Classification of cirrhotic liver in Gadolinium-enhanced MR images

Gobert Lee[a], Yoshikazu Uchiyama[a], Xuejun Zhang[a], Masayuki Kanematsu[b], Xiangrong Zhou[a], Takeshi Hara[a], Hiroki Kato[b], Hiroshi Kondo[b], Hiroshi Fujita[a] and Hiroaki Hoshi[b]

[a]Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Gifu 501-1194, Japan
[b]Department of Radiology, Gifu University School of Medicine, Gifu 501-1194, Japan

ABSTRACT

Cirrhosis of the liver is characterized by the presence of widespread nodules and fibrosis in the liver. The fibrosis and nodules formation causes distortion of the normal liver architecture, resulting in characteristic texture patterns. Texture patterns are commonly analyzed with the use of co-occurrence matrix based features measured on regions-of-interest (ROIs). A classifier is subsequently used for the classification of cirrhotic or non-cirrhotic livers. Problem arises if the classifier employed falls into the category of supervised classifier which is a popular choice. This is because the 'true disease states' of the ROIs are required for the training of the classifier but is, generally, not available. A common approach is to adopt the 'true disease state' of the liver as the 'true disease state' of all ROIs in that liver. This paper investigates the use of a nonsupervised classifier, the k-means clustering method in classifying livers as cirrhotic or non-cirrhotic using unlabelled ROI data. A preliminary result with a sensitivity and specificity of 72% and 60%, respectively, demonstrates the feasibility of using the k-means non-supervised clustering method in generating a characteristic cluster structure that could facilitate the classification of cirrhotic and non-cirrhotic livers.

Keywords: Computer-aided diagnosis (CAD), cirrhosis of the liver, magnetic resonance imaging (MRI), k-means clustering

1. INTRODUCTION

Cirrhosis of the liver is one of the leading causes of death by disease, killing more than 20,000 people in the United States each year. Cirrhosis of the liver is characterized by the presence of widespread fibrosis and regenerative nodules in the liver. The fibrosis and nodules formation causes distortion of the normal liver architecture, resulting in characteristic texture patterns. Magnetic resonance imaging (MRI) is an effect modality in assessing cirrhosis. Currently, MRI images are primarily interpreted by human observers. Computer-aided diagnosis (CAD) has drawn increasing attention over the last decade. It has the potential in improving radiologists' performance in MRI images interpretation and reducing inter- and intra-observer variations. However, there is an important issue in the training of a CAD system.

In interpreting hepatic MRI images for the diagnosis of cirrhosis or non-cirrhosis, radiologists search for manifestations of the disease in the liver region as well as in other abdominal areas. Manifestations in the liver such as formation of fibrosis and regenerative nodules, enlargement of the liver and an irregular or nodular liver boundary are primarily important while manifestations in other abdominal areas such as varices, splenomegaly and acites observed in the abdominal area are also very useful for the diagnosis of the disease. In addition, other information such as patient history and results of other tests also contribute to radiologists' diagnosis. In other words, radiologists come to a decision of cirrhosis or non-cirrhosis based on a range of global and local information and diagnosis was made with respect to a liver.

Computer aided diagnosis systems follow the suite and have their algorithms covering a range of clinically useful information. One of the key contributing factors is the formation of fibrosis and regenerative nodules and is commonly analyzed using co-occurrence matrix based texture features which are derived locally. That is, the co-occurrence matrix based texture features are measured on regions-of-interest (ROIs). A classifier is subsequently used for the classification of cirrhotic or non-cirrhotic livers. There are many popular choices for the classifier. A large number of these choices fall into the category of supervised classifier which includes Fisher's classifier, nearest neighborhood, artificial neural network, support vector machine and many others. Problem arises in the employment of classifiers in this category as the 'true disease states' of the ROIs are required for the training of the classifier but is, generally, not available. A common approach is to adopt the 'true disease state' of the liver as the 'true disease state' of all ROIs in that liver.

In this paper the use of a nonsupervised classifier, the k-means clustering method was investigated. The feasibility of using the k-means non-supervised clustering method in generating a characteristic cluster structure was evaluated and the performance of the generated cluster structure in classifying livers as cirrhotic or non-cirrhotic was evaluated. The extraction of texture features was reported previously [1] but is included in the following for the completeness of the paper.

