Workshop Report: The First ACM International Workshop on Multimedia Databases (MMDB 2003) at ACM CIKM 2003

Shu-Ching Chen
Distributed Multimedia Information System Laboratory
School of Computer Science
Florida International University
Miami, FL 33199, USA
chens@cs.fiu.edu

Mei-Ling Shyu
Department of Electrical & Computer Engineering
University of Miami
Coral Gables, FL 33124, USA
shyu@miami.edu

The motivation for organizing a workshop on "Multimedia Databases" at ACM International Conference on Information and Knowledge Management (ACM CIKM 2003) was the fact that more and more multimedia data are now available and effective management of these multimedia data is becoming a critical need. As a consequence, multimedia databases have emerged as an important research area. A multimedia database management system must provide the support for managing text, video, audio and image data, and it must also manage multimedia data types. In addition, the relationships between the objects and the representation of the spatial and/or temporal relationships may involve the extensions to the database management system.

The First ACM International Workshop on Multimedia Databases was held in New Orleans, Louisiana, USA, November 7, 2003. Its aim is to bring together university researchers, scientists, industry professionals, software engineers and graduate students who need to become acquainted with new theories and technologies in multimedia databases, and to all those who wish to gain a detailed technical understanding of what multimedia databases involve.

We received thirty-three submissions. Each of them was reviewed by at least three referees. The program committee, which was constituted of researchers and practitioners from the multimedia database research areas, selected twelve papers to be presented at the workshop. The authors were asked to address each and every comment made by the referees to improve the quality of the papers. These papers reflect current trends in the multimedia database research areas with topics such as Video Analysis, Video Retrieval, Video Summarizing, Multimedia Communications, Relevance Feedback, Multimedia Indexing, and Content-Based Image Retrieval for Multimedia Databases. From these twelve papers, the workshop offered three sessions for participants representing a variety of backgrounds and perspectives on multimedia databases.

M. Fayzullin presented a paper on behalf of V.S. Subrahmanian, A. Picariello, and M.L. Sapino on "The CPR Model for Summarizing Video." In this work, a model of video summarization based on three important parameters: Priority (of frames), Continuity (of the summary), and non-Repetition (of the summary) was described. Due to the fact that a summary must include high priority frames, be continuous, and be non-repetitive, an optimal summary is one that maximizes an objective function based on these three parameters. They developed formal definitions of all

these concepts and provide algorithms to find optimal summaries. The performance of these algorithms was also presented.

In "Automatic Generation of MPEG-7 Compliant XML Document for Motion Trajectory Descriptor in Sports Video," Yi Haoran, Deepu Rajan, and Chia-Liang Tien discussed the use of one of the visual descriptors related to motion (i.e., motion trajectory) in a video sequence and proposed a method of automatically extracting trajectories from video sequences and generating an XML document that conforms to the MPEG-7 schema. In particular, sports videos were applied to demonstrate the robustness of their algorithm because the trajectories are very random. They described the displacements of objects in time, where an object is defined as a spatiotemporal region or set of spatiotemporal regions. The MPEG-7 standard is a step towards standardizing the description of multimedia content so that quick and efficient identification of relevant content can be facilitated together with efficient management of information. The description definition language (DDL) is a schema language to represent valid MPEG-7 descriptors and description schemes. MPEG-7 instances are XML documents that conform to a particular MPEG-7 schema, as expressed in the DDL and that describe audiovisual content. Moreover, the MPEG-7 XM software is a working tool that enables the checking of the relative performance of various algorithms. Insofar as the motion trajectory is concerned, the current version of XM takes motion blobs that are already segmented as inputs and outputs the trajectory as an XML document. However, they went a step further by developing an automatic trajectory extraction algorithm that produces motion blobs, which are then tracked to generate the trajectory, which in turn is output as an XML document.

Wen Wen Hsieh and Arbee L.P. Chen addressed the design and implementation issues of a bowling information system on their paper entitled "Constructing a Bowling Information System with Video Content Analysis." This bowling information system contains three types of bowling game information including the bowling video content information, the game-related information, and the player information. The MPEG-7 Description Schemes were used to describe these types of information and the relationships among them that were obtained through an annotator by which the manual conceptual feature annotation (for the player and game-related information) and automatic perceptual feature extraction (for the video content information) were integrated. Several interesting events in the video such as strikes and the important frames were determined by automatically analyzing the video content. With an interactive user interface, a user's query was transformed into XQuery to retrieve the required information about the bowling games in order to learn the skills of bowling. In addition to the implementation of the system, experimental results demonstrating the effectiveness of the automatic video content information extraction were also reported.

