Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

Digital libraries: A testbed for multimedia technology

ZHUANG Yue-ting (庄越挺)

(*School of Computer Science, Zhejiang University, Hangzhou 310027, China*)

E-mail: yzhuang@cs.zju.edu.cn

Received Sept. 20, 2005; revision accepted Oct. 8, 2005

Abstract: A distinguishing feature of a digital library is that it has Terabyte volumes of multimedia resources. One challenge for researchers in the field of multimedia is to find a testbed for showing the potentials of multimedia technologies such as video summarization, semantic annotation, multimedia cross indexing and retrieval, and etc. Deeper research and wider applications of digital libraries revealed their indispensable role as testbed for multimedia technologies. This paper presents challenging issues of some key techniques used in digital libraries and their specific needs for multimedia technologies.

Key words: Digital library, Multimedia technology, CADAL (China-America Digital Academic Library)
doi:10.1631/jzus.2005.A1201 **Document code:** A **CLC number:** TP391

INTRODUCTION

One of the greatest attractions of a digital library is its multimedia resources humans can use for information retrieval to knowledge discovery and information integration. A problem of multimedia researchers is to find a so-called killer-Apps which can demonstrate well the potentials of research topics such as video summarization, semantic annotation, multimedia cross indexing and retrieval, etc. Deeper research and wider applications of digital libraries revealed their indispensable role as testbed for multimedia technologies.

This paper presents challenging issues of some key techniques used in digital libraries and their specific needs for multimedia technologies. CADAL (China-America Digital Academic Library) project is a collaborative project between universities and institutes in China and the USA. In CADAL, Terabyte volumes of multimedia data with various modality (such as text, image, video and animation) is available, which presents a challenge for the application of multimedia technologies. We have initiated several projects related to multimedia technology application in digital library, such as application in video summarization, multimedia semantic annotation, multimedia cross indexing and retrieval, Chinese calligraphy retrieval and 3-D visualization.

VIDEO MINING AND SUMMARIZATION

Video summarization is a compact representation of video content. How to concisely and informatively summarize video content is extremely important for large-scale multimedia browser and retrieval. For instance, it provides an effective way for rapid overview of the video database, fast access to users' video shots and episodes of interest browsing and retrieval of video content.

Usually users would browse video content according to different views such as characters, caption and frequent video patterns. For example, users would like to look at all of contents related to certain characters or similar frequent video shots. We propose a novel method for automatically generating multi-view video summarization.

Three kinds of content views are currently presented: video semantic-face, video caption and frequent shot patterns. First, since human faces in video and images bear significant semantic contents, we can use faces to index and summarize video contents, especially for video news. Although human faces are high-level semantic features, not all faces in video news are important and semantic. Only some important faces like semantic-face in video news carry important semantic contents. Semantic-face includes anchorperson face, interviewed person face and main news character face. Here we will present a new algorithm for semantic clustering and indexing of video news based on semantic-face.

Second, Video caption could also be used to generate video summarization with high-level semantics since it implied lots of semantics inherently. For example, captions in news broadcasts and documentaries usually annotate information of the reported events. There are two kinds of text in videos: video caption and scene text. Scene text is part of the environment and captured by the camera along with the rest of the scene. Because video caption contains more semantic information than scene text, the objective of our system is to locate the caption text.

Finally, we also present an approach of Frequent Shot Patterns (FSP) to automatically summarize video content. The frequent shot patterns in video are defined as a sequence of video shots that occur frequently in the time window. In this approach, video shots are grouped together through support vector clustering based on the shot boundary detection. Then video frequent shot patterns are extracted in order to remove the visual-content redundancy among video content clusters. At last, video summarization is generated by those frequent shot patterns with time interval constraint to cover the video scenario information.