2. METHODOLGY

2.1. Gradient magnitude image

The first challenge in computer analysis of MRI images is non-uniformity of image intensity observed in MRI images. Non-uniformity, or inhomogeneity, of image intensity refers to the spatial variation of image intensities which is not due to tissue differences. Specifically, the appearance of a slow varying shading effect over the image is typically seen. Figure 1 illustrates the image intensity non-uniformity apparent on a liver MRI image. Non-uniformity in the background poses a problem for conventional intensity based computer analysis of MRI images. Various methods have been proposed in the retrospective correction of image intensity non-uniformity in MRI images including parametric and nonparametric estimations of the non-uniform background [2-3]. Polynomial surface fitting, thin plate spline, homomorphic filtering, iterative algorithms and experimental calibration have also been studied [4-8]. It should be noted that when employing any background correction method, the introduction of biases and errors must be considered.
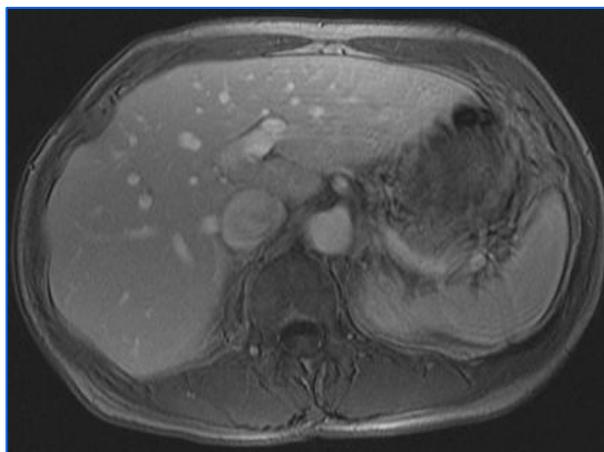
Fig. 1 Non-uniformity of image intensity in a liver MRI image. Note that the image intensity of tissues in the upper and lower sessions of the image is higher (brighter) than that of the same tissues in the middle session of the image.

Contrary to the common intensity-based approaches, a non-intensity based approach was used in this study for the analysis of the MRI images. In the analysis scheme, the gradient magnitude image $|G|$ was first generated. The gradient magnitude image $|G|$ was computed using the following formulation

$$|G| = \sqrt{G_x{}^{\wedge}2 + G_y{}^{\wedge}2} \; ,$$

where G_x and G_y are the partial derivatives of the raw MRI image with respect to x and y, respectively. The partial derivatives G_x and G_y were obtained using Sobel filtering. The filter kernels are shown in Figure 2. In the computed gradient magnitude image, the slow varying background was filtered out and the fibrosis structures were enhanced. All subsequent feature extractions were based on the gradient magnitude image.

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Fig. 2 The directional 3×3 Sobel filter kernels employed in computing the gradient magnitude image $|G|$.

2.2. Texture features

Figure 3 shows a typical histogram of the gradient magnitude of an ROI. The histogram depicts a unimodal distribution and is skewed to the right. The mode and the standard deviation of the distribution were estimated from the histogram profile of each and every ROI, giving rise to the first two texture features.
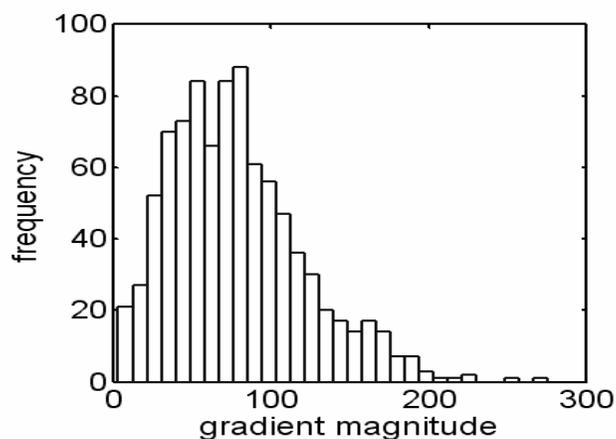
Fig.3 A typical histogram of the gradient magnitude of an ROI. The mode and the standard deviation estimated from the histogram profile of each and every ROI are two of the features measured for classification.

