a), (e), (i), and (m), MANTRA in (b), (f), (j), and (n), ASM+MI in (c), (g), (k), and (o), and ASM in (d), (h), (l), and (p). (a)-(h) show the results of the models on 1.5T T2-weighted prostate slices, in (i)-(l) are shown 3T T1-weighted prostate slices results, and finally in (m)-(p) are shown 1.5T DCE breast results.

(f), (j), and (n)), ASM+MI in the third column (Figures 2 (c), (g), (k), and (o)), and ASM in the fourth column (Figures 2 (d), (h), (l), and (p)). Figures 2 (a)-(h) show the results of the models on 1.5T T2-weighted prostate slices, Figures 2 (i)-(l) show 3T T1-weighted prostate slices results, and finally Figures 2 (m)-(p) show 1.5T DCE breast results. In all the cases, the MANTRA segmentation is most similar to the ground truth segmentation. The false edges that sometimes cause the models to deviate from the true prostate edge can be seen in Figures 2 (c) and (d), and in Figures 2 (i)-(l) the lack of a clear prostate edge at the top prevents the ASM+MI and ASM from finding the correct object border.

5   Concluding Remarks

We have presented a Multi-Attribute, Non-Initializing, Texture Reconstruction Based Active Shape Model (MANTRA) with the following strengths:

1. PCA-based texture models are used to better represent the border instead of simply using mean intensities as in the traditional ASM.
2. CMI is used as an improved border detection metric to overcome several inherent limitations with the Mahalanobis distance. The use of kNN entropic graphs makes it possible to compute CMI in higher dimensions.
3. Using multiple attributes gives better results than simply using intensities.
4. A multi-resolution approach is used to overcome initialization bias, and problems with noise at higher resolutions.

MANTRA was tested on over 230 clinical images, and outperformed the traditional ASM method. In addition, MANTRA was successful with different field strengths (1.5T and 3T) and on multiple protocols (DCE and T2). The incorporation of multiple texture features also increased results significantly, indicating that a multi-attribute approach is advantageous. Future work will attempt to discover and overcome limitations of the choice of features, and to extend MANTRA to be 3D (our tests show that a single CMI calculation for 2 neighborhoods of 64x64x10 pixels is on the order of 10^{-3} seconds, indicating that a 3D model can work in real time).

Acknowledgments

References

1. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
3. van Ginneken, B., Frangi, A.F., Staal, J.J., et al.: Active shape model segmentation with optimal features. IEEE Trans. Med. Imag. 21(8), 924–933 (2002)
4. Toth, R., Tiwari, P., Rosen, M., Kalyanpur, A., Pungabkar, S., Madabhushi, A.: A multimodal prostate segmentation scheme by combining spectral clustering and active shape models. In: SPIE, vol. 6914, pp. 69144S1–69144S12 (2008)
5. Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal shape and intensity cost path segmentation. IEEE Trans. Med. Imag. 26(8), 1115–1129 (2007)
6. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imag. 22(8), 986–1004 (2003)
7. Madabhushi, A., Feldman, M., Metaxas, D., Tomaszewski, J., Chute, D.: Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. IEEE Trans. Med. Imag. 24(12), 1611–1625 (2005)
8. Doyle, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: A boosting cascade for automated detection of prostate cancer from digitized histology. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 504–511. Springer, Heidelberg (2006)

9. Viswanath, S., Rosen, M., Madabhushi, A.: A consensus embedding approach for segmentation of high resolution in vivo prostate magnetic resonance imagery. In: SPIE (2008)

10. Chappelow, J., Madabhushi, A., Rosen, M., Tomaszeweski, J., Feldman, M.: A combined feature ensemble based mutual information scheme for robust inter-modal, inter-protocol image registration. In: ISBI 2007, pp. 644–647 (April 2007)

11. Tomazevic, D., Likar, B., Pernus, F.: Multifeature mutual information. In: Fitzpatrick, J.M., Sonka, M. (eds.) Proceedings of SPIE: Medical Imaging, vol. 5370, pp. 143–154 (2004)

12. Matsuda, H.: Physical nature of higher-order mutual information: Intrinsic correlations and frustration. Phys. Rev. E 62(3), 3096–3102 (2000)

13. Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Phys. Rev. E 69(6), 066138 (2004)

14. Cootes, T., Taylor, C., Lanitis, A.: Evaluating of a multi-resolution method for improving image search. In: Proc. British Machine Vision Conference, pp. 327–336 (1994)

15. Madabhushi, A., Metaxas, D.N.: Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Trans. Med. Imag. 22(2), 155–170 (2005)