Automated Segmentation and Retrieval Framework for Diagnosis of Breast Cancer From Mammograms

Ayush Shrivastava, Arpit Chaudhary, Devang Kulshreshtha, Rajeev Srivastava
Department of Computer Science And Engineering
Indian Institute of Technology (BHU) Varanasi

Introduction

Breast cancer is the most common type of cancer found among women. According to a survey, one in 22 women in India is likely to suffer from breast cancer. So, we present a fully automated CAD for segmentation and content based retrieval of mammograms. Models achieving state-of-the art accuracy aren't fully automated and due to poor segmentation process, the images are manually cropped. Our approach of segmentation segment out the ROI with good accuracy. The segmented ROI is further used for the proposed Content-Based Image Retrieval in mammograms.

Approach

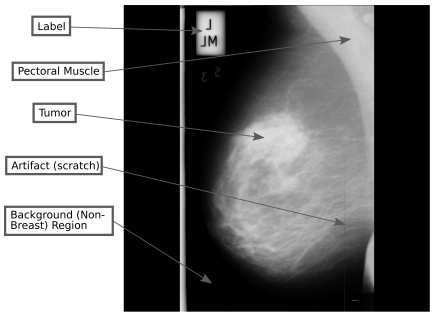

Figure 1: Anatomy of a mammogram

The automated mammogram retrieval system is broadly divided into two parts:

• **Segmentation:** Labels and tags are removed using Morphological Opening method. Pectoral muscle is removed using our proposed Pectoral Muscle Sliding Window Algorithm (PMSWA) which slides a window region from left to right and top to bottom until non pectoral muscle region is found. Dispersed Region Growing Algorithm(DRGA) is used for segmentation of mammogram which disperses seeds in different regions instead of a single bright region to take out ROI.

• **Retrieval:** Proposed idea is motivated by the recent works of Li-Lui and S.Liao in which LBP-based features are successfully exploited for feature representation on the texture classification problem.

Clustering without known predefined classes play an excellent role in mammogram image retrieval. This reduces the overall retrieval time without affecting the performance of the system.

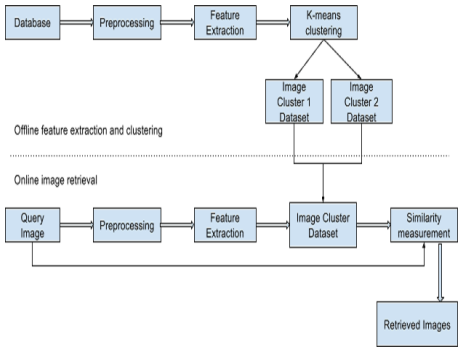

Figure 2: Proposed mammogram retrieval framework

• **Classification** Classification of mammogram into normal and abnormal classes are performed using Random Forest classifier with no. of trees set to 100.

Results

• **Dataset:** We have used MIAS database containing 330 mammogram images of 1024x1024 pixels and their corresponding ground truth.

• **Segmentation Quality:** Segmentation approach got 294 images correctly segmented and other 28 having slight deviation from correct ROI resulting in accuracy of 91.3

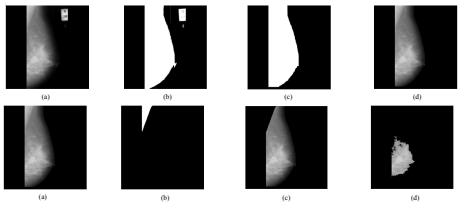

Figure 3: Suppression of artifacts and labels and Extraction of ROI

• **Retrieval Performance:** Performance analysis is done by estimating the mean precision rate and the computational time for retrieval. As shown in Figure 5, for any value of K the use of LBP outperforms SURF,SIFT and GLCM based feature extraction techniques.

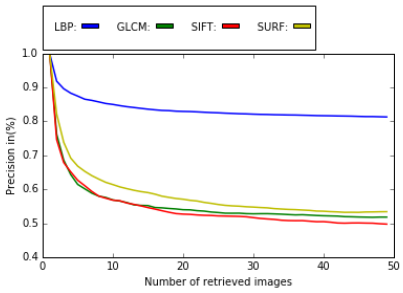

Figure 4: Comparative analysis

Figure 5: Computation time

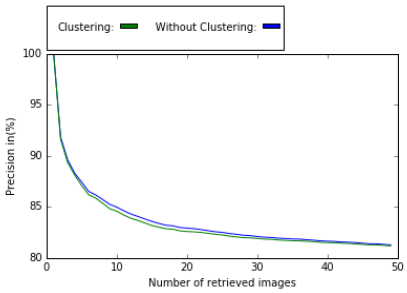

Figure 6: Retrieval performance clustering and non-clustering

There has been almost no effect on the precision for any value of K (number of retrieved images). Hence, our approach has not lead to significant decline in the accuracy.

Conclusions

In this work, we prosposed Pectoral Muscle Sliding Window Algorithm (PMSWA) for pectoral muscle removal and Dispered Region Growing Algorithm (DRGA) for segmentation of mammograms to segment ROI. On these segmented images, LBP feature extracion technique along with K-means clustering is used for content-based image retrieval.