The second lot of features was measured on co-occurrence matrices. Co-occurrence matrices were computed on the gradient magnitude image of every ROI. In statistics, a co-occurrence matrix is defined as the joint probability of two events. When applying to a two-dimensional image, the co-occurrence matrix is the probability of two elements (in the two-dimensional array/image) at a distance d and a direction θ take on values i and j. Thus, the co-occurrence matrix element in row i and column j is denoted as

$$Pij = \text{Prob}(i, j \mid d, \theta).$$

Fourteen feature measures proposed by Haralick *et al*. [9] were measured on co-occurrence matrices. The 14 features were angular second moment, contrast, correlation, entropy, intertia, inverse difference moment, sum average, sum variance sum entropy, difference average, difference variance, difference entropy, information measure 1 and information measure 2. As the optimal values for the co-occurrence matrix parameters were not known *a priori*, co-occurrence matrices with distance $d = 1, 5, 9$ and angle $\theta = 0, \pi/4, \pi/2, 3\pi/4$ were constructed. Features derived from co-occurrence matrices based on the four different angles were later averaged. This is because no directional information were expected. In addition, different co-occurrence matrices were obtained by varying the gradient magnitude image depth (e.g. 256 gray levels or 1024 gray levels). The gradient magnitude image was requantized using an image depth of $Q = 16, 32$ or 128 gray levels. Consequently, a total of 126 features were obtained when all combinations of all three parameters d, θ and Q were considered (Table 1).

Table 1. Feature indices. Table entries are indices to the co-occurrence matrix feature with quantization *Q=16, 32 and 128* and distance *d=1, 5 and 9*. Fourteen features proposed by Haralick et al. [9] were repeatedly measured on the co-occurrence matrices, giving rise to features 1 to 126. Feature 127 and 128 correspond to mode and standard deviation of a gradient magnitude distribution. Haralick's 14 features are indexed in the following order: (1) angular second moment (2) contrast (3) correlation (4) entropy (5) inertia (6) inverse difference moment (7) sum average (8) sum variance (9) sum entropy (10) difference average (11) difference variance (12) difference entropy (13) information measures 1 (14) information measures 2.

	$d = 1$	$d = 5$	$d = 9$
Q=128	[1, ...14]	[15, ...28]	[29, ...42]
Q= 32	[43, .. .56]	[57, ...70]	[71, ... .84]
Q= 16	[85, ... 98]	[99, ...112]	[113, ..126]

2.3. Monte Carlo method

Monte Carlo methods are stochastic sampling methods often used to find the approximation of solutions to models that are too complex to be solved analytically, or too costly to be solved otherwise. A Monte Carlo algorithm evaluates a deterministic model iteratively with the use of randomly generated numbers. In this study, a Monte Carlo method was employed for feature selection. The Monte Carlo method used a random number generator to select 100 feature subset of a fixed size p. The performances of the randomly generated feature subsets in classifying cirrhotic and non-cirrhotic livers were evaluated and the best feature subset was found. Note that the optimal size (p) of the feature subset was not known *a prior* and must be determined experimentally. Due to the relatively small database and the issue of dimensionality, the range of p was limited to $p = 2, 3, ...6$.

2.4. K-means clustering

After the Monte Carlo feature selection, a k-means clustering technique [10] was used to partition the selected feature vectors (points in the p-dimensional feature space) into k mutually exclusive clusters with the aim to distinguish cirrhotic livers from non-cirrhotic livers. A k-means clustering method is an unsupervised partitioning method of which 'true labels' of the feature vectors are not required. The partitions are generated such that objects within the same cluster are close to each other, and objects in different clusters are far from each other. Different distance measures can be used. In this study, the Euclidean distance measure was employed.

In performing a k-means clustering algorithm, knowledge of the number of clusters is required but is, generally, not known *a prior*. Hence, the number of clusters must be determined experimentally. The decision of the number of clusters was based on two criteria. The first criterion was the separability of the resulting clusters. This is to evaluate how well-separated the resulting clusters are. A separability measure was defined on each and every

point (feature vector) in the feature space. The separability measure is an index indicating how close a point is to points in its cluster than to those in the neighboring clusters. In particular, the separability value S is a normalized difference measure of the average distance of the i-th feature vector to all other feature vector in the same cluster and the average distance of the i-th feature vector to all other feature vector in the closest cluster other than its own cluster, with distance being the Euclidean distance. The separability measure of the i-th point is written as

$$S(i) = \frac{\min_k(avgdist_{other}(i,k)) - avgdist_{within}(i)}{\max(avgdist_{within}(i), \min_k(avgdist_{other}(i,k)))}, \tag{1}$$

where $avgdist_{within}$ is the average distance from the i-th point to the other points in its cluster and $avgdist_{other}$ is the average distance from the i-th point to points in another cluster k. The separability value ranges from +1 to -1. A value of 1 indicates that the point is very distant from neighboring clusters while a value of -1 indicates that the point is closer to points in its nearest neighboring cluster than to those in its cluster.

