Resolution effects on the morphology of calcifications in digital mammograms

Maria Kallergi, Li He, Marios Gavrielides, John Heine, and Laurence P. Clarke

Department of Radiology, College of Medicine, and H. Lee Moffitt Cancer Center & Research Institute at the University of South Florida
12901 Bruce B. Downs Blvd., Box 17, Tampa, FL 33612

Corresponding author: Maria Kallergi, Ph.D.
E-mail: kallergi@rad.usf.edu

ABSTRACT

The development of computer assisted diagnosis (CAD) techniques and direct digital mammography systems have generated significant interest in the issue of the effect of image resolution on the detection and classification (benign vs malignant) of mammographic abnormalities. CAD in particular seems to heavily depend on image resolution, either due to the inherent algorithm design and optimization, which is almost always resolution dependent, or due to the differences in image content at the various resolutions. This twofold dependence makes it even more difficult to answer the question of what is the minimum resolution required for successful detection and/or classification of a specific mammographic abnormality, such as calcifications. One may begin by evaluating the losses in the mammograms as the films are digitized with different pixel sizes and depths. In this paper we attempted to measure these losses for the case of calcifications at four different spatial resolutions through a simulation model and a classification scheme that is based only on morphological features. The results showed that a 60 µm pixel size and 12 bits per pixel should at least be used if the morphology and distribution of the calcifications are essential components in the CAD algorithm design. These conclusions were tested with the use of a wavelet-based algorithm for the segmentation of simulated mammographic calcifications at various resolutions. The evaluation of the segmentation through shape analysis and classification supported the initial conclusion.

I. INTRODUCTION

Approximately 178,700 women will be diagnosed with breast cancer this year in the United States alone [1]. Mammography is currently the only proven test that detects breast cancer at early stages and has led to reduced mortality rates [2]. CAD aims at the improvement and standardization of conventional mammogram interpretation. With the introduction of computer monitors in the reading of mammograms, CAD has also assumed the role of improving digital reading compensating for low monitor resolution and luminance [3]. The effect of image pixel size and depth on the performance of CAD methodologies is a topic of significant interest.
and one previously addressed in the literature [4-8]. Chan et al [5-7] have studied the effect of pixel size on the detection and classification of calcifications and showed that cluster detectability is not compromised by spatial resolutions of the order of 100 µm but that classification accuracy tends to decrease as the pixel size increases from 35 µm [6,7]. They noted that for classification, the effect depends on the features used and that morphological features showed the strongest dependence [7]. Two questions remain unanswered, however: (a) what exactly are the losses in the mammograms as pixel size increases and (b) how do these losses affect the step preceding classification, namely the segmentation of calcifications. This study addresses both questions for the case of calcifications by (a) determining the losses in the calcification morphology for various pixel sizes through a simulation model and (b) evaluating an automatic segmentation technique through shape analysis and classification.

The logic behind this work can be summarized in the following: Calcifications are found in almost 50% of all mammograms. Morphology and distribution of calcifications are two of the most important clinical elements in the differentiation between benign and malignant cases [11]. CAD methods that make use of morphological/distributional characteristics of calcifications are desirable. In a fully automatic CAD method, segmentation, detection, and classification steps are combined in a way that leads to the final diagnosis through the use of information found in the various outputs, particularly that of segmentation. Segmentation can be done manually by an expert [12] or automatically [7,9]. The former has the disadvantage that it is operator dependent and subject to the large variability observed among mammographers, something that is avoided with the automatic segmentation approach. The goal of any automatic segmentation technique in mammography is twofold: First to extract the individual calcifications preserving their distribution so that detection is successful and second to extract the individual calcifications preserving their shape so that morphology-based classification is successful.

The evaluation of the segmentation output (manual or automatic) can simply be done by comparing true and segmented objects of interest in terms of distribution, size and shape preservation. In the case of digital mammography and calcifications in particular, there are two problems with this approach: (a) There is no absolute truth available which prompts the need for simulation. Simulation can provide the so-needed absolute ground truth for calcification morphology and distribution [10]. (b) Errors in segmentation are not necessarily translated into errors in detection and/or classification. So, one should also look at the detection and classification outcomes if these are tightly connected to the segmentation.

Classical shape analysis is adequate in automatically describing the morphology and distribution of mammographic calcifications [13]. Features such as area, compactness, moments, eccentricity, spread, density, etc, are well known and representative morphological features. Hence, any trend observed in these features with increasing or decreasing resolution should be a representative behavior of this class of features and the morphology of the object of interest.