MULTIMEDIA SEMANTIC ANNOTATION

Multimedia resources are required to be managed, retrieved and integrated through high-level semantics, which is traditionally obtained from manual labelling. Well-annotated image collections include Corel image galleries (http://elib.cs.berkeley.edu/photos/corel/), most museum image collections (http://www.thinker.org), the Web archive (http://www.archive.org), etc. However, this approach is liable to be subjective, and requires a huge amount of human effort. As new multimedia resources increase dramatically everyday, an automatic annotation method becomes necessary and important.

We propose an approach to automatically annotate image content. First, Support Vector Machine (SVM) is used to classify images automatically; statistical learning method is then adopted to select the most appropriate keywords for an incoming image on the basis of the annotated image collections.

MULTIMEDIA CROSS INDEXING AND RETRIEVAL

Efficient multimedia retrieval is necessary for digital library. Most multimedia retrieval approaches available at present are dedicated to a certain media type. We developed a multimedia cross indexing and retrieval system for digital library which is distinguished from most current multimedia retrieval systems in that it integrates multi-modal data into a seamless system.

Differing from traditional query-by-example systems, the modality of query examples and that of query results need not be the same in cross indexing and retrieval systems. For example, users can query images by submitting an audio example in cross-media retrieval systems, and vice versa. Clearly, cross-media retrieval systems are especially powerful and should be paid more research attention.

The cross-reference search engine, based on cross-reference graph, serves as a bridge in the system. This system can not only calculate semantic level similarity between media object and the query but trigger single modal search engine performing content based retrieval based on support vector clustering. It can adjust the cross-reference graph model based on relevance feedback conducted by users, which progressively improves

the retrieval accuracy. As demonstrated by the experiments, the system can achieve satisfactory retrieval accuracy and efficiency on multi-modal datasets.

CHINESE CALLIGRAPHY CHARACTER RETRIEVAL AND 3-D VISUALIZATION

Original historical paper works of calligraphy comprise valuable legacy of civilization to mankind. They are fragile, and cannot be turned over and over again by many different people. A well-known protection method is restricting the access only to a few researchers. In order to widely share them with the general public in the CADAL project, many famous paper works are digitalized and published on the Portal of CADAL. The key challenge is how to manage large digitized calligraphy images to offer fast browsing and retrieval services and 3-D visualization of the writing process of the retrieved Chinese calligraphy character. We developed a Chinese calligraphy system, which provides shape-based Chinese calligraphy character retrieval and the 3-D visualization of the searched works and characters.

Shape-based Chinese calligraphy character retrieval: The digitized calligraphy requires extremely expensive manual annotation to retrieve them, because there is no existing technique to convert images of Chinese calligraphy character into computer font like OCR does. The key problem in calligraphic character retrieval is the matching of similar isolated characters.

In terms of image comparisons, there are two categories of methods: intensity-based (color and texture) and geometry-based (shape). For calligraphic character retrieval, shape is far more interesting than color or texture, though retrieval by shape is still a difficult aspect of content-base research. So the key issue is to find characters of similar shape based on user's query sample. There are 3 ways for a user to submit a query sample: importing an existing character image, sketching a character, and typing in a character from the keyboard. We propose a novel approach for efficient retrieval of Chinese calligraphy character. First, segment calligraphy pages into individual calligraphy characters using minimum-bounding box. Second, the individual calligraphy characters are normalized and shape feature is described by approximate points context. Third, character complexity, stroke density and stroke protrusion are computed to build hierarchical index. Fourth, compute the total matching cost of each candidate calligraphic character in the database with the query character. Lastly, the results are ranked according to the matching cost and the similar calligraphic characters are retrieved.

3-D visualization of the searched works and characters: In order to help people enjoy the art of calligraphy writing and find how it was written step-by-step, the writing process of a specified calligraphy character is animated by 3-D visualization method. Firstly, extract the strokes order of an offline Chinese calligraphic character. Then estimate the varied stroke's thickness. Finally, animate the writing process. Users can view the animation of calligraphic writing process by the browser.

(This is the abstract version for the invited talk in ICUDL'05)