Another criterion to determine the suitability of the choice of the number of cluster is to assess the resulting clusters in performing its task, in this case, discrimination of cirrhotic and non-cirrhotic livers. It should be mentioned here that there are more than one factor affecting the performance of the resulting clusters. For instance, if the selected features have no discriminative power, the resulting clusters will not be able to isolate cirrhotic livers from non-cirrhotic. However, all things being equal, a successful clustering should lead to the discovery of isolated groups of cirrhotic and non-cirrhotic livers.

Furthermore, it can be understood that not each and every ROIs obtained from cirrhotic livers display equally unequivocal signs and evidence of cirrhosis. The texture pattern in a diseased liver may vary from region to region. Hence, a successful discrimination of cirrhotic and cirrhotic livers may involve the clustering of the feature vectors into a number of clusters with multiple clusters of cirrhotic livers and non-cirrhotic livers. In addition, the multiple clusters of any disease state, i.e. cirrhotic or non-cirrhotic, are not necessarily adjacent to each other.

3. EXPERIMENTAL DATA

A database consists of 74 patient cases was employed in this study. Of the 74 patient cases, 25 were cirrhotic (11 grade 4 and 15 grade 3) and 49 were non-cirrhotic. The entire data set was randomly subdivided into two sets: the training set and the test set in a ratio of 6:4. Precisely, 15 cirrhotic cases and 30 non-cirrhotic cases were included in the training set and 10 cirrhosis cases and 19 non-cirrhosis cases were included in the test set. Each patient case comprises T1- and T2- weighted MRI images as well as gadolinium-enhanced images acquired in four different phases, namely, pre-contrast, hepatic arterial, portal venous and equilibrium phases. All MRI images were acquired in the axial plane. The size of the axial images was 512×512 pixels with a depth resolution of 16 bits (65536 gray scale levels). The slice interval was 10 mm. Depending on the size of the coverage, an MRI sequence typically includes a stack of 18-22 slice images. Gadolinium-enhanced equilibrium phase MRI images were employed in this study.

For each patient, a number of non-overlapping, regions-of-interest (ROIs) within the liver region were manually extracted from each of the equilibrium phase images. The size of the ROIs was 32×32 pixels. The number of ROIs varies from patient to patient depending on a number of factors such as the size of the liver and the presence of other space occupying lesions, for example, hemangioma and hepatocellular carcinoma, but was generally in the range of 20 to 70.

