The Holy Grail of Multimedia Information Retrieval: So Close or Yet So Far Away?

BY ALAN HANJALIC
Guest Editor

RAINER LIENHART
Guest Editor

WEI-YING MA
Guest Editor

JOHN R. SMITH
Guest Editor

I. WHAT IS MULTIMEDIA INFORMATION RETRIEVAL?

The tremendous interest of users in multimedia information retrieval is driving significant research focus in this problem space. The research objectives are to develop technology breakthroughs that enable fast, natural, intuitive, and personalized "content-based" access to vast multimedia data collections. Research is building on the synergy of many scientific disciplines, such as signal processing, pattern recognition, machine learning, information retrieval, information theory, natural language processing, human–computer interaction, and psychology. The earliest research on image query-by-example, has succeeded in inspiring and mobilizing a significant and growing amount of research worldwide on a complex set of problems and solutions to address user's needs for efficient and effective multimedia information retrieval.

Papers cover the main aspects of multimedia information retrieval research, assess the applicability of the obtained results in real-life scenarios, and address the many future challenges in this field.

Multimedia information retrieval refers to a set of theories, algorithms, and systems that aim at extracting pertinent descriptors or metadata related to multimedia content and allowing search, retrieval, and other user functions. Experience shows that users typically need to access multimedia content at the *semantic* level. In other words, the descriptors or metadata should allow for direct interpretation by humans. As a result, a significant focus of multimedia information retrieval research is to develop techniques that "bridge the semantic gap" using algorithms that extract machine-level audio–visual feature descriptors and produce semantic-level descriptors. Here, semantic-level can involve a description of the visual scenes, e.g., indoors versus outdoors, nature versus cityscape, crowd scenes; objects, e.g., car, face, building; events, e.g., sports, car chases; or more abstract concepts, e.g., happiness. We will further refer to the machine-level data interpretation as *semantic inference*. We will use the term *multimedia* to refer to both to the diversity of modalities (e.g., images, video, text, music, speech) but also to the complex, compound (multimodal) data sets (e.g., a video with an

Digital Object Identifier: 10.1109/JPROC.2008.916338

accompanying sound track and closed captions). Finally, we will use the term *retrieval* to cover a vast diversity of the ways for the users to access multimedia. In the following sections, we briefly address approaches for multimedia retrieval.

A. Query by Example

Query by example (QBE) involves using an example of the content of interest to illustrate the user's information need. QBE emerged during the mid 1970s and has its origin in text-based information retrieval. Conceptually, the example (*query*) can be in the form of multimedia, e.g., images, video, text, etc. However, the extension from QBE using text to other modalities has long been considered difficult due to the richness of multimedia. Recently, new ideas have emerged that are likely to prove the opposite (J. Sivic and A. Zisserman: "Efficient Visual Search for Objects in Video," L. Kennedy *et al.*: "Query-Adaptive Fusion and Multimodal Search," X. Xie *et al.*: "Mobile Search with Multimodal Queries," and M. Casey *et al.*: "Current and Future Directions in Music Information Retrieval").

B. Retrieval Through Semantic Indexing

Semantic indexing involves using techniques that bridge the semantic gap to allow users to search at the semantic level. To make this possible, links between text-based search terms and extracted semantic descriptors need to be established. In the past, the problem was avoided by relying on manual annotation of multimedia repositories to make them searchable. However, with the tremendous growth of data and high costs associated with manual annotation, machine-based techniques are needed. Techniques based on machine learning and pattern classification are being developed to extract semantic descriptors and support effective searching by users (A. Hauptmann *et al.*: "Video Retrieval Based on Semantic Concepts," L. Xie *et al.*: "Event Modeling and Pattern Mining in Multimedia Streams").

C. Interactive Retrieval

Relevance feedback (RFB) is a variation of QBE that involves multiple interations with a user at search time. In RFB, the query is refined over stages in which the user indicates which retrieved examples match or do not match the user's need. It is expected that for RFB to be useful relevant matches should be found after a small number of iterations. In designing techniques to solve this problem, RFB can be viewed as an online version of bridging the semantic gap, in which positive and negative examples of semantic concept, in this case a user-defined concept, are collected at query time. There are also variations of RFB, such as active learning, that use different approaches for determining what input to collect from the user. Overall, the challenge is to harness machine learning at query time to provide efficient multimedia information retrieval (T. Huang *et al.*: "Active Learning and Relevance Feedback for Interactive Multimedia Retrieval" and M. Kankanhalli and Y. Rui, "Application Potential of Multimedia Information Retrieval").

