Classification of mammographic lesions into BI-RADS™ shape categories using the Beamlet Transform

Mehul P. Sampat[a], Alan C. Bovik[b] and Mia K. Markey[♦a]
[a]Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA
[b]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ABSTRACT
We present a new algorithm and preliminary results for classifying lesions into BI-RADS shape categories: round, oval, lobulated, or irregular. By classifying masses into one of these categories, computer aided detection (CAD) systems will be able to provide additional information to radiologists. Thus, such a tool could potentially be used in conjunction with a CAD system to enable greater interaction and personalization. For this classification task, we have developed a new set of features using the Beamlet transform, which is a recently developed multi-scale image analysis transform. We trained a k-Nearest Neighbor classifier using images from the Digital Database for Digital Mammography (DDSM). The method was tested on a set of 25 images of each type and we obtained a classification accuracy of 78% for classifying masses as oval or round and an accuracy of 72% for classifying masses as lobulated or round.

Keywords: BI-RADS™ categories, Computer aided diagnosis, Beamlet Transform, k-Nearest Neighbor.

1. INTRODUCTION

1.1. Motivation

The American Cancer Society estimates that 215,990 women will be diagnosed with breast cancer and 40,110 women will die of the disease in the U.S. in 2004 [1]. Early detection of breast cancer increases the treatment options and the survival rate.

Currently, the most effective and commonly used tool for early detection of breast cancer is screening mammography. In screening mammography, two views of each breast are recorded: the craniocaudal (CC) view, which is a top to bottom view, and a mediolateral oblique (MLO) view, which is a side view taken at an angle. Radiologists visually analyze mammograms for signs of breast cancer. Some of the common signs of cancer are calcifications, masses, and architectural distortions.

A number of descriptors are used to characterize these abnormalities. They are described and reported according to the Breast Imaging Reporting and Data System (BI-RADS™) [2]. BI-RADS™ is a mammography lexicon developed by the American College of Radiology (ACR), for the description of mammographic lesions. The BI-RADS™ lexicon includes descriptors such as the margin of a mass and the distribution of calcifications and it defines final assessment categories to describe the radiologist's level of suspicion about the mammographic abnormality. If a suspicious abnormality is detected, a diagnostic mammographic examination is carried out to decide the future course of action required. Based on the level of suspicion of the abnormality following the diagnostic examination, a recommendation is made for routine follow up, short-term follow up, or biopsy. The descriptors used to define masses are shape and margin. Masses can be round, oval, lobulated, or irregular in shape. The mass boundary or margins may be circumscribed, microlobulated, obscured, ill-defined, or spiculated. Some mass features are more worrisome than others are. For example, irregular, spiculated masses are more likely to be malignant than round, circumscribed masses.

Computer-Aided Detection (CAD) systems have been developed to aid radiologists in detecting mammographic lesions that may indicate the presence of breast cancer [3-6]. These systems act only as a second reader and the final decision is made by the radiologist. Computer-Aided Diagnosis (CADx) systems for aiding in the decision between follow-up and biopsy are not available commercially, although

♦ mia.markey@mail.utexas.edu; phone: +1.512.471.1771; fax: +1.512.471.0616; http://www.bme.utexas.edu/research/informatics/

numerous researchers have developed methods to classify masses as malignant or benign. However, to the best of our knowledge, very little work has been done to classify masses into one of the four BI-RADS™ shape categories. Classifying masses into the different shape categories has a number of advantages. A computer-aided detection (CAD) system could then not only detect the spatial location of the mass, but would also give additional information regarding its shape to the radiologist. This method could potentially be used to personalize a CAD system for each user; for example, Dr. X wants to be prompted on irregular or spiculated masses only.

Several studies have shown that CAD improves the rate of breast cancer detection. However, a very large and recent study by Gur *et al.* [7] showed that CAD did not increase the detection rate of breast cancer. This may be due to the fact that existing systems typically employ a "one-size-fits-all" approach that provides the same information in the same manner to all users. It is possible that systems that interact in a flexible manner with radiologists and are "personalized" will be more useful and useable. For example, Korsch *et al.* [8] assessed the benefits of a CADx system, which could adjust its output, based on the user's expectation of disease prevalence and showed that this increased the correlation between the radiologist estimated and computer-estimated probabilities of malignancy.

Another potential application of this method would be to build upon research on BI-RADS™ based CADx approaches. In these approaches, the BI-RADS™ descriptors of abnormalities (assigned by radiologists) are used to classify abnormalities as malignant or benign and have shown to be quite accurate [9-18].