4. RESULTS AND DISCUSSIONS

4.1. Number of clusters in k-means clustering

In determining the suitable number of clusters, a range of values were investigated and the results of partitioning the feature vectors into k clusters where $k = 2, 3, ...10$ were analyzed. Criteria for the decision of the number of clusters were described in Section 2.4.1. Figure 4a shows a separability plot of the results of k-means clustering with k equals to two. The y-axis of the horizontal bar graph is the ROI (feature vector) index by cluster, whereas the x-axis is the separability value S. It is noted that both cluster 1 and cluster 2 have portion of their members attaining low separability values (< 0.5). Moreover, some of the members in cluster 2 have negative separability values, indicating that those points are closer to points in cluster 1 than to those in its cluster. The separabilility plot with feature vectors clustered into three clusters is shown in Figure 4b. The result was not much different from the one using two clusters. When the number of clusters was increased to four (Figure 4c), a small cluster appeared. The appearance of a small cluster or small clusters persists with the increase in k. It was mentioned earlier in this manuscript that cirrhosis is the result and end stage of chronic liver diseases. Hence, feature vectors that are due to cirrhosis are thought to be extreme from those that are fibrosis or normal and the number of feature vectors that are evidence of cirrhosis would be small. In addition, due to the limited size of the data set, the value of k was capped at 10 Based on these assumptions and limitations, the smallest two clusters in a 10-cluster structure were taken as a candidate for cirrhotic cluster set. Figure 4d shows the k-cluster structure with $k = 10$.
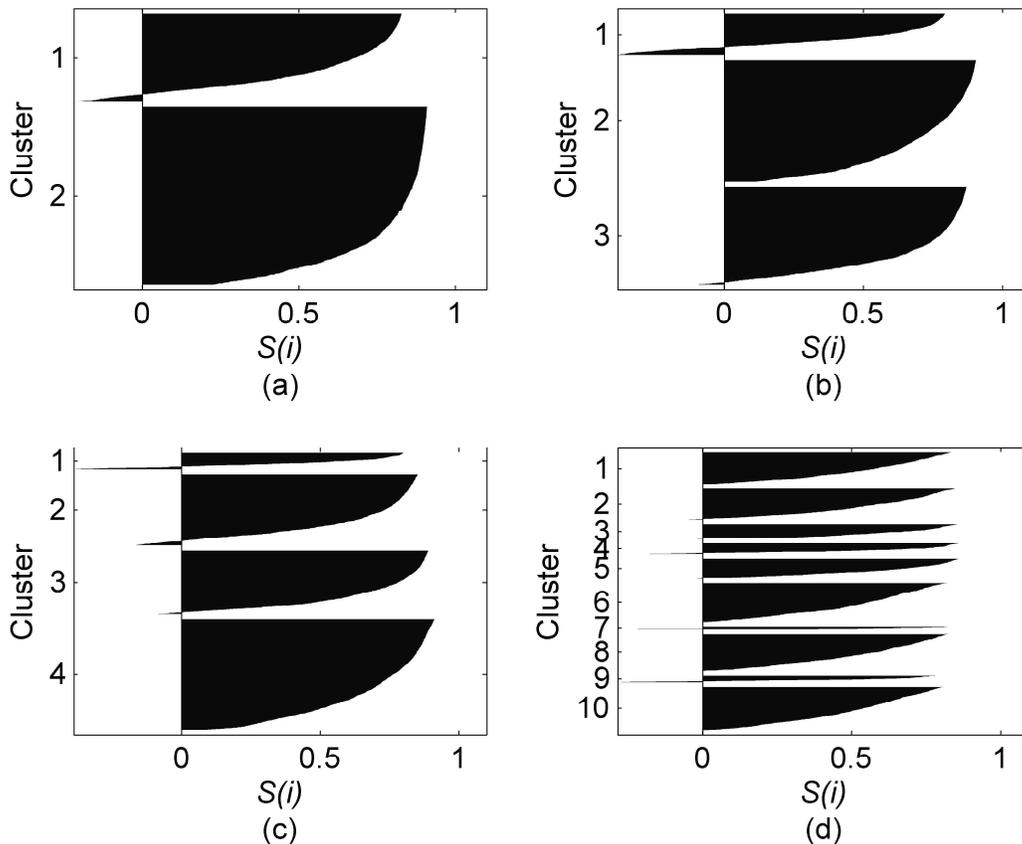
Fig. 4 Separability plot of a k-cluster structure generated by the k-means clustering algorithm with (a) $k = 2$, (b) $k = 3$, (c) $k = 4$, and (d) $k = 10$.

With respect to the above candidate cirrhotic clusters, the ability of the candidate cluster set in facilitating the classification of cirrhotic and non-cirrhotic livers was evaluated. In other words, the performance of the resulting clusters in differentiating cirrhotic and non-cirrhotic livers was evaluated. Results are shown in the next section. In addition, classification using the two-cluster structure shown in Figure 4a was also analyzed. In that investigation, it was found that each liver, whether cirrhotic or non-cirrhotic, has some ROIs included in cluster 1 and some in cluster 2. This means that no discrimination was achieved when partitioning the feature vectors (ROIs) into two groups. This is not surprising as it is unrealistic to expect each and every ROIs obtained from cirrhotic livers display equally unequivocal signs and evidence of cirrhosis. This issue of sample-dependence in the diagnosis of cirrhosis is also a challenge in liver biopsy.

