

# Preface

## 1 Workshop Theme

Digital multimedia differs from previous forms of combined media in that the bits that represent text, images, animations, and audio, video and other signals can be treated as data by computer programs. One facet of this diverse data in terms of underlying models and formats is that it is synchronized and integrated, hence it can be treated as integral data records. Such records can be found in a number of areas of human endeavour.

Modern medicine generates huge amounts of such digital data. Another example is architectural design and the related architecture, engineering and construction (AEC) industry. Virtual communities (in the broad sense of this word, which includes any communities mediated by digital technologies) are another example where generated data constitutes an integral data record. Such data may include data about member profiles, the content generated by the virtual community, and communication data in different formats, including e-mail, chat records, SMS messages, videoconferencing records.

Not all multimedia data is so diverse. An example of less diverse data, but data that is larger in terms of the collected amount, is that generated by video surveillance systems, where each integral data record roughly consists of a set of time-stamped images – the video frames. In any case, the collection of such integral data records constitutes a multimedia data set. The challenge of extracting meaningful patterns from such data sets has led to the research and development in the area of multimedia data mining. This is a challenging field due to the nonstructured nature of multimedia data. Such ubiquitous data is required, if not essential, in many applications. Multimedia databases are widespread and multimedia data sets are extremely large. There are tools for managing and searching within such collections, but the need for tools to extract hidden useful knowledge embedded within multimedia data is becoming critical for many decision-making applications. The tools needed today are tools for discovering relationships between data items or segments within images, classifying images based on their content, extracting patterns from sound, categorizing speech and music, recognizing and tracking objects in video streams, relations between different multimedia components, and crossmedia object relations.

## 2 Multimedia Workshop Series

This book is a result of two workshops: Multimedia Data Mining (MDM/KDD 2002) held in conjunction with ACM SIGKDD 2002 in Edmonton, Canada in July 2002, and Knowledge Discovery from Multimedia and Complex Data (KDMCD 2002) held in conjunction with PAKDD 2002 in Taipei, Taiwan in May 2002. These two workshops brought together crossdisciplinary experts for

the analysis of digital multimedia content, multimedia databases, spatial data analysis, analysis of data in collaborative virtual environments, and knowledge engineers and domain experts from different applied disciplines related to multimedia data mining. The book reveals a variety of topics that come under the umbrella of multimedia data mining and mining complex data: mining spatial multimedia data; mining audio data and multimedia support; mining image and video data; frameworks for multimedia mining; multimedia mining for information retrieval; and applications of multimedia mining. These workshops were a continuation of other successful multimedia workshops held in conjunction with the KDD conference in 2000 and 2001. The multimedia workshop series events have been attended by industry as well as academia, and attendees have shown a sustained interest in this area.

### 3 Papers

The book attempts to address the above-mentioned issues, looking at specific issues in pattern extraction from image data, sound, and video; suitable multimedia representations and formats that can assist multimedia data mining; and advanced architectures of multimedia data mining systems. The papers in the book are not presented in a specific order. This is perhaps to reflect the fact that multimedia mining is at the confluence of many disciplines, often tackling different subjects and problems from different angles at the same time. The papers provide an interesting coverage of different issues and some technical solutions.

In “Subjective Interpretation of Complex Data: Requirements for Supporting Kansei Mining Process” Bianchi-Berthouze and Hayashi present the continuation of Bianchi-Berthouze’s work on modeling visual impressions from the point of view of multimedia data mining. The new work describes a data warehouse for the mining of multimedia information, where a unique characteristic of the data warehouse is its ability to store multiple hierarchical descriptions of the multimedia data. Such a characteristic is necessary to allow mining not only at different levels of abstraction but also according to multiple interpretations of the content. The proposed framework could be generalized to support the analysis of any type of complex data that relate to subjective cognitive processes, whose content interpretation would be greatly variable.

In “Multimedia Data Mining Framework for Raw Video Sequences,” Oh and Bandi present a general framework for real-time video data mining from “raw videos” (e.g., traffic videos, surveillance videos). The focus within the presented framework is on the motion as a feature, and how to compute and represent it for further processing. The multilevel hierarchical segment clustering procedure used category and motion.

In the paper “Object Detection for Hierarchical Image Classification” by Khan and Wang, the authors discuss the indexing of images according to meanings rather than objects that appear in images. The authors propose a solution to the problem of creating a meaning-based index structure through the design of a concept-based model using domain-dependent ontologies. Aiming at an accurate

identification of object boundaries, the authors propose an automatic scalable object boundary detection algorithm based on edge detection and region growing techniques, and an efficient merging algorithm to join adjacent regions using an adjacency graph to avoid the over-segmentation of regions. They implemented a very basic system aimed at the classification of images in the sports domain, in order to illustrate the effectiveness of the algorithm.

