Image Similarity and Asymmetry to Improve Computer-Aided Detection of Breast Cancer

Dave Tahmoush[1,2] and Hanan Samet[1]

[1] Computer Science Department, Center for Automation Research, Institute for Advanced Computer Science, University of Maryland, College Park, USA
tahmoush@cs.umd.edu, hjs@umiacs.umd.edu
[2] Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland, USA

Abstract. An improved image similarity method is introduced to recognize breast cancer, and it is incorporated into a computer-aided breast cancer detection system through Bayes Theorem. Radiologists can use the differences between the left and right breasts, or asymmetry, in mammograms to help detect certain malignant breast cancers. Image similarity is used to determine asymmetry using a contextual and then a spatial comparison. The mammograms are filtered to find the most contextually significant points, and then the resulting point set is analyzed for spatial similarity. We develop the analysis through a combination of modeling and supervised learning of model parameters. This process correctly classifies mammograms 84% of the time, and significantly improves the accuracy of a computer-aided breast cancer detection system by 71%.

1 Introduction

Breast cancer remains a leading cause of cancer deaths among women in many parts of the world. In the United States alone, over forty thousand women die of the disease each year [1]. Mammography is currently the most effective method for early detection of breast cancer [2]. For two-thirds of the women whose initial diagnosis of their mammogram is negative but who actually have breast cancer, the cancer is evident upon a second diagnosis of their mammogram [2]. Computer-aided detection (CAD) of mammograms could be used to avoid these missed diagnoses, and has been shown to increase the number of cancers detected by more than nineteen percent [3]. Measuring asymmetry, which consists of a comparison of the left and right breast images [4], is a technique that could be used to improve the accuracy of CAD. An automated prescreening system only classifies a mammogram as either normal or suspicious, while CAD picks out specific points as cancerous [5]. One of the most challenging problems with prescreening is the lack of sensitive algorithms for the detection of asymmetry [6]. This paper presents a simple and effective algorithm for the detection of asymmetry and extensions to improve upon it. We improve on our earlier results [7] and incorporate image similarity into a CAD system.

Mammograms are an excellent candidate for image similarity techniques to be effective because there are images of both the left and right breasts, which should be similar if there is no cancer present. Image similarity has been often utilized for

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 221–228, 2006.

content-based image retrieval (CBIR) from image databases [8, 9, 10, 11]. Both contextual and spatial comparisons are used [8]. Medical image databases have also used image similarity, from rule-based systems for chest radiographs [12] to anatomical structure matching for 3D MR images [13]. However, the focus is often on the non-cancerous structures, while it is the cancerous structures that are of principle interest here. In this paper we combine the image similarity concept of contextual then spatial comparison to the problem of detecting breast cancer in mammograms.

The majority of work on CAD analysis of mammograms has focused on determining the contextual similarity to cancer, finding abnormalities in a local area of a single image [14, 15]. This paper focuses on combining this with a spatial comparison in order to complete an image similarity measure. The majority of work has used methods ranging from filters to wavelets to learning techniques, but a detailed discussion of various imaging techniques is beyond the scope of this paper. Problems arise in using filter methods [14] because of the range of sizes and morphologies for breast cancer, as well as the difficulty in differentiating cancerous from non-cancerous structures. The size range problem has been addressed by using multi-scale models [15]. Similar issues affect wavelet methods, although their use has led to reported good results [16] with the size range issue being improved through the use of a wavelet pyramid [17]. Learning techniques have included support vector machines [18] and neural networks [16].

Detecting breast cancer is challenging because the cancerous structures have many features in common with normal breast tissue. This means that a high number of false positives or false negatives are possible. Asymmetry can be used to help reduce the number of false positives so that true positives are more obvious. Previous work utilizing asymmetry has used wavelets or structural clues to detect asymmetry with correct results as often as 77% of the time [4, 19]. Additional work has focused on bilateral or temporal subtraction, which is the attempt to subtract one breast image from the other [20, 21]. This approach is good because it does try to utilize the multiple images taken with the same machine by the same technician and analyzed using the same process in an effort to reduce the systematic differences that can be introduced. However, bilateral subtraction is hampered by the necessity of exact registration and natural asymmetry of the breasts. We introduce a measure of asymmetry that is more approximate in nature and seems more robust to the large amount of noise in the data, using learning to determine a highly constrained number of model parameters. Minimizing the number of parameters that are learned makes the model less subject to overfitting the noise in the data at the possible expense of accuracy.

Comparing multiple mammograms using learning techniques has been shown to be effective in CBIR [10, 22]. Our application lends itself well to supervised learning because the data set has already been screened for cancer and thus classified by expert radiologists. However, care must be taken since the expert classification is known not to be perfect [2].