ABSTRACT

Micorcalcification is an early index of breast cancer. Computer Aided Diagnose (CAD) system can help users to diagnose microcalcification clusters. When users have doubt on the output of a CAD system, it is not able to provide a second opinion. To solve this problem, we develop a content-based image retrieval (CBIR) system of mammographic microcalcification clusters. The goal of the proposed CBIR system is to provide a set of pathology verified images relevant to the query image as the reference for users. The two essential aspects of CBIR are the feature extraction and the definition of similarity. In this thesis, thirty-five features are extracted from each cluster. Since not all features are useful, the Sequential Forward Selection (SFS) technique is applied to select 6 most useful features for input feature vector of the CBIR system. In this thesis, we design investigate three types of similarity indices to evaluate the system performance. The first is to calculate the angle between the feature vectors extracted from the query image and the images in the database. The second is to calculate the Euclidean distance between the feature vectors extracted from the query image and the images in the database. The last, we applied a general regression neural network (GRNN) based learning algorithms. To provide an objective measurement for the desire output, we apply the angle between the feature vectors extracted from the query image and the images in the database as the desire output value for GRNN training. In addition, we formulate a linguistic query system, which allow users to query images by inputting features though semantic query. By applying the images from the Nijmegen Digital Mammogram Database, experimental results showed that the GRNN, with precision rate 0.9405 and recall rate 0.8095, out performs the similarity indices formulated by the angle and Euclidean distance.

Keywords: Content-Based Image Retrieval; Microcalcification Clusters; General Regression Neural Network; Linguistic Query