Chapter 4

Mammogram Retrieval: Image Selection Strategy of Relevance Feedback for Locating Similar Lesions

Chee-Chiang Chen
Tungs' Taichung MetroHarbor Hospital, Taiwan

Pai-Jung Huang
Tungs’ Taichung MetroHarbor Hospital, Chi Mei Foundation Hospital, Taipei Medical University Hospital, & Taipei Medical University, Taiwan

Chih-Ying Gwo
Ching Yun University, Taiwan

Yue Li
Nankai University, China

Chia-Hung Wei
Ching Yun University, Taiwan

ABSTRACT

Content-based image retrieval (CBIR) has been proposed by the medical community for inclusion in picture archiving and communication systems (PACS). In CBIR, relevance feedback is developed for bridging the semantic gap and improving the effectiveness of image retrieval systems. With relevance feedback, CBIR systems can return refined search results using a learning algorithm and selection strategy. In this study, as the retrieving process proceeds further, the proposed learning algorithm can reduce the influence of the original query point and increase the significance of the centroid of the clusters comprising the features of those relevant images identified in the most recent round of search. The proposed selection strategy is used to find a good starting point and select a set of images at each round to show that search result and ask for the user’s feedback. In addition, a benchmark is proposed to measure the learning ability to explain the retrieval performance as relevance feedback is incorporated in CBIR systems. The performance evaluation shows that the average precision rate of the proposed scheme was 0.98 and the learning ability reach to 7.17 through the five rounds of relevance feedback.
1. INTRODUCTION

Content-based image retrieval (CBIR) refers to the retrieval of images whose contents are similar to a query example, using information derived from the images themselves, rather than relying on accompanying text indices or external annotation (El-Naqa, Yang, Galatsanos, Nishikawa, & Wernick, 2004). One of the key challenges in CBIR is bridging the gap between low-level representations and high-level semantics. The semantic gap exists because low-level features are formulated in the system design process while high-level queries are used at the starting point of the retrieval process (Lew, Sebe, & Eakins, 2002). Relevance feedback is developed for bridging the semantic gap and improving the effectiveness of image retrieval systems (El-Naqa et al., 2004; Wei & Li, 2008). With relevance feedback, CBIR systems can return refined search results using a learning algorithm and selection strategy.

Content-based image retrieval has been proposed by the medical community for inclusion into picture archiving and communication systems (PACS) (Lehmann et al., 2004). The idea of PACS is to integrate imaging modalities and interfaces with hospital and departmental information systems in order to manage the storage and distribution of images to radiologists, physicians, specialists, clinics, and imaging centres (Huang, 2003). A crucial requirement of PACS is to provide an efficient search function for accessing images that are relevant to the query example. The contents of medical images provide useful information, which can be used to search for other images containing similar content.

An enormous number of digital mammograms have been generated in hospitals and breast screening centres in recent years. As hospitals and breast screening centres are connected together through PACS, content-based approaches can be applied to efficiently retrieve mammograms from distributed databases. However, content-based retrieval approaches are usually developed for specific contents of medical images. Given this motivation, the goal of this study is to develop a novel and complete scheme for incorporating the relevance feedback into a content-based mammogram retrieval system.

The rest of the paper is organized as follows. Prior work is reviewed and described in Section 2. An overview of the proposed content-based retrieval framework is described in Section 3. The proposed learning algorithm and selection strategy for relevance feedback is presented in Sections 4 and 5, respectively. Section 6 evaluates the retrieval performance, and Section 7 discusses the results of the experiments. Section 8 makes a conclusion on this study.

2. PRIOR WORK

To incorporate relevance feedback into content-based image retrieval, two main approaches are developed: the query point movement approach and re-weighting approach. The concept behind the first approach is to modify the query that is originally submitted by the user. It is assumed that there exists at least one image which completely conveys the intentions of the user, and its high-level concept has been modeled in low-level feature space (Kushki, Androutsos, Plataniotis, & Venetsanopoulos, 2004). The query point movement approach is to move the point of the query toward the region of the feature space that contains the ideal image (Zhong, Hongjiang, Li, & Shaoping, 2003). Based on this concept, the classic Rocchio algorithm was originally developed to improve the effectiveness of information retrieval system (Rocchio, 1971). The MARS system has applied the query point movement approach as one of methods for relevance feedback (Rui, Huang, & Mehrotra, 1998). The method used in the MARS system is called \( tf \times idf \) (term frequency-inverse document frequency), which generates pseudo-document vectors from image