The increased use of video data sets for multimedia-based applications has created a demand for strong video database support, including efficient methods for handling the content-based query and retrieval of video data. Video query processing presents significant research challenges, mainly associated with the size, complexity and unstructured nature of video data. A video query processor must support video operations for search by content and streaming, new query types,

and the incorporation of video methods and operators in generating, optimizing and executing query plans. Walid Aref, Ann Christine Catlin, Ahmed K. Elmagarmid, Moustafa Hammad, Ihab Ilyas, Mirette Marzouk and Thanaa Ghanem addressed these query processing issues in "Video Query Processing in the VDBMS Testbed for Video Database Research" under two contexts: (1) as applied to the video data type and (2) as applied to the stream data type. In this paper, the query processing functionality of the VDBMS video database management system (designed to support a full range of functionality for video as an abstract database data type), and two query operators for the video data type to implement the rank-join and stop-after algorithms were described. As videos may be considered streams of consecutive image frames, video query processing can be expressed as continuous queries over video data streams. The stream data type was therefore introduced into the VDBMS system, and system functionality was extended to support general data streams. From this viewpoint, they presented an approach for defining and processing streams, including video, through the query execution engine. In addition, the implementation of several algorithms for video query processing (expressed as continuous queries over video streams) such as fast forward, region-based blurring and left outer join was given. A description of the window-join algorithm as a core operator for continuous query systems and the shared execution as an optimization approach for stream query processing were also presented.

In "Facilitate Knowledge Communications in Multimedia e-Learning Environments," Weihong Huang, Ting Tao, Mohand-Said Hacid, and Alain Mille proposed a contextual knowledge management framework to improve the current isolated learning information management status, by making the most use of related learning information, and enabling flexible knowledge communication and transfer between instructors and learners. The motivation of their work is because the current multimedia database and information retrieval technologies mainly focus on low-level single type media servicing, which cannot fulfill the increasing requirements of semantic-based knowledge retrieval and communication services in highly interactive multimedia e-Learning environments. New multimedia computing and knowledge management technologies are needed to facilitate the knowledge communications in e-Learning environments. This paper explored a set of novel knowledge management techniques from a contextual point of view in order to work with related existing e-Learning content description standards and multimedia information retrieval techniques and consequently enable a semantic-based interactive learning environment. Based on knowledge communication model analysis in university e-learning environments, a contextual multimedia knowledge representation model was presented with the discussion of the corresponding knowledge retrieval techniques, and a case study in a virtual scenario-based learning environment to show how their presented framework supports intelligent multimedia e-Learning.

Recently, various schemes for broadcasting continuous media data such as music or movies have been studied. These schemes reduce the waiting time for playing the data under the continuity condition, i.e., playing continuous media data without any intermittence until the end of the data. These schemes usually employ multiple channels to broadcast the data. However, most clients are not able to receive data from multiple channels concurrently. In "Scheduling Methods for

Broadcasting Multiple Continuous Media Data," Tomoki Yoshihisa, Masahiko Tsukamoto, and Shojiro Nishio proposed methods for reducing client waiting time for multiple data via a single channel. In our proposed methods, we divide each data into several segments and produce a schedule that includes precedent segments more frequently. By changing the number of segments according to the playback ratio, client waiting time is effectively reduced.

Douglas R. Heisterkamp and Jing Peng introduced a novel KVA-File (kernel VA-File) that extends VA-File to kernel-based retrieval methods in "Kernel VA-Files for Relevance Feedback Retrieval" to address the problems that many data partitioning index methods perform poorly in high dimensional space and do not support relevance feedback retrieval. Though the vector approximation file (VA-File) approach could overcome some of the difficulties of high dimensional vector spaces, but cannot be applied to relevance feedback retrieval using kernel distances in the data measurement space. A key observation is that kernel distances may be non-linear in the data measurement space but is still linear in an induced feature space. It is this linear invariance in the induced feature space that enables KVA-File to work with kernel distances. An efficient approach to approximating vectors in an induced feature space is presented with the corresponding upper and lower distance bounds. In their work, an effective indexing method was provided for kernel-based relevance feedback image retrieval methods, and their experimental results validated the efficacy using large image data sets with approximately 100,000 images with 463 dimensions of measurement.

Guang-Ho Cha's paper entitled "Bitmap Indexing Method for Complex Similarity Queries with Relevance Feedback" proposed a novel indexing method that is designed to efficiently handle complex similarity queries as well as relevance feedback in high-dimensional image and video databases. Similarity indexing and searching are known to be challenging for high-dimensional applications such as multimedia databases. Especially, they become more difficult when multiple features have to be indexed together. Moreover, few indexing methods are currently available to effectively support disjunctive queries for relevance feedback. In this work, every dimension was treated independently in order to provide the indexing method with the flexibility in controlling multiple features and multiple query objects. The efficiency of this method was realized by a specialized bitmap indexing that represents all objects in a database as a set of bitmaps. The percentage of data accessed in this method was inversely proportional to the overall dimensionality, and thus the performance deterioration with the increasing dimensionality does not occur. There are three main contributions in this work. First, a novel way to index high-dimensional data was provided. Second, complex similarity queries were efficiently handled. Lastly, disjunctive queries driven by relevance feedback were efficiently treated. Speedups of 10 to 15 over the linear scan were also reported from the empirical results.

