IRM PRESS
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.irm-press.com
ITB11368

This chapter appears in the book, *Managing Multimedia Semantics*, edited by Uma Srinivasan and Surya Nepal

Chapter 1

Toward Semantically Meaningful Feature Spaces for Efficient Indexing in Large Image Databases

Anne H.H. Ngu, Texas State University, USA

Jialie Shen, The University of New South Wales, Australia

John Shepherd, The University of New South Wales, Australia

ABSTRACT

The optimized distance-based access methods currently available for multimedia databases are based on two major assumptions: a suitable distance function is known a priori, and the dimensionality of image features is low. The standard approach to building image databases is to represent images via vectors based on low-level visual features and make retrieval based on these vectors. However, due to the large gap between the semantic notions and low-level visual content, it is extremely difficult to define a distance function that accurately captures the similarity of images as perceived by humans. Furthermore, popular dimension reduction methods suffer from either the inability to capture the nonlinear correlations among raw data or very expensive training cost. To address the problems, in this chapter we introduce a new indexing *technique called Combining Multiple Visual Features (CMVF) that integrates multiple visual features to get better query effectiveness. Our approach is able to produce low-dimensional image feature vectors that include not only low-level visual properties but also high-level semantic properties. The hybrid architecture can produce feature vectors that capture the salient properties of images yet are small enough to allow the use of existing high-dimensional indexing methods to provide efficient and effective retrieval.*

INTRODUCTION

With advances in information technology, there is an ever-growing volume of multimedia information from emerging application domains such as digital libraries, World Wide Web, and Geographical Information System (GIS) systems available online. However, effective indexing and navigation of large image databases still remains one of the main challenges for modern computer system. Currently, intelligent image retrieval systems are mostly similarity-based. The idea of indexing an image database is to extract the features (usually in the form of a vector) from each image in the database and then to transform features into multidimensional points. Thus, searching for "similarity" between objects can be treated as a search for close points in this feature space and the distance between multidimensional points is frequently used as a measurement of similarity between the two corresponding image objects.

To efficiently support this kind of retrieval, various kinds of novel access methods such as Spatial Access Methods (SAMs) and *metric trees* have been proposed. Typical examples of SAMs include the *SS*-tree (White & Jain, 1996), R^+-tree (Sellis, 1987) and grid files (Faloutsos, 1994); for *metric trees*, examples include the *vp*-tree (Chiueh, 1994), *mvp*-tree (Bozkaya & Ozsoyoglu, 1997), *GNAT* (Brin, 1995) and *M*-tree (Ciaccia, 1997). While these methods are effective in some specialized image database applications, many open problems in image indexing still remain.

Firstly, typical image feature vectors are high dimensional (e.g., some image feature vectors can have up to 100 dimensions). Since the existing access methods have an exponential time and space complexity as the number of dimensions increases, for indexing high-dimensional vectors, they are no better than sequential scanning of the database. This is the well-known "dimensional curse" problem. For instance, methods based on *R*-trees can be efficient if the fan-out of the *R*-tree nodes remain greater than two and the number of dimensions is under five. The search time with linear quad trees is proportional to the size of the hyper surface of the query region that grows with the number of dimensions. With grid files, the search time depends on the directory whose size also grows with the number of dimensions.

Secondly, there is a large semantic gap existing between low-level media representation and high-level concepts such as person, building, sky, landscape, and so forth. In fact, while the extraction of visual content from digital images has a long history, it has so far proved extremely difficult to determine how to use such features to effectively represent high-level semantics. This is because similarity in low-level visual feature may not correspond to high-level semantic similarity. Moreover, human beings perceive and identify images by integrating different kinds of visual features in a "nonlinear" way. This implies that assuming each type of visual feature contributes equally to the recognition