In this paper, we propose a new algorithm for classifying masses as round or oval and round or lobulated. For this task, we have developed a new set of features. To obtain these features we first compute the Beamlet transform of the image. The Beamlet transform is a recently developed multi-scale image transform [19]. The organization of the paper is as follows: Section 1.2 briefly describes the Beamlet transform and the motivations for using it. Section 2 describes the new features we have developed for this task. This is followed by a description of the data sets and experiment setup. Section 3 describes the preliminary results achieved and the conclusion and discussion are presented in Section 4.

1.2. Beamlet Transform

Our aim is to classify masses into the four BI-RADS™ shape categories. A number of general purpose shape descriptors have been developed by various researchers [20]. The computation of these features usually requires the pre-processing step of segmentation to be performed on the object. However, segmentation is a difficult task and shape features are sensitive to segmentation errors. Masses generally have fuzzy boundaries and this increases the likelihood of segmentation errors. Thus, we did not want to use the general purpose shape descriptors but have tried to develop features specifically designed for masses, which would be robust to segmentation errors.

To compute features, we decided to new use a newly developed multi-scale image analysis transform called the Beamlet transform. There are a number of reasons why this particular transform is ideal for the task of computing discriminative features for masses. It is well known that masses occur across a range of scales. The diameter of masses varies from 4 mm to 4 cm [21]. Masses may also occur in any orientation and thus a transform like the Beamlet transform, which can capture both the multi-scale and multi-orientation aspects of masses, is appropriate. We now describe some details of the Beamlet transform.

In this section, we give a brief introduction to the Beamlet transform. A Beamlet can be viewed as line segment through an image. A Beamlet dictionary is a collection of line segments, which occupy a range of locations, scales and orientations. These can be used to generate multi-scale approximations to the collection of all line segments.

The Beamlet transform of an image is a collection of all line integrals over line segments, which are present in the Beamlet dictionary. Readers familiar with the Radon transform may note that the Beamlet transform can be viewed as a multi-scale Radon transform. A more detailed description of the Beamlet transform can be obtained in [19].

Using the Beamlet transform as a basis a number of image analysis tools have been developed [19]. One such method is known as the Beamlet Decorated Recursive Dyadic Partition (BD-RDP), which is described below. A recursive dyadic partition (RDP) of an image is any partition obtained by dividing a square image into four equal-sized squares with their sides half in length of the original square and then

recursively dividing each new square that is generated. If each new square is divided into four parts, the partition is called a complete RDP; otherwise, if squares are selectively divided, based on certain criteria, the partition formed is called an incomplete RDP. A complete and incomplete RDP are shown in Figure 1. Note that the complete RDP is a 3-level RDP as squares have been recursively divided three times. This idea is similar to that of quad-tree decomposition.

A BD-RDP is an RDP in which some of the terminal squares of the RDP contain a Beamlet, which is associated with that square. The Beamlets used can only be chosen from the Beamlet dictionary. Since the Beamlet dictionary is a collection of line segments at various locations, orientations and scales, the BD-RDP can be viewed as a mechanism to represent the boundary of an object using these line segments. Two shapes and their BD-RDP representations are shown in Figure 2. Note that for one object, a 1-level incomplete RDP was required whereas for the second one, a 2-level incomplete RDP was necessary.

Given an image containing an object, a number of BD-RDPs can be generated and the computation of the optimal BD-RDP to represent the object has been modeled as an optimization problem [19]. Two Regions of Interest (ROI) containing a round and oval mass and their corresponding BD-RDPs are shown in Figure1. The advantage of such a representation is that any object of an arbitrary shape can be described with just a few Beamlets.

2. METHODS

2.1. Data Description

The images for this study were obtained from the Digital Database for Screening Mammography (DDSM) (http://marathon.csee.usf.edu/Mammography/Database.html) [22]. The DDSM is the largest publicly available data-set of digitized mammograms. To evaluate the performance of the algorithm, a dataset consisting of 200 regions of interests (ROI) from digitized mammograms depicting breast masses in the Digital Database for Screening Mammography (DDSM) were used. There were 50 cases in each of the four BI-RADS™ shape categories: round, oval, lobulated, or irregular. To compute a BD-RDP, an image the size of the image must be a power of two. For this, all ROIs were resized to a size of 128-by-128 pixels. The aspect ratio of the original image was maintained.