D. Personalized and Adaptive Content Delivery

Personalized and adaptive content delivery goes beyond relying solely on processing the multimedia content to utilizing context to provide effective multimedia information retrieval. The goal is to maximize the overall *quality of experience* (QoE) for the user. QoE can be seen as a measure for the ability to protect the user from the information overload associated to large data collections and maximize the relevance of the retrieved or delivered content. For example, this can be done by better understanding user's context and possibly anticipating the user's needs depending on user preferences, usage environment, device capabilities, and so on (S. Venkatesh *et al.*: "You Tube and I Find—Personalizing Multimedia Content Access," F. Pereira *et al.*: "Multimedia Retrieval and Delivery: The Role of Metadata Standards").

II. PROGRESS SO FAR

The research progress in the field of multimedia information retrieval since the first shy steps made in the late 1980s has been immense. This progress went through two major phases. The initial phase lasting until the mid 1990s was characterized by explorative research efforts that targeted the discovery of the basic approaches to represent and process multimedia data for the purpose of data interpretation at the semantic level. This phase resulted: 1) in the first insights into the number and type of features that can best be used to represent image, video, and audio data for the purpose of semantic inference and 2) in numerous, though heavily heuristic, methods to enable basic machine-level indexing and retrieval of data. Typical examples are image indexing and retrieval methods like *Find all images containing a large dark green area near the top of the image*, or video segmentation into basic temporal units (shots).

The second phase brought a shift towards higher semantic levels (e.g., from shot-based towards scene-based video segmentation) and the awareness of a high potential of the multimodal semantic inference. It also introduced sophisticated probabilistic inference mechanisms and many innovative ideas of cross-modal indexing and retrieval into the field. This is also the phase of rapidly growing scale at which the semantic inference is performed and intensive integration and exchange of algorithms and datasets, which all together contributed to a rapid increase in significance and relevance of the reported results (M. Kankanhalli and Y. Rui: "Application Potential of Multimedia Information Retrieval"). The steadily increasing research output in the field has already conquered many already established conferences and journals

in terms of paper submissions and has led to a number of successful dedicated conferences (e.g., ACM Conference on Image and Video Retrieval), workshops (e.g., IEEE Workshop on Content-Based Access of Image and Video Libraries, ACM Workshop on Multimedia Information Retrieval), and benchmarking platforms (e.g., U.S. NIST's TRECVID) that have been growing both in terms of the volume and importance of the reported research results ever since. Also to be mentioned at this place are the intensive standardization efforts that emerged from the research and that have been conducted under the umbrella of different standardization bodies, among which the *moving picture experts group* (MPEG) (the MPEG-7 and MPEG-21 standards) and the *TV-Anytime Consortium* (the TV-Anytime Standard) (F. Pereira *et al.*: "Multimedia Retrieval and Delivery: The Role of Metadata Standards").

All in all, solutions have been proposed that seem to bring us rather close to the holy grail of the multimedia information retrieval research, namely *getting the access to the content we like quickly and easily whenever we like it and wherever we are*. But, are we really so far?

III. SPECIAL ISSUE

After almost two decades of exploding interest in the multimedia information retrieval field, and an overwhelming interdisciplinary effort towards making large multimedia content archives easily accessible in the applications ranging from home media libraries, via multimedia lectures archives, content navigation for broadcast TV and video on demand, to advanced automated surveillance systems, there has been enough research momentum generated in this field to reflect on the overall progress and envision future challenges. The papers contributed to this special issue are carefully selected to cover the main aspects of the multimedia information retrieval research, to highlight the successes but also to critically analyze the achievements booked so far and assess the applicability of the obtained results in real-life scenarios. This special issue does not only provide insights into the current possibilities for building (semi) automated methods and algorithms for segmenting, abstracting, indexing, representing, browsing, and retrieving multimedia content in various applications context, but it also addresses the many challenges that are likely to drive the research in this field in years to come. In the remainder of this section we summarize the content of this special issue by providing and integrating (brief or slightly modified) abstracts of the papers published in this issue.

