Mammographic Mass Detection Using Unsupervised Clustering in Synergy with a Parcimonious Supervised Rule-Based Classifier

Michel Bruynooghe

University Louis Pasteur of Strasbourg,
Yorckstrasse 49, 76185 Karlsruhe, Germany
Michel.Bruynooghe@t-online.de

Abstract. We develop a novel CAD detection system that can help a radiologist to detect masses in mammograms. The proposed algorithm concurrently detects the breast boundary and the pectoral muscle. Then, a clustering and morphology based segmentation algorithm is applied to the enhanced mammography image to separate the mass from the normal breast tissues. This technique outlines the shape of candidate masses in mammograms. To maximize detection specificity, we develop a two-stage hybrid classification network. First, an unsupervised classifier is used to classify suspicious opacities as suspect or not. Then, a few supervised interpretation rules are applied to further reduce the number of false detections. Using a private mammography database and the publicly available USF/DDSM database, experimental results demonstrate that a sensitivity of 94% (resp. 80%) can be achieved at a specificity level of 1.02 (resp. 0.69) false positives per image. Even in dense mammograms, the CAD algorithm can still correctly detect subtle masses.

1 Introduction

Classical computer-assisted detection of masses in mammographic images generally requires a multistage algorithm that includes detection of candidate masses, pattern recognition techniques to classify the candidate objects, and a method to eliminate false detections and to determine if a mass exists. Wavelet or morphological techniques are generally used to enhance the mammography image. Supervised classifiers such as neural networks, fuzzy neural classifiers, bayesian classifiers or rule-based classifiers are applied to discriminate between normal and suspicious objects. These classical supervised classifiers generally require the learning of a rather large number of parameters.

Specificity levels of automatic mass detection methods in mammography are generally rather low. To improve detection scores, te-Brake, Karssemeijer and Hendriks [4] have introduced features that are related to image characteristics that radiologists use to discriminate real lesions from normal tissue. Approximately 75% of all cancers were detected in at least one view at a specificity level of 1.0 false positive per image.

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 68–75, 2006.

Yates, Evans and Brady [22] have proposed a pre-processing step for improving te-Brake's mass-detection algorithm. Their method, that is based on wavelets and phase congruency, removes the locally linear fine detail structure, whilst retaining the larger underlying mass structure. The resulting ROC curve has shown that this removal technique improves the mass detection rate. Hong and Brady [11] have proposed a segmentation method for delineating regions of interest in mammograms. Their algorithm concurrently detects the breast boundary, the pectoral muscle, and dense regions that include candidate masses. A topographic representation called the iso-level contour map has been used to estimate the saliency of each suspicious region. This method has achieved a satisfactory performance as a prompt system for mass detection.

An adaptive, multiscale method for mass detection using a circular ring template and a nonparametric test has been presented by Khan et al. [12]. Training data corresponding to the local background intensity level is extracted from the outer ring of the template while test data for mass detection is obtained from the inner disk. Experimental results show a sensitivity of 1.0 and a false positive rate of 1.40 per image on 30 images. Detection algorithms often fail to detect masses with a partial loss of region, that are located on the edge of the film. To overcome this problem, Hatanaka et al. [8] have proposed to identify partial loss masses by their similarity to a sector-form model in the template matching process. The true-positive fraction is 0.97 (resp. 0.84) when the number of false positives (FP) is 1.20 (resp. 1.49) per mammogram on 335 (resp. 1075) digitized mammograms. Petrick et al. [18] have designed an object-based region-growing technique to improve mass segmentation. As a preprocessing step, this segmentation method utilizes the density-weighted contrast enhancement (DWCE) filter to adaptively enhance the contrast between the breast structures and the background. Each suspicious opacity is classified as a mass or normal tissue based on morphological and texture features. This segmentation scheme detected 90% (resp. 80%) of 253 biopsy-proven breast masses at a specificity level of 4.2 (resp. 2.0) false positive per image.

Cheng and Muyi-Cui [6] have recently presented a novel fuzzy neural network approach to detect malignant mass on mammograms. They analyzed 670 ROIs from mammograms of the DDSM database. The true-positive fraction is 0.92 when the number of FPs is 1.33 per mammogram. But, this FP score is underestimated because only a few ROIs have been analyzed per mammogram instead of full mammograms. Hence, it cannot be compared to other specificity scores. Heath et al. [10] have introduced a mass detection algorithm by relative image intensity that estimates the degree to which a surrounding region of a point decreases in intensity. This algorithm requires neither the training of parameters nor the normalization of images. Detection performances and FROC curves have been estimated using datasets from the MIAS and DDSM databases. Experimental results show a sensitivity of 0.65 (resp. 0.70) and a false positive rate of 1.75 (resp. 1.60) per image on 246 MIAS images (resp. 160 USF images).

Performances of other recently published CAD algorithms for mass detection are presented in figure 3. Published detection scores show that the simultaneous