All images were scanned with a single scanner and they contained a single lesion and were randomly selected. More details on the set of images used can be obtained on our website (http://www.bme.utexas.edu/research/informatics/index.aspx). Tables 2, 3 and 4 show detailed descriptions of various parameters of the masses used for training and testing. The parameters reported, are the density, subtlety and pathology of the masses. The density is a BI-RADS™ descriptor and can take values from 1 to 4. A density value of 1 indicates fatty breast tissue while a value of 4 indicates dense breast tissue. The subtlety parameter is not a BI-RADS™ descriptor and can take values in the range 1 to 5, where 1 indicates a “subtle lesion” and 5 indicates an “obvious lesion”. This parameter is a subjective measure of the subtlety of a lesion, to an expert radiologist and this may indicate the difficulty in finding the lesion. (http://marathon.csee.usf.edu/Mammography/DDSM/ddsm_terminology.html)

The pathology of masses was obtained from a biopsy examination and was either malignant or benign. For a small number of oval (3) and lobulated (2) masses, the pathology results were not available and were labeled as unproven. For each experiment, 100 images were used, and 50 images were used to empirically optimize the parameters of the algorithm. Once the parameters were set, the algorithm was tested on the remaining 50 images.

2.2. Feature Extraction

In this section, we describe the features we developed for the task. The features we propose are derived from the BD-RDP of each object. That is, the BD-RDP of each image is computed and the features are computed from the properties of the Beamlets of the representation. The Beamlet transform and the BD-RDP were computed using the BeamLab toolbox [23]. This is a MATLAB® (The MathWorks, Natick, MA) based toolbox and can be downloaded from: http://www-stat.stanford.edu/~beamlab/ .

Intuitively, shapes of different categories would produce unique BD-RDP representations and the differences in these representations could be used to differentiate shapes of various categories. One way in which the BD-RDP could be used to discriminate between shapes is shown in Figure 2.

For round masses, the distance of the Beamlets from the center of the mass should be uniform, whereas, for oval or lobulated masses one would expect more variability in the distances of the Beamlets from the center of the mass. Thus, a histogram of the distance of Beamlets from the centroid of the mass was computed and the standard deviation (Dist_std), variance (Dist_var), skewness (Dist_skew) and kurtosis (Dist_kurt) of the histogram were calculated.

The other features are based on the observation that the normalized histogram of orientations of the Beamlets used to describe round masses is different from that for oval and lobulated masses. Thus, we computed a normalized histogram of orientation of beamlets of each image. The standard deviation (Ori_std) of this histogram and six histogram bin values (Ori_1 to Ori_6) were used as features. Thus, a total of 10 features are obtained for each ROI. A summary of the features is given in Table 1.

Feature Names	**Description**
Dist_std, Dist_var, Dist_skew and Dist_kurt	Moments of the histogram of distance of Beamlets from the center of the mass.
Ori_std and Ori_1 to Ori_6	Standard deviation of the histogram of Beamlet orientations; Ori_1 to Ori_6 are the histogram bin values of the orientation histogram.

Table 1: A summary of the different features that were developed for classification into BI-RADS™ shape categories. These were based on the histogram of the distances of Beamlets from the centroids and the histograms of Beamlet orientations.

2.2. Classification and Experimental Setup

The features were used in a k nearest neighbor (k-NN) classifier to predict the shape value. The k-NN classifier is a supervised non-parametric classification method [24]. Briefly, given a set of training samples and a test sample *'t'*, the k-NN method computes the k-nearest training samples to *'t'* in terms of a distance metric and then assigns the sample *'t'* to the class that occurs most frequently among the k-nearest training samples [24]. The distance metric we used was the Euclidean distance.

We initially attempted to classify masses into the four BI-RADS™ shape categories. The data set of 200 images was divided into training and testing sets of 100 images each. The training data was used to empirically optimize the parameters of the algorithm such as the number of neighbors 'k'. Using a leave-one-classification method on the training data, the optimal choice for k was 9. Once the parameters of the algorithm were set using the method described above, it was tested on the remaining 100 test samples. The testing was done using a leave-one-out classification k-NN (k=9). However, we did not achieve very good results on this classification task and observed that the classifier performed significantly better on classifying round and oval masses that it did on classifying lobulated and irregular masses. Thus, we tried to classify masses into two shape categories at a time.

Two experiments were conducted. In the first one, the aim was to classify masses as round or oval. In the second one, the aim was to classify masses as round or lobulated. Thus, there were a total of 100 images for each experiment. Half of these were used for training purposes and the other half was used for testing. The training data was used in the same manner as described above and once the parameters of the algorithm were set, it was tested on the remaining 50 test samples. The testing was done using a leave-one-out classification k-NN (k=9).

