Chapter 4
Issues and Techniques to Mitigate the Performance Gap in Content-Based Image Retrieval Systems

Ahma J. M. Traina
University of São Paulo (USP) at São Carlos, Brazil

Caetano Traina Jr.
University of São Paulo (USP) at São Carlos, Brazil

Robson Cordeiro
University of São Paulo (USP) at São Carlos, Brazil

Marcela Xavier Ribeiro
Federal University of Sao Carlos, Brazil

Paulo M. Azevedo-Marques
University of São Paulo (USP) at Ribeirão Preto, Brazil

ABSTRACT

This chapter discusses key aspects concerning the performance of Content-based Image Retrieval (CBIR) systems. The so-called performance gap plays an important role regarding the acceptability of CBIR systems by the users. It provides a timely answer to the actual demand for computational support from CBIR systems that provide similarity queries processing. Focusing on the performance gap, this chapter explains and discusses the main problems currently under investigation: the use of many features to represent images, the lack of appropriate indexing structures to retrieve images and features, deficient query plans employed to execute similarity queries, and the poor quality of results obtained by the CBIR system. We discuss how to overcome these problems, introducing techniques such as how to employ feature selection techniques to beat the “dimensionality curse” and how to use proper access methods to support fast and effective indexing and retrieval of images, stressing the importance of using query optimization approaches.

DOI: 10.4018/978-1-60960-780-7.ch004
INTRODUCTION

A nowadays quest to database administrators and systems’ managers is how to benefit from all the data stored along the years in large clinical and medical facilities. One of the main challenges for medical systems is how to efficiently take advantage of all the information gathered by these systems, in order to improve 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 (Deselaers, Keysers, & Ney, Information Retrieval). 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 image of John Doe taken on December 5, 2010, 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 dealing with simple data, 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 (Fan, Gao, Luo, & Jain, 2008; Hare et al., 2006; Hauptmann, Yan, & Lin, 2007). Applied to images, the semantic gap corresponds to “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 utmost 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 chapter we highlight the main problems that lead to performance gaps and present a survey of existing techniques aiming at bridging it. The remainder of this chapter is organized as follows. Section 2 discusses the main performance gaps 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. Section 3 presents the main research efforts being pursued to improve performance, regarding the inner structures supporting CBIR. Section 4 illustrates recent techniques being devel-
Related Content

Simulation Modeling as a Decision-Making Aid in Economic Evaluation for Randomized Clinical Trials
www.igi-global.com/chapter/simulation-modeling-decision-making-aid/53678?camid=4v1a

E-Health as the Realm of Healthcare Quality
www.igi-global.com/chapter/health-realm-healthcare-quality/53577?camid=4v1a

Modeling Colorectal Cancer: A Stability Analysis Approach
Svetoslav Nikolov, Mukhtar Ullah, Momchil Nenov, Julio Vera Gonzalez, Peter Raasch and Olaf Wolkenhauer (2013). *Medical Advancements in Aging and Regenerative Technologies: Clinical Tools and Applications* (pp. 53-75).
www.igi-global.com/chapter/modeling-colorectal-cancer/71976?camid=4v1a

Medical Image Segmentation and Tracking Through the Maximisation or the Minimisation of Divergence Between PDFs
www.igi-global.com/chapter/medical-image-segmentation-tracking-through/46687?camid=4v1a