Based on the above assumptions, we studied the effect of pixel size on the morphological characteristics of calcifications and on a wavelet-based segmentation. Both effects were studied...
by generating digital mammograms with simulated calcification clusters at four different spatial resolutions, by analyzing the calcifications and the clusters with shape analysis, and by classifying the raw and segmented calcification clusters using morphological features only and a neural network. The following sections describe our approach in detail and present the results and conclusions of the study.

II. SIMULATION MODEL

A model was developed for simulating benign and malignant calcification clusters of types described in the Breast Imaging Reporting and Data System (BIRADS) of the American College of Radiology (ACR) [9,10]. Details of the simulation model can be found elsewhere [10]. Briefly, 30 normal mammograms (single views) were digitized with an ImagClear R3000 scanner (DBA, Melbourn, FL) at 30 µm and 16 bits per pixel. Computer generated clusters of objects with morphologies representing different calcification shapes (round, coarse, dystrophic, pleomorphic, amorphous, etc) were superimposed on the digitized mammograms. Fifteen of the clusters consisted of typically benign types of calcifications and 15 consisted of typically malignant types. For our discussion, the mammograms containing the cluster will be referred to as simulated cases while the raw objects will be referred to as raw binary data. The intensity profile of each object within a cluster was modeled by a Gaussian filter with amplitude and standard deviation dependent on the background (normal tissue) on which the object was embedded. The simulated cases were evaluated in a blind study by two mammographers. The clusters that were identified as simulated were modified and reevaluated. The initial, computer binary pattern of the object group generated at a resolution of 30 µm provided the absolute ground truth [10].

Three lower resolution sets of images were mathematically generated from the 30 µm binary data: 60, 90, and 120 µm with the same pixel depth of 16 bits. Three sets of lower resolution images were also generated from the simulated 30 µm cases in a similar way.

III. SEGMENTATION METHODOLOGY

Several wavelet-based segmentation methods have been developed in our laboratory for the segmentation of calcifications from digitized mammograms at various resolutions. The algorithm used in this study included a symmlet wavelet with 12 coefficients [14]. The images were decomposed down to the 4th band, and the 2nd and 3rd bands were kept as the output images which were further similarly thresholded based on their histograms. The algorithm was initially optimized for 60 µm images and then it was applied to different resolution data. The reason was that we wanted to test and evaluate the robustness of the method across different image sets and determine the factors that affect its performance. The entire mammogram can be processed with our method. In this study, however, only 512×512 sections of the mammograms containing the cluster were used to facilitate the subsequent shape analysis and classification; these sections will still be referred to as simulated cases.
Hence, after the wavelet-based segmentation of the simulated cases, we had a total of 8 sets of 30 images: Four sets of raw binary data (at 30, 60, 90, 120 µm pixel sizes) and four sets of wavelet-segmented data (at 30, 60, 90, 120 µm). All 8 sets were analyzed and classified with the methods described below in order to determine the effect of pixel size on morphology and automatic segmentation.

IV. SHAPE ANALYSIS AND CLASSIFICATION

Six shape factors were calculated for each segmented calcification: area ($A$), compactness ($C$), moments ($M$), Fourier descriptors ($FD$), eccentricity ($\varepsilon$), and spread ($S$). It has been shown that these features are robust descriptors of the morphology of calcifications [13]. In addition, with the exception of the area, they are scale invariant features, a property of particular importance in our study. The mean and standard deviation (SD) of each factor were estimated for each cluster (binary or segmented) at the four resolutions. Hence 12 features were calculated for each cluster.

To determine the effect of pixel size on the morphological characteristics of calcifications (either raw or segmented), the 12 features of the raw binary data were compared at each resolution. The deviation (% error) from ground truth (30 µm raw binary data) is an indication of the resolution effects on the morphological characteristics. Another indication was provided by the success of the classification of these clusters as benign or malignant based on the 12 cluster features. These features were used as inputs to a three layer, feed-forward neural network [9]. A leave-one-out subsampling methodology was used in the evaluation of the classifier due to small sample size. For the segmented cases, the true positive (TP) detection rate was estimated at all four resolutions. A case was considered as TP if at least half of the calcifications in the cluster were segmented.

V. RESULTS AND DISCUSSION

A visual comparison of the raw binary data and segmentation results showed that in the higher resolution data (a) the margins of the individual calcifications, (b) the relative magnitudes of the calcifications in the cluster, © the number of the calcifications in the cluster, and (d) the size of the calcifications were better preserved.