3. RESULTS

In this section, we present the preliminary results of our experiments. In one experiment, we attempted to classify masses as oval vs. round and in the other, lobulated vs. round. In Figure 3 we show two examples of a BD-RDP for a round and oval mass respectively. We note that in the BD-RDP for the round mass, most of the lines segments (Beamlets) occur at approximately the same distance from the centroid of the round mass (ignoring the segments in the periphery); where as for the oval mass the distance of the Beamlets from the centroid varies considerably. This difference is captured by the features (Dist_std and Dist_var), which were described in Section 2.2.

As mentioned in section 2.2, we had initially attempted to classify masses into the four BI-RADS™ shape categories but we were not very successful in this task. Table 5 shows the four-by-four confusion matrix for this experiment. The classification accuracy achieved was 46%. The classifier was most accurate in classifying round masses and least accurate in classifying irregular masses.

Thus, we choose to classify masses into two shape categories at a time and conducted two experiments.In the first experiment, the classification accuracy obtained was 78% and the accuracy obtained in the second experiment was 72%. The confusion matrix for an 'N' class classification task is an N-by-N matrix in which the columns represent the actual category of the objects and the rows represent the category assigned by the classification method. If a diagonal confusion matrix (only diagonal entries are non-zero) is achieved, that shows that the classifier is perfect since it classified all objects into the correct categories. In our case, we attempted to classify masses into two classes and the corresponding 2-by-2 confusion matrices for both experiments are shown in Table 6.

From these confusion matrices, we can analyze the errors that occurred. For example, we see that the algorithm incorrectly classifies many lobulated masses as round masses. A lobulated mass is defined by BI-RADS™ as "A mass that has contours with undulations." [2]. One possible explanation for this error is that the features failed to capture this subtle characteristic of lobulated masses and incorrectly classified them as round masses. We also observe that the algorithm performs much better on classifying masses as round or oval. This can be explained by the fact that the differences in round versus oval masses are more obvious that the differences between round versus lobulated masses. Thus, we need to design new features to identify lobulated masses.

The above method of evaluation of classifier performance uses the classification assigned by a radiologist as the ground truth. However, it is known that inter-observer variability is present between multiple observers and a number of studies have been carried out to measure this variability, in which, the kappa statistic has been used to measure this variability. Baker *et al*. [25] have measured the inter-observer variability between two radiologists while classifying masses into the BI-RADS™ categories and reported that the kappa value was 0.65. Kerlikowske *et al.* [26] reported a kappa value of 0.40 and Berg *et al.* [27] reported an inter-observer kappa value of 0.28 for mass shape. Thus, we see that there is a significant range in values of kappa reporting the agreement between radiologists.

We also used the same metric to measure the variability between our method and the radiologist (ground truth). We note that the while the studies reported above, computed the kappa statistic for classification into all four shape categories, while we have focused on classification into any two shape categories at a time only. The kappa value for classification as round versus oval was 0.52 and the kappa value for classification as round versus lobulated was 0.44. For the 4-class classification experiment, the kappa value was 0.28.

Kappa values in the range 0.41 to 0.6 suggest that the agreement is moderate, whereas, Kappa values in the range 0.21 to 0.4 suggest that the agreement is fair [28]. We are working on developing new features for classifying masses into all four BI-RADS™ shape categories and on improving the accuracy of the algorithm.

4. CONCLUSION AND DISCUSSION

In this work, we have presented preliminary results for the classification of masses into the BI-RADS™ shape categories was presented. The motivation was to design new features, which would be robust to segmentation errors. To capture the multi-scale and multi-orientation aspects of masses, we used the Beamlet transform. The masses were then classified into 2 categories using a k-NN classifier (k = 9) We obtained classification accuracies of 78% for classifying masses as oval or round and 72% for classifying masses as lobulated or round. For future work, we would like to develop features, which provide good discrimination among all four shapes simultaneously. We will now discuss some limitations of the proposed algorithm and some ideas on how these issues can be resolved.

A limitation of our current algorithm is that after the computation of a BD-RDP, there were a few Beamlets in the periphery, which did not correspond to the mass region. These could cause errors in the classification task. We see that this occurs because the BD-RDP is sensitive to linear tissue structures and tries to account for them. For example, this is observed in the BD-RDP for the round mass shown in Figure 3. In this image, one sees "stray" Beamlets in the upper and lower corners on the left side of the image. These line segments could adversely affect the performance of the algorithm and could be easily discarded before computing the features. These could be removed by using a convolving each ROI with a smoothing filter like a two-dimensional Gaussian filter.