In their paper “Mining High-level User Concepts with Multiple Instance Learning and Relevance Feedback for Content-Based Image retrieval,” Huang, Chen, Shyu and Zhang propose a framework that incorporates multiple instance learning into the user relevance feedback to discover users’ concept patterns. These patterns reflect where the user’s most interested region is allocated and how to map the local feature vector of that region to the high-level concept pattern of the user. This underlying mapping could be progressively discovered through the proposed feedback and learning procedure. The role of the user in the retrieval system is in guiding the mining process according to the user’s focus of attention.

In “Associative Classifiers for Medical Images,” Antonie, Zaiiane and Coman present a new approach to learning a classification model. Their approach is based on discovering association rules from a training set of images where the rules have the constraint to always include a class label as a consequent. Their association rule-based classifier was tested on a real dataset of medical images. The approach, which included a significant (and important) preprocessing phase, showed promising results when classifying real mammograms for breast cancer detection.

In “An Innovative Concept for Image Information Mining,” Datcu and Seidel introduce the concept of “image information mining” and discuss a system that implements this concept. The approach is based on modeling the causalities that link the image-signal contents to the objects and structures of interest to the users. Their approach consists first of extracting image features using a library of algorithms; then unsupervised grouping in a large number of clusters, data reduction by parametric modeling of the clusters, and supervised learning of user semantics, the level where, instead of being programmed, the system is trained using a set of examples.

The paper “Multimedia Data Mining Using P-Trees,” by Perrizo, Jockcheck, Perera, Ren, Wu and Zhang, focuses on a data structure that provides an efficient, lossless, data-mining-ready representation of the data. The Peano Count Tree (P-tree) provides an efficient way to store and mine sets of images and related data. The authors show the effectiveness of such a structure in the context of multimedia mining.

In “Scale Space Exploration for Mining Image Information Content,” Ciucu, Heas, Datcu and Tilton describe an application of a scale-space-clustering algorithm for the exploration of image information content. The clustering considers the feature space as a thermodynamical ensemble and groups the data by minimizing the free energy, with the temperature as a scale parameter. They analyze the information extracted by the grouping and propose an information represen-

tation structure that enables exploration of the image content. This structure is a tree in the scale space showing how the clusters merge.

In “Videoview: A Content-Based Video Description Scheme and Database Navigation,” Guler and Pushee introduce a unified framework for a comprehensive video description scheme and present a browsing and manipulation tool for video data mining. The proposed description scheme is based on the structure and the semantics of the video, incorporating scene, camera, and object and behaviour information pertaining to a large class of video data. The navigator provided a means for visual data mining of multimedia data: an intuitive presentation, interactive manipulation, the ability to visualize the information and data from a number of perspectives, and the ability to annotate and correlate the data in the video database.

In “The Community of Multimedia Agents,” Wei, Petrushin and Gershman present a work devoted to creating an open environment for developing, testing, learning and prototyping multimedia content analysis and annotation methods. Each method is represented as an agent that could communicate with the other agents registered in the environment using templates based on the descriptors and description schemes in the emerging MPEG-7 standard. This environment enables researchers to compare the performance of different agents and combine them in more powerful and robust system prototypes.

“Multimedia Mining of Collaborative Virtual Workspaces: An Integrative Framework for Extracting and Integrating Collaborative Process Knowledge” by Simoff and Biuk-Aghai is an extended paper from a presentation at the MDM/KDD 2001 workshop. The authors focus on the knowledge discovery phase that comes after the data mining itself, namely the integration of discovered knowledge and knowledge transfer. They present a framework for this integration and transfer, and show its use in a particular domain: collaborative virtual workspaces.

In “STIFF: A Forecasting Framework for Spatio-Temporal Data,” Li and Dunham cope with complex data containing both spatial and temporal characteristics. They propose a framework that uses neural networks to discover hidden spatial correlations and a stochastic time series to capture temporal information per location. This model is used for predictions. They show how their model can be used to predict water flow rate fluctuations in rivers.

In “Mining Propositional Knowledge Bases to Discover Multi-level Rules,” Richards and Malik describe a technique for recognizing knowledge and discovering higher-level concepts in the knowledge base. This technique allows the exploration of the knowledge at and across any of the levels of abstraction to provide a much richer picture of the knowledge and understanding of the domain.

In “Meta-classification: Combining Multimodal Classifiers,” Lin and Hauptmann present a combination framework called “meta-classification,” which models the problem of combining classifiers as a classification problem itself. They apply the technique on a wearable “experience collection” system, which unobtrusively records the wearer’s conversation, recognizes the face of the dialogue partner, and remembers his/her voice. When the system sees the same person’s

face or hears the same voice, it can then use a summary of the last conversation to remind the wearer. To identify a person correctly from a mixture of audio and video streams, classification judgments from multiple modalities must be effectively combined.