Table 1 lists the classification and detection results obtained from the 8 sets of images. The classification of the clusters dropped with increasing pixel size in agreement with Chan et al [7]. However, similar performances were obtained from the 30 and 60 µm data. This result is significant because it established that the minimum required resolution for detection and classification can be 60 µm instead of 30 µm, thus reducing digital image size, computational time, and storage requirements significantly. This seems to be a general result for morphology-based algorithms and not dependent on the segmentation methodology used in this study. It may also be true for algorithms using other features for classification, e.g. texture, since statistical analysis of the digitized mammograms at the various resolutions suggested that images with pixel
sizes in the 30-70 µm range contain similar information. Differences were observed for larger pixel sizes.

A relative comparison of the values of the 12 shape descriptors obtained for the binary and the segmented sets of images showed that as the pixel size increased: (a) The area of the individual calcifications did not change significantly but some microcalcifications were entirely missed. (b) The compactness decreased significantly and in some cases it was not possible to calculate because calcifications were reduced to single or double pixel areas and were disregarded for this measurement. The decrease in $C$ indicated that at low resolutions more and more objects approximated a circle because of loss of detailed boundary information. © Moments increased for large, usually benign, calcifications as pixel size was increased from 30 to 90 µm but dropped for the 120 µm set. A steady decrease was observed for small, usually malignant, calcifications. (d) Fourier descriptors increased steadily with resolution. (e) Eccentricity remained the same for 30, 60, and 90 µm images and increased slightly for the 120 µm data. (f) Spread did not change significantly with resolution. (g) The SD of the features was generally larger for the malignant, an indication of larger variability, than the benign clusters. SDs increased with increasing pixel size but the benign/malignant relation remained unchanged. (h) Finally, the % error of the measurements from the ground truth data (30 µm raw binary images), increased significantly as the pixel size increased.

The comparison of the shape features obtained from the 60 µm segmented data to those obtained from the 60 µm raw binary images showed that the automatic segmentation does not fully preserve the morphology of the calcifications yielding a lower success rate in classification using the same features. This resolution was of primary interest because the algorithm was optimized for this particular image set and hence it was expected to have the best performance. The classification error rate was indeed the lowest for this image set. As expected, the performance at other resolutions, lower or higher, dropped significantly as indicated by the significantly higher classification error rates in Table 1 suggesting significant losses in the morphology and distribution of the segmented calcifications.

| Table 1: Detection and classification results for the 8 sets of images discussed earlier. The error rate of the classifier is listed as well as the TP detection rate. The threshold value used to separate the benign from the malignant classes is also listed. Various threshold were evaluated. The one listed here corresponds to the best classifier performance. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Image set**   | **Resolution**  | **True Positive** | **Classification** | **Classification** |
|                 | **(µm)**       | **rate**         | **error rate (%)** | **Threshold**    |
| Raw binary      | 30             | 100              | 7                | 0.5             |
| clusters        | 60             | 100              | 8                | 0.5             |
|                 | 90             | 100              | 10               | 0.5             |
|                 | 120            | 100              | 16               | 0.3             |
It may be argued that changes in certain morphological features are due to the mathematical averaging used for the generation of the lower resolution images. Other interpolations may produce different results and should be considered. Losses, however, are expected with any approach. The question is whether they will affect the classification outcome. One may also argue that the results depend on the way the shape features are calculated. For example there are different ways in calculating the compactness, the perimeter, or the outline of an object. Our experiments showed that different absolute numbers may be obtained but the overall behavior remains unchanged. The final argument comes for the segmentation method. Namely, results depend on the method used, its optimization at different resolutions, and the features entering the classification. This may be true but the first part of our study addressed the issue in a more general way, i.e., independent of segmentation algorithm. Binary data can be thought as corresponding to segmentation data with a technique that is not introducing additional errors other than the errors due to pixel averaging, something that is inherent to digital images. Furthermore, this study was limited to the use of shape features as being the most pertinent to the issue of morphology. Other features may yield different results but their evaluation was beyond the scope of this work. The second part of our study supported the conclusions of the first one for a specific segmentation method optimized for a specific resolution and showed the need for separate optimization for different types of images at least as far as wavelet-based segmentation is concerned. However, current and ongoing evaluation of segmentation results obtained from methods optimized for each resolution separately show similar behavior. Namely, segmentation data at resolutions greater than 60 µm do not hold sufficient morphological information for the calcifications to provide adequate classification. Hence, it is expected that CAD algorithms based on morphological characteristics derived from segmented data will exhibit similar behavior.

REFERENCES