Another reason for the errors could be variability between the masses in the training and testing sets. For example, the density of the breast tissue can vary from being heterogeneously dense to being

completely fatty. Note that the parameters of the algorithm were empirically optimized over the training data. If most samples in the training data and testing data were of different categories, the parameters would not be ideal for the testing data. We plan to design new features, which are not affected (that is, are invariant) to properties like mass density and subtlety. Currently, we have not accounted for major variations in mass density and subtlety and will do so in the future. The Beamlet transform was computed at many scales and because of this choice very small line segments in the BD-RDP representation of each mass are observed. These small segments could be combined into a larger smoother segment, which may form a better representation of the mass.

The classifier (k-NN), used is a supervised parametric classification technique in which a test case, is classified by examining the classes of the 'k' nearest training samples. Another manner in which the output of the classifier could be used is that a likelihood could be assigned for each class. For example, if for a test case, 4 out of 5 nearest training samples belonged to the category round, then one could say that the likelihood or probability that the test case is a round mass is 0.8. Other classifiers such as, decision trees can also be applied for this task

Pathology/Density	1	2	3	4	Total
Test_Benign	3	4	4	2	13
Test_Malignant	2	7	2	1	12
Train_ Benign		10	4	5	19
Train_ Malignant	1	5			6
Total	6	26	10	8	50

Pathology/Subtlety	2	3	4	5	Total
Test_ Benign		2	3	8	13
Test_ Malignant				12	12
Train_ Benign	1	11	6	1	19
Train_ Malignant			3	3	6
Total	1	13	12	24	50

Table 2: These tables show the properties of the round masses used for training and testing. The first tables shows the pathology and density parameters of the masses and the second table shows the pathology and subtlety values.

Pathology/Density	1	2	3	4	Total
Test_ Benign		6	4		10
Test_ Malignant	6	6	2		14
Train_ Benign	2	16	3		21
Train_ Malignant		1	2		3
Train_Unproven				2	2
Total	8	29	11	2	50

Pathology/Subtlety	2	3	4	5	Total
Test_ Benign	1	1	3	5	10
Test_ Malignant				14	14
Train_ Benign	1	6	7	7	21
Train_ Malignant		2		1	3
Train_Unproven		2			2
Total	2	11	10	27	50

Table 3: These tables show the properties of the lobulated masses used for training and testing. The first tables shows the pathology and density parameters of the masses and the second table shows the pathology and subtlety values.

Pathology/Density	1	2	3	4	Total
Test_ Benign	7	7	2		16
Test_ Malignant	3	5	1		9
Train_ Benign	1	3	4	1	9
Train_ Malignant	11	2			13
Train_Unproven		1	2		3
Total	22	18	9	1	50

Pathology/Subtlety	3	4	5	Total
Test_ Benign	1	4	11	16
Test_ Malignant		1	8	9
Train_ Benign	4	2	3	9
Train_ Malignant		1	12	13
Train_Uproven	2	1		3
Total	7	9	34	50

Table 4: These tables show the properties of the oval masses used for training and testing. The first tables shows the pathology and density parameters of the masses and the second table shows the pathology and subtlety values.

