Chapter 12
Content-Based Access to Medical Image Collections

Juan C. Caicedo
National University of Colombia, Colombia

Jorge E. Camargo
National University of Colombia, Colombia

Fabio A. González
National University of Colombia, Colombia

ABSTRACT
Medical images are a very important resource for the clinical practice and operation. Thousands of them are daily acquired in hospitals to diagnose the health state of patients. However, once they have been archived as part of a large image database, it is very difficult to retrieve the same image again since it requires remembering dates or names. Furthermore, in such cases in which the same image is not required, but a physician is looking for images with particular contents, current technologies are not able to offer such functionality. The ability to find the right visual information in the right place, at the right time, can have great impact in the medical decision making process. This chapter presents two computational strategies for accessing a large collection of medical images: retrieving relevant images given an explicit query and visualizing the structure of the whole collection. Both strategies take advantage of image contents, allowing users to find or identify images that are related by their visual composition. In addition, these strategies are based on machine learning methods to handle complex image patterns, semantic medical concepts, image collection visualizations and summarizations.

INTRODUCTION
Large amounts of medical images are produced daily in hospitals and health centers. For instance, the University Hospital of Geneva reported a production of 70,000 images per day during 2007 in the Radiology department alone (Pitkanen et al., 2008). The management of those large image collections is a challenging task nowadays, mainly due to the ability of accessing the image database for obtaining
useful information. The computational power required to archive and process the image database has been raising during the last few years, making it possible to store large collections of medical images in specialized systems such as Picture Archiving and Communication Systems (PACS). These systems may be extended to archive more digital images according to the hospital’s needs and usually they also support the work flow in the radiology department as well as in other specialized services. However, even though the capacity is expanded, the functionality of these systems remains static and still provides very basic operations to query and search for medical images.

The contents of a large image collection in medicine may be used as a reference set of previously evaluated cases, with annotations associated to diagnosis and evolution of patients. Then, a physician that is attending a new patient may check out medical records from other patients, evaluated by other experts, and hence clinical decisions can be highly enriched by the information stored in the database. In addition, clinical training in medical schools may be supported by these real reference collections, allowing students and professors to access the experience accumulated from thousands and thousands of cases previously diagnosed. The actual problem is how to query and explore the collection in an effective way, that is, with an immediate relevant response. The first approach that may be considered is the use of textual annotations through a standard information retrieval system, so users are enabled to write several keywords associated to an information need. However, a collection of medical images does not necessarily have complete and descriptive annotations for all images, so this method would prevent full access to the database. Furthermore, users are not always quite aware of their information needs in terms of keywords, and this would lead to a trial-and-error loop for finding right answers from the system. The good news are that physicians may have example images from the current case to query the system, which include the kind of visual patterns that they are interested in.

Content-based Image Retrieval (CBIR) is an interesting alternative to support the decision making process in a clinical work-flow. CBIR systems are designed to search similar images using visual contents instead of associated data (Datta et al., 2008). So, given an example image, the system should be able to extract visual features, structure them and find the semantically matching images in the database. This approach is known as the Query-By-Example (QBE) paradigm for image search, which has been widely studied. In general, a CBIR system has to consider two main aspects in order to provide that functionality: (1) image content representation and (2) similarity measures. Content representation is related to image processing methods and feature extraction algorithms, and aims to identify characteristic image descriptors such as points, regions or objects. Ideally, image descriptors should clearly match real life objects and concepts, but in practice it is very difficult to get such a result due to the semantic gap, i.e. the lack of coincidence between extracted features and human interpretations (Smeulders et al., 2000). On the other hand, similarity measures are needed to accurately distinguish images that share the same features, so that the system will be able to recommend the most similar images to a user, if an example query image is provided.

There are several scenarios in which a physician may have an example image to query the system, for instance, when attending new patients or reading electronic papers. However, having an example image at hand to query the system whenever relevant images are needed is not always the case, for instance, in the middle of a lecture. This problem is known as the zero page problem of the query-by-example paradigm in information retrieval (La Cascia et al., 1998). In such cases, the common alternative is to provide keyword-based search functionality, but the problem of image annotations pops up again. So, the question is: how to supply content-based access to a large medical image collection for physicians, professors and students that would like to obtain examples of biological structures, diseases or diagnosis,