Content based retrieval for mammographic mass

Abstract

Content-based image retrieval (CBIR) is a technique that uses visual contents such as color, shape, texture and spatial layout to retrieve image from large scale image database, given a query image. To automate CBIR system for medical images, algorithms are to be developed that can automatically detect regions of interest (ROIs) and retrieve semantically and perceptually similar images to provide evidence-based decision support. Developing automated algorithms for retrieval of biomedical images by structural contents is a significant research challenge, because ROIs are commonly irregular, overlapping, partially occluded or highly localized. Our work is based on content based retrieval of cervical spine images and mammographic mass.

Extensive research work is going on at National Library of Medicine (NLM) on indexing and content based retrieval of spine x-ray images, collected in the second National Health and Nutrition Examination Survey (NHANES II) conducted by U.S. National Center for Health Statistics (NCHS). An algorithm is proposed to automatically locate contour points of a geometric 8-point model of spine vertebra for region based partial shape feature extraction (at the anterior corners) in an unambiguous way. Multiple distance metrics are investigated for retrieval. The retrieval performance of proposed spine image retrieval technique has been compared with different existing techniques when tested on a subset of NHANES II collection of data.

ROI in a mammogram image typically consists of gray level variations in highly localized regions of the image. A method to automatically localize (i.e detect and segment) mass in mammograms with reduced computational load is proposed. Free response receiver operating characteristic (FROC) plots for the proposed mass detection approach and two existing mass detection methods are compared when tested on direct radiography (DR) images and scanned films from mini-MIAS database. The effectiveness of the proposed method is found to be significantly better. More shape and texture based features reported in literature, that characterize mass are investigated for automatically segmented mass contours in mammograms. Features for retrieval are selected by using Fisher's ratio criterion. Based on automatically segmented mass contours, *precision vs recall* plot with selected features and that obtained using benchmark feature set are compared. Multiple distance metrics are investigated for retrieval.

Keywords: Content-based image retrieval; precision; semantic features; feature extraction; cervical spine; geometric model; mammographic mass; mass detection; segmentation; Free response receiver operating characteristic; Fisher's ratio.