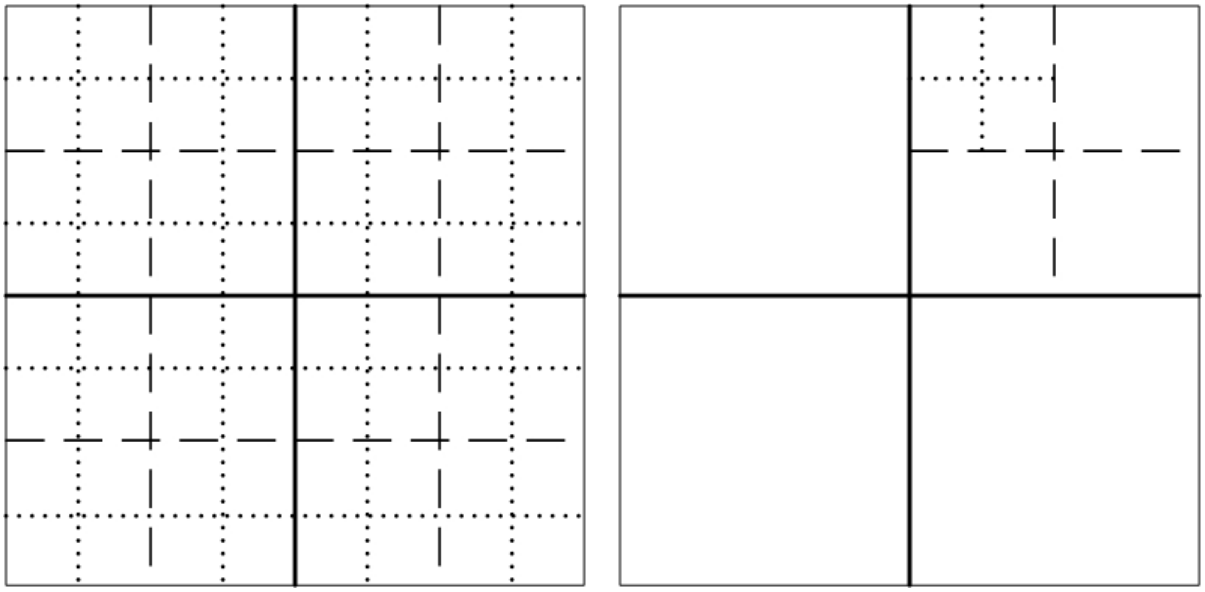

Figure 1:
A complete Recursive Dyadic Partition (RDP) (left) and An incomplete RDP (right). A recursive dyadic partition (RDP) of an image is any partition obtained by dividing a square image into four equal-size squares, which are half the size of the original square, and then recursively dividing each new square that is generated. If each new square is divided into four parts, the partition is called a complete RDP; otherwise, if squares are selectively divided, based on certain criteria, the partition formed is called an incomplete RDP. The complete RDP is a 3-level RDP since squares has been divided recursively at three levels. Lines of different formatting denote the three levels. (Adapted from [19])

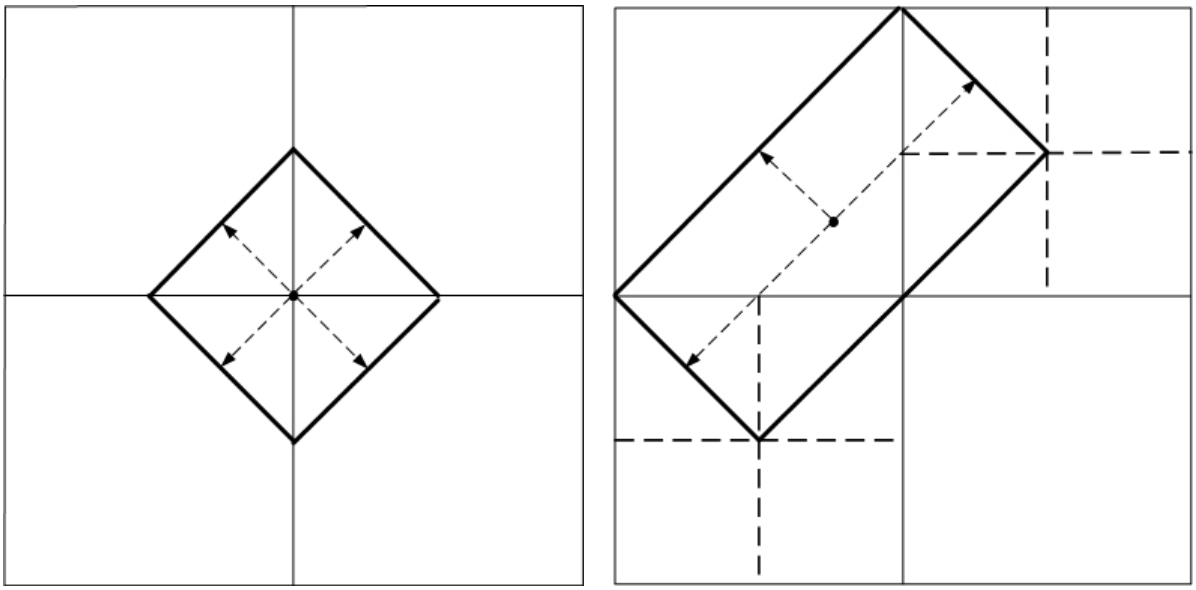

Figure 2: A schematic explanation of the features:
Any shape can be approximated with a BD-RDP. Two objects and the corresponding BD-RDP are in this figure. Note that on the left, one level of decomposition was required to represent the object, where on the right two levels of decomposition were required. The dotted lines with arrows show the distance of the beamlets from the centroid of the objects. (The centroid is represented by the 'black circle'). We note that in the first shape the distance of the all of the beamlets from the centroid is equal. Whereas for the second shape, the distance of beamlets from the centroid varies. This difference can be used to differentiate between the two shapes.