In “Partition Cardinality Estimation in Image Repositories,” Fernandez and Djeraba deal with the problem of automatically identifying the number of clusters to be discovered when considering clustering in large repositories of artefacts, such as images. They present an approach that estimates automatically the best partition cardinality (the best number of clusters) in the context of content-based accessing in image repositories. They suggest a method that reduces drastically the number of iterations necessary to extract the best number of clusters.

In “A Framework for Customizable Sports Video Management and Retrieval,” Tjondronegoro, Chen and Pham propose a framework for a customizable video management system that allows the system to detect the type of video to be indexed. The system manages user preferences and usage history to make the system support specific requirements. The authors show how the extracted key segments can be summarized using standard descriptions of MPEG-7 in a hierarchical scheme.

In “Style Recognition Using Keyword Analysis,” Lorensuhewa, Pham and Geva are interested in supervised classification where the learning sample could be insufficient. They present a framework for the augmentation of expert knowledge using knowledge extracted from multimedia sources such as text and images, and they show how this framework can be applied effectively.

## 4 Conclusion

The book discussion revised the scope of multimedia data mining outlined during the previous workshops of the MDM/KDD series, clearly identifying the need to approach multimedia data as a “single unit” rather than ignoring some layers in favor of others. The authors acknowledged the high potential of multimedia data mining methods in medical domains, design and creative industries. There was an agreement that the research and development in multimedia mining should be extended in the area of collaborative virtual environments, 3D virtual reality systems, the music domain and e-business technologies. The papers show that many researchers and developers in the areas of multimedia information systems and digital media turn to data mining methods for techniques that can improve indexing and retrieval in digital media. There is a consensus that multimedia data mining is emerging as its own distinct area of research and development. The work in the area is expected to focus on algorithms and methods for mining from images, sound and video streams. The paper authors identified that there is a need for: (i) development and application of specific methods, techniques and tools for multimedia data mining; and (ii) frameworks that provide consistent methodology for multimedia data analysis and integration of discovered knowledge back into the system where it can be utilized.

## 5 Acknowledgements

We would like to acknowledge the Program Committee members of MDM/KDD 2002 and KDMCD 2002 who invested their time in carefully reviewing papers for this volume: Frederic Andres (NII, Japan), Marie-Aude Aufaure (INRIA, France), Bruno Bachimont (INA, France), Nadia Bianchi-Berthouze (University of Aizu, Japan), Nozha Boujemaa (INRIA, France), Terry Caelli (University of Alberta, Canada), Liming Chen (ECL, France), Claude Chrisment (University of Toulouse, France), Chitra Dorai (IBM, USA), Alex Duffy (University of Strathclyde, UK), William Grosky (Wayne State University, USA), Howard J. Hamilton (University of Regina, Canada), Jiawei Han (University of Illinois at Urbana-Champaign, USA), Mohand-Saïd Hacid (Claude Bernard University, France), Alexander G. Hauptmann (Carnegie Mellon University, USA), Wynne Hsu (National University of Singapore, Singapore), Odej Kao (Technical University of Clausthal, Germany), Paul Kennedy (University of Technology, Sydney, Australia), Latifur Khan (University of Texas, USA), Inna Kolyshkina (Pricewaterhouse Coopers, Australia), Nabil Layaida (INRIA Rhône Alpe, France), Brian Lovell (University of Queensland, Australia), Mike Maybury (MITRE Corporation, USA), Gholamreza Nakhaeizadeh (DaimlerChrysler, Germany), Mario Nascimento (University of Alberta, Canada), Ole Nielsen (Australian National University, Australia), Monique Noirhomme-Fraiture (FUNDP, Belgium), Vincent Oria (New Jersey Institute of Technology, USA), Jian Pei (SUNY, Buffalo, USA), Valery A. Petrushin (Accenture, USA), Jean-Marie Pinon (INSA, France), Mohamed Quafafou (IAAI, France), Zbigniew Rass (UNC, Charlotte, USA), Simone Santini (San Diego Super Computer Center, USA), Florence Sedes (University of Toulouse, France), Pramod Singh (University of Technology, Sydney, Australia), Dong Thi Bich Thuy (University of Hô Chi Minh City, Vietnam), Duminda Wijesekera (George Mason University, USA).

We would also like to thank others who contributed to the Multimedia Workshop series, including the original PC members who reviewed the first set of workshop papers. We are grateful to the SIGKDD and PAKDD organizing committee. Finally we would like to thank the many participants who brought their ideas, research and enthusiasm to the workshops and proposed many new directions for multimedia mining research.

June 2003

Osmar R. Zaiane  
Simeon J. Simoff  
Chabane Djeraba