Inherent Fusion: Towards Scalable Multi-Modal Similarity Search

Petra Budikova, Masaryk University, Brno, Czech Republic
Michal Batko, Masaryk University, Brno, Czech Republic
David Novak, Masaryk University, Brno, Czech Republic
Pavel Zezula, Masaryk University, Brno, Czech Republic

ABSTRACT

The rapid growth of unstructured data, commonly denoted as the Big Data challenge, requires new technologies that are capable of dealing with complex data objects such as multimedia. In this work, the authors focus on the content-based retrieval approach, which is able to organize such data by exploiting the similarity of data content. In particular, they focus on solutions that are able to combine multiple similarity measures during the query evaluation. The authors introduce a classification of existing approaches and analyze their performance in terms of effectiveness, efficiency, and scalability. Further, they present a novel technique of inherent fusion that combines the efficiency of fast indexed retrieval with the effectiveness of ranking methods. The performance of all discussed methods is evaluated by extensive experiments with user participation.

KEYWORDS

Content-Based Retrieval, Evaluation, Image Retrieval, Late Fusion, Multi-Modal Search, Scalability, Similarity Searching

INTRODUCTION

Due to the rapid growth of both the amount and the diversity of digital data, many organizations are nowadays facing the Big Data problem, i.e. the situation when they have potential access to a wealth of information, but they do not know how to get value out of it (Zikopoulos & Eaton, 2011). This inspired research and new solutions on different levels of data processing, including data modeling, storing, and analysis (Batini et al., 2015; Maté et al., 2015; Meged & Gelbard, 2012). In this paper, we focus on the storage and retrieval of unstructured data that cannot be straightforwardly organized in relational databases, since they are not searched by exact match but rather by similarity. This is e.g. the case of images, where pixel-to-pixel matching does not make sense but searching by visual similarity is desired in many situations, e.g. medical image analysis, entertainment, security, and surveillance. For such applications, the content-based retrieval techniques were developed (Alemu et al., 2009; Datta et al., 2008).

The fundamental idea of content-based data management is to organize complex data objects such as multimedia using their content instead of descriptive metadata that are used in traditional data management systems. As illustrated in Figure 1, a content-based image search system can retrieve images that are visually similar to a given example. A principal advantage of the content-based paradigm is the fact that the multimedia object content is always available, whereas the metadata are often sparse, erroneous, or not available at all. Depending on the type of the data to be processed,
different salient features can be extracted from the complex objects and used for indexing and retrieval of the original data. In case of images, we can use e.g. global features such as MPEG7 color, shape, or texture, local image features describing individual points of interest, face descriptors, etc. The relevance of individual data items with respect to a given query is then determined by the similarity of the extracted features, which is computed by a suitable distance function (Zezula et al., 2006). In this paper, we shall call each salient feature and the associated distance function a modality of the content-based similarity search.

In the first content-based multimedia retrieval systems, a single modality was utilized to organize and search the data. However, this proved to be insufficient for several reasons: 1) each modality only reflects a specific perspective of the complex object, which may not agree with the actual users’ subjective view (this is often denoted as the semantic gap problem); 2) a particular modality may not be applicable in some situations; 3) in large-scale applications, a single modality is typically not distinctive enough to distinguish relevant objects from irrelevant ones. Therefore, latest data management techniques focus on a multi-modal retrieval that combines multiple orthogonal views on objects (Datta et al., 2008; Jain & Sinha, 2010).

Following these observations, a number of multi-modal retrieval systems have been proposed in the past decade. In this paper, we are mainly interested in image retrieval, in particular a general-purpose image retrieval that could be used e.g. in a web search engine. This task appears in many real-world applications and therefore has attracted many researchers from different communities. As a result, diverse multi-modal image search techniques can be found in the literature. However, to achieve real improvements and mature solutions, it is also necessary to have a cooperation and comparison between individual approaches. Unfortunately, this is rather scarce in this area due to the lack of commonly accepted benchmarking platforms (Lew et al., 2006). The research groups tend to work with their own special datasets, application settings, etc., making the presented results virtually incomparable.

Figure 1. Content-based image retrieval: similar images to the query (left) were selected from a 20M image collection
The Information Quality of Databases

Impact of Flow and Brand Equity in 3D Virtual Worlds