True Classification **Predicted Classification**	*Round*	*Oval*	*Lobulated*	*Irregular*
Round	16	5	2	7
Oval	4	15	10	7
Lobulated	1	4	8	4
Irregular	4	1	5	7

Table 5: This table shows the confusion matrices for classifying masses into the four BI-RADS™ shape categories. The columns represent the true class of the mass and the rows represent the predicted class of the mass. The classification accuracy was 46% and the kappa value was 0.28

True Classification **Predicted Classification**	*Round*	*Oval*
Round	20	6
Oval	5	19

True Classification **Predicted Classification**	*Round*	*Lobulated*
Round	19	8
Lobulated	6	17

Table 6: The tables show the confusion matrices for both experiments. In the first experiment, masses were classified as Round or Oval and in the second experiment, they were classified as Round or Lobulated. The columns represent the true class of the mass and the rows represent the predicted class of the mass. The classification accuracy for the first experiment was 78% and it was 72% for the second experiment.

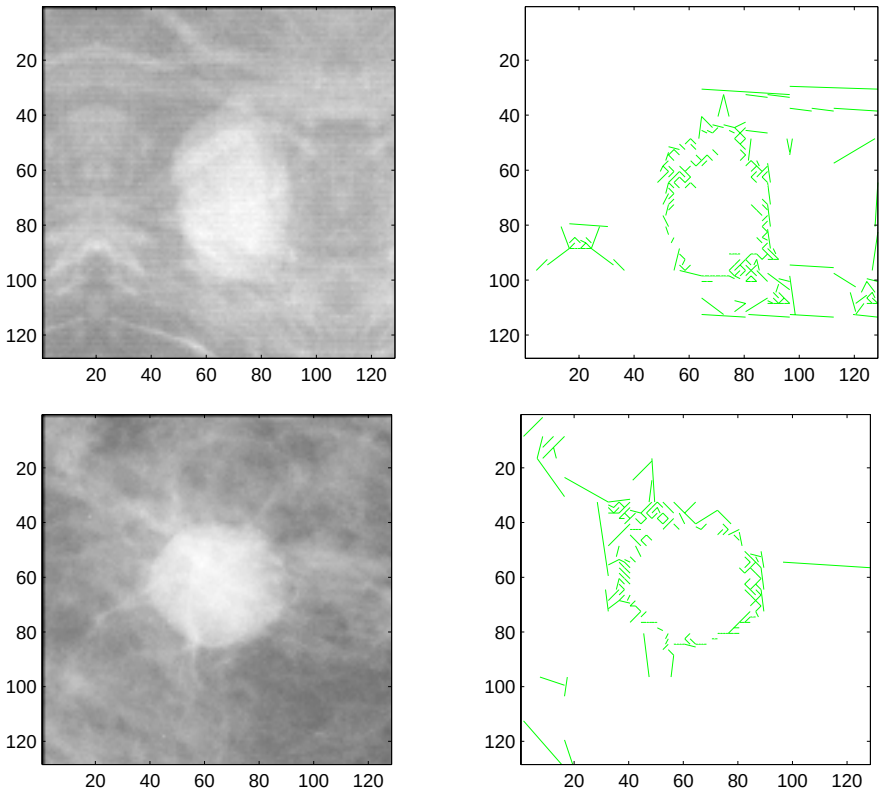

Figure 3: An oval (above) and round (below) mass and their corresponding Beamlet-Decorated Recursive Dyadic Partition (BD-RDPs). Features were extracted from the BD-RDP to classify the mass into a BI-RADS™ shape category. A number of features are extracted from the BD-RDP. For example, some of the features are computed from the histograms of the distance of the Beamlets to the center.

REFERENCES

[1] "Cancer Facts and Figures 2004," American Cancer Society, Atlanta 2004.

[2] American College of Radiology, *ACR BI-RADS - Mammography, Ultrasound & Magnetic Resonance Imaging*, Fourth ed. Reston, VA: American College of Radiology, 2003.

[3] M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed., 2nd ed, 2005.

[4] M. L. Giger, "Computer-aided diagnosis of breast lesions in medical images," *Computing in Science & Engineering*, vol. 2, pp. 39-45, 2000.

[5] C. J. Vyborny, M. L. Giger, and R. M. Nishikawa, "Computer-aided detection and diagnosis of breast cancer," *Radiologic Clinics of North America*, vol. 38, pp. 725-40, 2000.

