Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 49073, Pages 1–3
DOI 10.1155/ASP/2006/49073

Editorial
Information Mining from Multimedia Databases

Ling Guan,[1] Horace H. S. Ip,[2] Paul H. Lewis,[3] Hau San Wong,[2] and Paisarn Muneesawang[1]

[1] *Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3*
[2] *Department of computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong*
[3] *Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

Received 7 September 2005; Accepted 7 September 2005

Welcome to the special issue on "Information mining from multimedia databases." The main focus of this issue is on information mining techniques for the extraction and interpretation of semantic contents in multimedia databases. The advances in multimedia production technologies have resulted in a rapid proliferation of various forms of media data types on the Internet. Given these high volumes of multimedia data, it is thus essential to extract and interpret their underlying semantic contents from the original signal-based representations without the need for extensive user interaction, and the technique of *multimedia information mining* plays an important role in this automatic content interpretation process.

Due to the spatio-temporal nature of most multimedia data streams, an important requirement for this information mining process is the accurate extraction and characterization of salient events from the original signal-based representation, and the discovery of possible relationships between these events in the form of high-level association rules. The availability of these high-level representations will play an important role in applications such as content-based multimedia information retrieval, preservation of cultural heritage, surveillance, and automatic image/video annotation. For these problems, the main challenges are in the design and analysis of mapping techniques between the signal-level and semantic-level representations, and the adaptive characterization of the notion of saliency for multimedia events in view of its dependence on the preferences of individual users and specific contexts.

The focus of the first two papers is on the automatic analysis and interpretation of video contents. X.-P. Zhang and Chen describe a new approach to extracting objects from video sequences which is based on spatio-temporal independent component analysis and multiscale analysis. Specifically, spatio-temporal independent component analysis is first performed to identify a set of preliminary source images which contain moving objects. These data are then further processed using wavelet-based multiscale analysis to improve the accuracy of video object extraction. Liu et al. propose a new approach for performing semantic analysis and annotation of basketball video. The technique is based on the extraction and analysis of multimodal features which include visual, motion, and audio information. These features are first combined to form a low-level representation of the video sequence. Based on this representation, they then utilize domain information to detect interesting events, such as when a player performs a successful shot at the basket or when a penalty is imposed for rule violation, in the basketball video.

The topic of the next two papers is on video analysis in the compressed domain. Hesseler and Eickeler propose a set of algorithms for extracting metadata from video sequences in the MPEG-2 compressed domain. Based on the extracted motion vector field, these algorithms can infer the correct camera motion, allow motion detection within a limited region of interest for the purpose of object tracking, and perform cut detection. In the next paper, Fonseca and Nesvadba introduce a new technique for face detection and tracking in the compressed domain. In particular, face detection is performed using DCT coefficients only, and motion information is extracted based on the forward and backward motion vectors. The low computational requirement of the proposed technique facilitates its adoption on mobile platforms.

The next two papers describe new information mining techniques based on the extraction and characterization of audio features. Radhakrishnan et al. propose a content-adaptive representation framework for event discovery using audio features from "unscripted" multimedia such as sports and surveillance data. Based on the assumption that interesting events occur infrequently in a background of uninteresting events, the audio sequence is regarded as a time series,

and temporal segmentation is performed to identify subsequences which are outliers based on a statistical model of the series. In the next paper, Chu et al. introduce a hierarchical approach for modeling the statistical characteristics of audio events over a time series to achieve semantic context detection. Specifically, modeling at the two separate levels of audio events and semantic context is proposed to bridge the gap between low-level audio features and semantic concepts. Different characteristic events in action movies are modeled using hidden Markov models, and both generative and discriminative approaches are adopted at the semantic context level to perform event fusion for detection of characteristic scenes.

The next four papers investigate techniques for bridging the semantic gap between low-level representation and high-level interpretation in different types of multimedia applications. To avoid the need for manual labeling of regions in the supervised learning of visual concepts in content-based indexing systems, Lim and Jin propose a hybrid learning framework for the discovery of semantically meaningful local image regions, such that representative samples of these regions can be generated with minimal human intervention. Supervised learning is first applied to train image classifiers based on a small subset of labeled images. This is followed by the discovery of local semantic regions through the clustering of image blocks with high classifier outputs. In other words, supervised and unsupervised learning techniques are combined to identify visual patterns which are representatives of each semantic class.

In the next paper, Tong et al. describe a new keyword propagation approach for image retrieval based on a recently developed manifold-ranking algorithm. Specifically, a keyword model is constructed based on a small subset of labeled images by the manifold-ranking algorithm, through which all images in the database are softly annotated. The distinguishing characteristic of this approach is its emphasis on the exploration of relationship between all labeled and unlabeled images in the learning stage, instead of constructing a separate classifier for each keyword in conventional approaches.

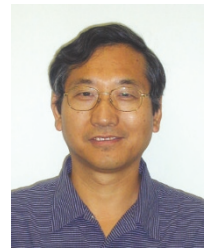
An alternative approach for bridging the semantic gap in image retrieval is to include an intermediate level between the low-level and high-level representations, as proposed by Raicu and Sethi in their paper. Based on latent semantic indexing techniques from the field of information retrieval, they introduce a new type of image feature, which consists of specific patterns of colors and intensities, for capturing the latent association between visual feature elements within an image, and across different images in the database. This intermediate level of representation will facilitate the learning of associations between image features and semantic concepts.