Mario A. Nascimento and Jie Luo presented an approach based on a hierarchical tree that encodes the color feature of image regions and is stored in the format of an index sequence in "Content Based Sub-image Retrieval via Hierarchical Tree Matching." Their approach deals with the problem of finding images that contain a given query image, the so-called content-based sub-image retrieval. The image regions are obtained by moving windows of different sizes over a

fixed grid partition, so as to maximize the probability that a cell covers an image's regions of interest. The resulting approach is simple yet quite effective and efficient. In their experiments using 10,000 images and disk-resident meta-data, 60% (85%) of the time the relevant image (the one where the query sub-image was extracted from) was found among the first 20 (50) retrieved images in about a tenth of a second.

In "Robust Content-Based Image Searches for Copyright Protection," Sid-Ahmed Berrani, Laurent Amsaleg, and Patrick Gros described a novel content-based image retrieval scheme for copyright protection. Its goal is to detect matches between a set of doubtful images and the ones stored in the database of photographies legal holders. If an image was stolen and used to create a pirated copy, it tries to identify from which original image the copy was created. Thanks to local differential descriptors, the matching process takes into account a large set of variations that might have been applied to pirated images. The high cost and the complexity of this image recognition scheme requires a very efficient retrieval process since many individual queries must be performed before being able to construct the final result. Hence, their work proposed to use a novel search method that trades the precision of each individual search for reduced query execution time. This imprecision has only little impact on the overall recognition performance since the final result is a consolidation of many partial results. However, it dramatically accelerates queries. This result was corroborated by a theoretically study, and the experimental results showed the efficiency and the robustness of their method.

Mei-Ling Shyu, Shu-Ching Chen, Min Chen, Chengcui Zhang, and Kanoksri Sarinnapakorn discussed an effective image database retrieval framework using a new mechanism called the Markov Model Mediator (MMM) in their paper entitled "Image Database Retrieval Utilizing Affinity Relationships." This framework focused on bridging the gap between the low-level features and high-level semantic contents of images since this gap has become the bottleneck of Content-Based Image Retrieval (CBIR). To meet this demand, their MMM mechanism takes into consideration not only the low-level image features, but also the high-level concepts learned from the history of user's access pattern and access frequencies on the images in the database, which is efficient in two aspects. Firstly, the overhead for real-time training is avoided in the image retrieval process because the high-level concepts of images are captured in the off-line training process, and secondly, before the exact similarity matching process, Principal Component Analysis (PCA) is applied to reduce the image search space. A training subsystem for this framework was implemented and their experimental results demonstrated that the MMM mechanism can effectively assist in retrieving more accurate results from image databases.

Finally, James C. French on behalf of Xiangyu Jin presented a novel technology which can return semantically related images in different visual clusters by merging the result sets of multiple queries in their paper "Improving Image Retrieval Effectiveness via Multiple Queries." The conventional approaches to image retrieval are based on the assumption that relevant images are physically near the query image in some feature space, which is the basis of the cluster hypothesis. However, semantically related images are often scattered across several visual clusters. Although traditional Content-based Image Retrieval (CBIR) technologies may utilize the

information contained in multiple queries (gotten in one step or through a feedback process), this is only a reformulation of the original query. As a result these strategies only get the images in some neighborhood of the original query as the retrieval result. This severely restricts the system performance. To overcome such a problem, their proposed technology merged the result sets of multiple queries, which was furthermore validated via several experiments.

In fact, a particular strength of the workshop was the result that it brought an international community of researchers from academic institutions, opening up a valuable set of interdisciplinary conversations. All participants were given the workshop proceedings with the full papers. In addition to these spectacular sessions, the workshop also held one keynote speech by Prof. Shih-Fu Chang at Columbia University. The title of this keynote speech was "Mining Spatio-Temporal Patterns and Knowledge Structures in Multimedia Collection." This talk addressed the issues of the detection and recognition of semantic events for multimedia indexing, and how mining of patterns of different types at different levels offers rich benefits including the automatic discovery of salient events in a new domain, automatic alert generation from massive real-time data, and discovery of novel event relationships.

The long-term goal of MMDB is to build up a reputable and respectable workshop for the international community. Without the contribution and support from the international community, MMDB could never be classified as a significant and successful workshop as what the MMDB 2003 has achieved! Furthermore, given the high quality of the papers, several selected papers are invited for the special issue on "Multimedia Databases" in the Multimedia Tools and Applications Journal (Kluwer).

Many people contributed to the success of this event. We would like to thank all the participants, paper presenters, and session chairs for their valuable cooperation with the demands placed on them. We wish to thank the program committee members and the external referees for their great work. We are grateful to Professor Shih-Fu Chang at Columbia University for accepting to give a keynote presentation. Finally, we would like to express our gratitude towards the CIKM organizers.

Once again, we are looking forward to meeting you in ACM MMDB 2004 (*http://www.cs.fiu.edu /mmdb04*).

Shu-Ching Chen and Mei-Ling Shyu
ACM MMDB 2003 Program co-chairs