[6] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, and Y. Jiang, "Computer-aided diagnosis in radiology: potential and pitfalls," *European Journal of Radiology.*, vol. 31, pp. 97-109, 1999.

[7] D. Gur, J. H. Sumkin, H. E. Rockette, M. Ganott, C. Hakim, L. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system.[see comment]," *Journal of the National Cancer Institute*, vol. 96, pp. 185-90, 2004.

[8] K. Horsch, M. L. Giger, C. E. Metz, C. J. Vyborny, G. Newstead, and R. S. Schmidt, "Prevalence-modified estimation of computer-determined probabilities of malignancy for CAD," presented at Radiological Society of North America Annual Meeting, Chicago, IL., 2003.

[9] M. K. Markey, J. Y. Lo, and C. E. Floyd, Jr., "Differences between computer-aided diagnosis of breast masses and that of calcifications," *Radiology*, vol. 223, pp. 489-93, 2002.
[10] M. K. Markey, J. Y. Lo, G. D. Tourassi, and C. E. Floyd, Jr., "Cluster analysis of BI-RADS descriptions of biopsy-proven breast lesions," presented at Medical Imaging 2002: Image Processing, San Diego, 2002.
[11] J. A. Baker, P. J. Kornguth, J. Y. Lo, and C. E. Floyd, Jr., "Artificial neural network: improving the quality of breast biopsy recommendations," *Radiology*, vol. 198, pp. 131-5, 1996.
[12] J. Y. Lo, J. A. Baker, P. J. Kornguth, J. D. Iglehart, and C. E. Floyd, Jr., "Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features," *Radiology*, vol. 203, pp. 159-63, 1997.
[13] C. E. Floyd, Jr., J. Y. Lo, and G. D. Tourassi, "Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions," *American Journal of Roentgenology*, vol. 175, pp. 1347-52, 2000.
[14] J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford, and C. E. Floyd, Jr., "Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon," *Radiology*, vol. 196, pp. 817-22, 1995.
[15] J. Y. Lo, J. A. Baker, P. J. Kornguth, and C. E. Floyd, Jr., "Computer-aided diagnosis of breast cancer: artificial neural network approach for optimized merging of mammographic features," *Academic Radiology*, vol. 2, pp. 841-50, 1995.
[16] J. Y. Lo, J. A. Baker, P. J. Kornguth, and C. E. Floyd, Jr., "Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks," *Academic Radiology.*, vol. 6, pp. 10-5, 1999.
[17] J. Y. Lo, M. K. Markey, J. A. Baker, and C. E. Floyd, Jr., "Cross-institutional evaluation of BI-RADS predictive model for mammographic diagnosis of breast cancer," *American Journal of Roentgenology*, vol. 178, pp. 457-63, 2002.
[18] M. K. Markey, J. Y. Lo, R. Vargas-Voracek, G. D. Tourassi, and C. E. Floyd, Jr., "Perceptron error surface analysis: a case study in breast cancer diagnosis," *Computers in Biology & Medicine*, vol. 32, pp. 99-109, 2002.
[19] D. L. Donoho and X. Huo, "Beamlets and Multiscale Image Analysis," in *Springer Lecture Notes in Computational Science and Engineering*, vol. 20, J. Barth, T. Chan, and R. Haimes, Eds., 2001, pp. 149-196.
[20] M. Tuceryan and A. K. Jain, "Texture Analysis," in *The Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds.: World Scientific Publishing Co.,, 1998, pp. 207-248.
[21] D. B. Kopans, *Breast Imaging*: Lippincott-Raven Publishers, Philadelphia, 1998.
[22] M. Heath, K. W. Bowyer, and D. Kopans, "Current status of the Digital Database for Screening Mammography," presented at Digital Mammography, Dordrecht, 1998.
[23] D. L. Donoho and X. Huo, "BeamLab and Reproducible Research," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, pp. 391-414, 2004.
[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.
[25] J. A. Baker, P. J. Kornguth, and C. E. Floyd, Jr., "Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description," *AJR. American Journal of Roentgenology*, vol. 166, pp. 773-8, 1996.
[26] K. Kerlikowske, D. Grady, J. Barclay, S. D. Frankel, S. H. Ominsky, E. A. Sickles, and V. Ernster, "Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System," *Journal of the National Cancer Institute*, vol. 90, pp. 1801-9, 1998.
[27] W. A. Berg, C. Campassi, P. Langenberg, and M. J. Sexton, "Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment," *AJR. American Journal of Roentgenology*, vol. 174, pp. 1769-77, 2000.
[28] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159-74, 1977.