The focus of the paper by Falelakis et al. is on a new approach for balancing between the computational cost (complexity) of semantic identification, and the accuracy (validity) of the identification results. Based on the availability of a semantic encyclopedia for identifying the semantic entities in multimedia documents, hierarchical semantic concepts are modeled by means of finite automata. In this way, efficient approaches are designed for semantic search and indexing, taking into account the tradeoff between computational cost and achieved validity of the identification.

Motivated by the increased adoption of the MPEG-7 standard in mobile multimedia applications, Kofler-Vogt et al. introduce a data structure, in the form of a B-tree, for indexing XML-based MPEG-7 data, and propose an associated coding scheme which allows the streaming of this index tree in a limited-bandwidth environment. The resulting improved efficiency based on the proposed approach will help to facilitate the performance of multimedia content search on mobile platforms.

We would like to take this opportunity to express our thanks to the contributing authors and the reviewers for their efforts, and we hope that the work described in the papers of this issue will inspire new research directions in multimedia information mining.

Ling Guan
Horace H. S. Ip
Paul H. Lewis
Hau San Wong
Paisarn Muneesawang



Ling Guan received his B.S. degree in electronic engineering from Tianjin University, China, in 1982, M.S. degree in systems design engineering at University of Waterloo, Canada, in 1985, and Ph.D. degree in electrical engineering from University of British Columbia, Canada, in 1989. From 1993 to 2000, he was on the Faculty of Engineering at the University of Sydney, Australia. Since May 2001, he has been a Professor in electrical and computer engineering at Ryerson University, Canada. In 2001, he was appointed to the position of Tier I Canada Research Chair. He is the recipient of Ontario Outstanding Researcher's Award in 2002, and IEEE Transactions on Circuits and Systems for Video Technology Best Paper Award in 2005. He held visiting positions at British Telecom (1994), Tokyo Institute of Technology (1999), Princeton University (2000), and Microsoft Research Asia. Dr. Guan has authored/coauthored more than 200 scientific publications in multimedia processing and communications, computer vision, machine learning, and adaptive image/signal processing. He served as Associate Guest Editor of numerous international journals, including Proceedings of the IEEE, IEEE Signal Processing Magazine, and two IEEE Transactions. He was the founding General Chair of IEEE Pacific-Rim Conference on Multimedia, and currently serves as the General Chair of 2006 IEEE International Conference on Multimedia and Expo to be held in Toronto, Canada.

Horace H. S. Ip received his B.S. (first-class honours) degree in applied physics and Ph.D. degree in image processing from University College London, United Kingdom, in 1980 and 1983, respectively. Presently, he is the Chair Professor of the Computer Science Department and the founding Director of the AIMtech Centre (Centre for Innovative Applications of Internet and Multimedia Technologies) at City University of Hong Kong. His research interests include image processing and

analysis, pattern recognition, hypermedia systems in education, and computer graphics. Professor Ip is the Chairman of the IEEE (Hong Kong Section) Computer Chapter, and the founding President of the Hong Kong Society for Multimedia and Image Computing. He has published over 160 papers in international journals and conference proceedings. Professor Ip is a Member of the IEEE, a Fellow of the Hong Kong Institution of Engineers (HKIE), Fellow of the Institution of Engineers (IEE), UK, and Fellow of the International Association for Pattern Recognition (IAPR).

Paul H. Lewis received the B.S. degree in physics from Imperial College, London, and a Ph.D. degree in physics from London University in 1972. He is a Professor in the Intelligence, Agents, Multimedia Group in the School of Electronics and Computer Science at the University of Southampton in the UK. His main research interests are in the area of image and video content analysis, semantic analysis and applications to intelligent multimedia information handling, and data mining. Particular application areas include the medical domain and cultural heritage systems.

Hau San Wong is currently an Assistant Professor in the Department of Computer Science, City University of Hong Kong. He received the B.S. and M.Phil. degrees in electronic engineering from the Chinese University of Hong Kong, and the Ph.D. degree in electrical and information engineering from the University of Sydney. He has also held research positions in the University of Sydney and Hong Kong Baptist University. His research interests include multimedia signal processing, neural networks, and evolutionary computation. He is the coauthor of the book *Adaptive Image Processing: A Computational Intelligence Perspective*, which is a joint publication of CRC Press and SPIE Press, and was an Organizing Committee Member of the 2000 IEEE Pacific-Rim Conference on Multimedia and 2000 IEEE Workshop on Neural Networks for Signal Processing, which were both held in Sydney, Australia. He has also coorganized a number of conference special sessions, including the special session on "Image Content Extraction and Description for Multimedia" in 2000 IEEE International Conference on Image Processing, Vancouver, Canada, and "Machine Learning Techniques for Visual Information Retrieval" in 2003 International Conference on Visual Information Retrieval, Miami, Fla.

Paisarn Muneesawang received the B.Eng. degree from Mahanakorn University of Technology, Thailand, in 1996. He received the M.Eng.Sc. degree in electrical engineering from the University of New South Wales in 1999, and the Ph.D. degree in electrical and information engineering from the University of Sydney in 2002. In 2003-2004, he held Postdoctoral Research Fellow position at Ryerson Multimedia Research Laboratory, Ryerson University, Canada. He was a faculty member at Naresuan University, Thailand, from 1996 to 2004. Since February 2005, he has been an Assistant Professor in College of Information Technology at the University of United Arab Emirates. His research interests include multimedia signal processing, information system, computer vision, and machine learning.