

# Integrated Techniques for Information Retrieval

KHALIL SHIHAB<sup>1</sup> and NIDA CHALABI<sup>2</sup>

<sup>1</sup>School of Computing & Design, Swinburne University of Technology – Sarawak Campus

<sup>2</sup>Department of Computer Science, Sultan Qaboos University, OMAN

<sup>1</sup>kshihab@swinburne.edu.my; <sup>2</sup>nida@squ.edu.om

**Abstract** - *Image retrieval systems become more important and popular because of the increased demands on these systems. Recent attempts have adapted semantic ontology and metadata languages. We, however, noted that the integration of case-based and concept-based image retrieval is of great benefit in this area. The underlying technique uses high-level features to find the most possible assignment to the presented image description. It applies fuzzy reasoning to convert the quantitative attributes into qualitative terms for indexing and retrieval. To facilitate the storage and retrieval, we used XML documents as image repository. This allows high-quality museum images and associated information to be made available over networks.*

**Keywords:** Concept-based Image Retrieval, Case-based Reasoning, Fuzzy Logic, XML Documents.

## 1 Introduction

From prehistoric times, human communication has depended upon the creation and use of image-base information. The exponential growth of web images has resulted in an enormous growth of interest in image retrieval [1].

Many different methods and techniques have been proposed for modeling and retrieving images. They largely fall into two categories, namely, concept-based methods and content-based image retrieval (CBIR) methods. But most existing search engines are using keywords-based image retrieval.

CBIR relies on retrieving stored images from a collection by comparing low features that are automatically extracted from the images themselves [2]. It was introduced in the early 1980s. In CBIR, images are indexed by their visual content, such as color, texture, shapes. Extensive experiments on CBIR systems show that low-level contents often fail to describe the high level semantic concepts in user's mind [3]. Therefore, the performance of CBIR is still far from user's expectations.

In some cases, semantics can be easily derived from our daily language. For example, garden, can be described as 'open area, plants, and green region'; and plants can also be described as 'flowers, trees, and grass. For using such simple semantics, concepts can be defined by its primitives (descriptors) that are for the medium and low level features. These descriptors form a simple vocabulary, the so-called 'object-ontology' which provides a qualitative definition of high-level query concepts.

In order to improve the retrieval accuracy of content-based image retrieval systems, research focus has been shifted from designing sophisticated low-level feature extraction algorithms to reducing the 'semantic gap' between the visual features and the richness of human semantics. In 2003, during the second Iraq war, more than 15,000 items were stolen from Baghdad's Iraq Museum. A large number of the famous artifacts in history and treasures like the beautiful carved-ivory Mona Lisa of Nimrod survived ten centuries, only to fall victim to chaos and looters, some sent by international art dealers.

This paper describes the design and prototype implementation of a novel architecture for integrated metadata and concept based indexing and retrieval of museum information. The system constitutes a virtual museum preserving some works that are lost and providing more versatile access to the images and information of lost treasures from Baghdad museum, in particular.

One of the limitations of the current image retrieval systems is the use of sound knowledge representation paradigm. This is because the domains of these systems can be hardly represented by logical formalization. Therefore, we used case-based reasoning that has been proven more effective in such weak-theory domains [2].

In order to maintain a close match between the user queries and the retrieved images, we therefore used an integrated technique based on similarity matching and fuzzy reasoning for indexing and retrieval of images. We also adopted XML case-representation to facilitate the image storage and retrieval process [3].

## 2 Concept-Based Image Indexing and Retrieval

In order to improve the retrieval accuracy of content-based image retrieval systems, research focus has been shifted from designing sophisticated low-level feature extraction algorithms to reducing the 'semantic gap' between the visual features and the richness of human semantics.

The solution for image retrieval historically has been to develop text-based ontology and classification schemes for image description. Images are indexed using different