4.2. Classification of cirrhotic and non-cirrhotic livers

Classification of cirrhotic and non-cirrhotic livers were based on the feature subsets selected by the Monte Carlo method for each subset size p where $p = 2, 3, \ldots 6$. K-means clustering with $k = 10$ was used to find natural clusters of the feature vectors and the smallest two clusters were taken as the diseased clusters. In addition, only ROIs (feature vectors) attained a separability value of $S >= 0.5$ were assigned as test-positive. Based on the test-positive ROIs, the true-positive fraction (TPF) and the false-positive fraction (FPF) were computed and were given as

$$\text{TPF} = \frac{\text{number of cases with at least one test-positive ROI and are actually positive}}{\text{total number of true positive (cirrhotic) cases}},$$

and

$$\text{FPF} = \frac{\text{number of cases with at least one test-positive ROI and are actually negative}}{\text{total number of true negative (non-cirrhotic) cases}}.$$

Table 2 shows the results of classification. Only results with TPF and (1-FPF) both attain values close to or greater than 60%, or one of the quantities attains a value close to or greater than 50% and the other quantity attains a value close to or greater than 70% are included. Moreover, results with six feature subsets are not included in the table because all feature subsets tested attained their sensitivities and specificities both below 60%. From the table, the best result was found with a sensitivity and specificity of 72% and 60%, respectively, using two features. Both of the two features were co-occurrence matrix based features. They are feature 3 and feature 112, that is, correlation measured on co-occurrence matrix with $d = 1$ and $Q = 128$ and information measure 2 measured on co-occurrence matrix with $d = 5$ and $Q = 16$, respectively.

Table 2. Classification results using k-means clustering

Number of features (p)	Sensitivity (TPF)	Specificity (1-FPF)	
2	72%	60%	Feature subset =[3, 112]
3	72%	47%	
	64%	60%	
4	72%	51%	
	72%	47%	
	60%	61%	
5	72%	51%	
	56%	63%	

5. CONCLUSIONS

A k-means unsupervised clustering technique was investigated for the differentiation of cirrhotic and non-cirrhotic livers where texture features were measured on ROIs of the livers but the 'true disease states' of the ROIs were not known. Using two clusters selected from a ten-cluster structure, classification results with a sensitivity and specificity of 72% and 60%, respectively, were achieved, indicating the feasibility of k-means clustering in generating a clustering configuration that could facilitate the differentiation of cirrhotic and non-cirrhotic livers. Furthermore, in determining the appropriate number of clusters, it was found that a commonly used two-cluster configuration with one cluster labeled cirrhosis and the other cluster labeled non-cirrhosis was not a suitable candidate as ROIs of each and every liver were found in both the cirrhosis and non-cirrhosis clusters. This is not surprising as the degree of fibrosis varies locally within a liver.

ACKNOWLEDGMENTS

REFERENCES

1. G. N. Lee, X. Zhang, M. Kanematsu, *et al*., "Classification of cirrhotic liver on MR images using texture analysis," *Int'l J Computer Assisted Radiology and Surgery*, 1, Supplement 1, 379-381 (2006).
2. M. Styner, C. Brechbuhler, G. Szekely, *et al.,* "Parametric estimate of intensity inhomogeneities applied to MRI," *IEEE Trans. Med. Imag*., 19(3):153-165 (2000).
3. J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. on Med. Imag*., 17:87-89 (1998)
4. M. Tincher, C. R. Meyer, R. Gupta, *et al*., "Polynomial modeling and reduction of RF body coil spatial inhomogeneity in MRI*," IEEE Trans. Med. Imag*., 12(2):361-365 (1993).
5. C. R. Meyer, P. H. Bland, and J. Pipe, "Retrospective correction of intensity inhomogeneities in MRI," *IEEE Trans. Med. Imag.,* 14:36-41 (1995).
6. B. Dawant, A. Zijidenbos, and R. Margolin, "Correction of intensity variations in MR images for computer-aided tissue classification," IEEE Trans. Med. Imag., 12:770-781 (1993).
7. B. H. Brinkmann, A. Manduca, and R. A. Robb, "Optimized homo-morphic unsharp masking for MR grayscale inhomogeneity correction," *IEEE Trans. Med. Imag*., 17:161-171 (1998).
8. Y. M. Kadah and X. Hu, "Algebraic reconstruction for magnetic resonance imaging under B_o inhomogeneity," *IEEE Trans. Med. Imag*., 17 (3):362-370 (1998).
9. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans Syst Man Cybern*., 3(6):610-621 (1973).
10. A. K. Jain, M. N. Murty, and P.J. Flynn. "Data Clustering: A Review," *ACM Computing Surveys*, 31(3):264-323 (1999).