How to Cope with the Performance Gap in Content-Based Image Retrieval Systems

Agma J. M. Traina, University of São Paulo at São Carlos, Brazil
Caetano Traina Jr., University of São Paulo at São Carlos, Brazil
Cristina D. A. Ciferri, University of São Paulo at São Carlos, Brazil
Marcela X. Ribeiro, University of São Paulo at São Carlos, Brazil
Paulo M. Azevedo-Marques, University of São Paulo at Ribeirão Preto, Brazil

ABSTRACT
This paper discusses the main aspects regarding the performance gap in Content-based Image Retrieval (CBIR) systems, which is an important issue regarding their acceptability. We also detail the main problems that lead to the performance gap: the use of many features to represent images, the lack of appropriate indexing structures for images and features, deficient query plans employed to execute similarity queries, and sometimes the poor quality of results obtained by the CBIR system. We present guidelines to overcome these problems by employing feature selection techniques to beat the "dimensionality curse", by using proper access methods to support fast and effective indexing and retrieval of images, by stressing the importance of using query optimization approaches and by including the user during the tuning of the CBIR system through relevance feedback techniques.

Keywords: content-based image retrieval (CBIR); feature selection; image indexing; performance gap; semantic gap

INTRODUCTION
One of the main challenges in medical systems is how to efficiently take advantage of all the information gathered by these systems, in order to benefit the diagnosis and treatment of patients in a timely manner. This challenge is even bigger when considering the large volume of images that are daily produced by the devices during the process of image diagnosing in hospitals and medical centers. The procedure of finding a particular image in a database considering
only its intrinsic characteristics is called Content-based Image Retrieval (CBIR). The core of CBIR systems is the definition of which characteristics or features should be employed to properly identify a given image. Traditionally, features considering the color distribution, texture and shape of the objects/regions of the image, as well as the relationship among image objects are employed to characterize an image (Ortega-Binderberger & Mehrotra, 2004). The features are grouped in a feature vector, which is employed by the CBIR system to search the database to find the images most similar to a given one. For example, a CBIR system can answer queries such as: “Given the Thorax-XRay of John Doe taken on January 5, 2008, which are the 10 images most similar to it?”. Therefore, CBIR systems are expected to retrieve images assessing their similarity regarding the extracted features, in contrast to the practice of comparing elements by equality or ordering in traditional systems.

Database Management Systems (DBMS) are largely employed when the data is simple, as numbers and small character strings. For this kind of data, there are several highly effective techniques to represent search conditions and to achieve fast and precise answers. However, when the data is more complex, such as images from medical exams, there are several issues not yet fully addressed by the existing technology, leading to large divergences between what the user wants to retrieve and what the current technological state of the art can provide. This dichotomy is often called a gap.

One of the most well-known and prominent examples is the semantic gap, extensively mentioned in the literature (Y. Liu, Zhang, Lu, & Ma, 2007; A. J. M. Traina, Marques, & Traina, 2006). Applied to images, the semantic gap is “the disparity or discontinuity between human understanding of images and the comprehension that is obtainable from computer algorithms” (Deserno, Antani, & Long, 2008). However, as it was pointed out in (Deserno et al., 2008), there are several other gaps that affect CBIR systems, and the so-called performance gap is one of foremost importance. The term Performance gap refers mainly to the following potential problems:

- divergence between what the user expects from the system and what the system provides in terms of effective search resources available (such as ways to express and refine queries);
- effective use of the resources available (such as time and memory to answer a query); and
- integration of the CBIR tools to other facilities in the health center (such as to other software systems and imaging equipment).

In this paper we highlight the main problems that lead to the performance gap and present a survey of existing techniques aimed at bridging it. The remainder of this paper is organized as follows. The next section discusses the main faces of the performance gap that occur in CBIR systems, presenting a general architecture of those systems and identifying the performance issues that can arise from each of its components. The performance gap section presents the main research efforts being pursued to improve the system performance, both regarding the inner structures supporting CBIR as well as regarding techniques that empower the final user to interact with the system. The discussion and examples section illustrates recent techniques being developed to cope with the most important aspects of the performance gap, showing how the performance gap is being bridged. Finally, conclusions of the concepts presented in the paper are presented.

STORING AND SEARCHING IMAGES IN A CBIR SYSTEM

To identify how and where the performance issue of a CBIR system affects its usability, let us describe the main steps executed when the user poses a query in such a system. In fact, the CBIR process starts long before a query is
The E-Learning Experience in Medical School of Casablanca
www.igi-global.com/chapter/the-e-learning-experience-in-medical-school-of-casablanca/115117?camid=4v1a