In the first paper of this issue, "Efficient Visual Search for Objects in Video," J. Sivic and A. Zisserman describe an approach to generalize the concept of a text-based search to nontextual information. In particular, they elaborate on the possibilities of retrieving objects or scenes in a movie with the ease, speed, and accuracy with which Google retrieves web pages containing particular words, by specifying the query as an image of the object or scene. In their approach each frame of the video is represented by a set of viewpoint invariant region descriptors. These descriptors enable recognition to proceed successfully despite changes in viewpoint, illumination, and partial occlusion. Vector quantizing these region descriptors provides a visual analogy of a word, which they define as a "visual word." Efficient retrieval is then achieved by employing methods from statistical text retrieval, including inverted file systems, and text and document frequency weightings. This paper also gives the reader an outstanding overview of the current state-of-the-art in visual object search, in particular regarding the problems of: 1) building visual vocabularies for very large-scale retrieval; 2) retrieval of 3-D objects; and 3) thorough verification and ranking using spatial structure of objects.

An important recently gained insight regarding the optimization of the QBE search performance is that applying a single standard search method for all possible queries is inadequate. In other words, a successful retrieval system needs to be able to interpret a query, extrapolate its intention, and adapt its optimal search strategy accordingly. In their paper "Query-Adaptive Fusion and Multimodal Search," L. Kennedy, S.-F. Chang, and A. Natsev conduct a broad survey of query-adaptive search strategies in a variety of application domains. While these query-adaptive approaches can range from metasearch over text collections to multimodal search over video databases, the authors propose that all such systems can be framed and discussed in the context of a single, unified framework. In their analysis the authors concentrate on the domain of video search, where search cues are available from a rich set of modalities, including textual speech transcripts, low-level visual features, and high-level semantic concept detectors. The relative efficacy of each of the modalities is highly variable between many types of queries. The authors observe that the state-of-the-art in query-adaptive retrieval frameworks for video collections is highly dependent upon the definition of classes of queries, which are groups of queries that share similar optimal search strategies, while many applications in text and web retrieval have included many advanced strategies, such as direct prediction of search method performance and inclusion of contextual cues from the searcher. The authors conclude that such advanced strategies previously developed for text retrieval have a broad range of possible applications in future research in multimodal video search.

The popularity of mobile devices has grown rapidly over the last couple of years, and it is expected that more and more people will also search the Web while they are on the move. In addition to text-based keyword queries, mobile devices can

support richer and hybrid queries, such as images, audio, video, and their combinations. In their paper "Mobile Search with Multimodal Queries," X. Xie, L. Lu, M. Jia, H. Li, F. Seide, and W.-Y. Ma discuss mobile search systems that support image and audio queries, covering typical designs for mobile visual and audio search, as well as the further research and development opportunities and challenges related to mobile search. To illustrate current possibilities for mobile search, the paper presents an in-depth study of two real systems developed at Microsoft: product image categorization and mobile ringtone search, which use image and audio queries, respectively. Experimental results on large-scale real-life data demonstrate their effectiveness and efficiency.

How to find video clips showing or being related to "Bill Clinton?" How to retrieve images containing a "sunset," "beach," or a "landscape?" How to find all "romantic" music pieces, or TV news items about "United Nations?" The terms within the quotes are examples of high-level semantic concepts that represent the most intuitive abstraction level at which multimedia information is explored and searched for by a human. In their paper "Video Retrieval Based on Semantic Concepts" A. Hauptmann, M. Christel, and R. Yan present an approach that uses many intermediate semantic concepts with the potential to bridge the semantic gap between what a color, shape, and texture-based "low-level" image analysis can extract from video and what users really want to find, most likely using text descriptions of their information needs. Semantic concepts such as cars, planes, roads, people, animals, and different types of scenes (outdoor, night time, etc.) can be automatically detected in the video with reasonable accuracy. But then, how can they be used automatically and how does a user (or a retrieval system) translate the user's information need into a selection of related concepts that would help find the relevant video clips from the large list of available concepts? The authors illustrate how semantic concept retrieval can be automatically exploited by mapping queries into query classes and through pseudo-relevance feedback. This paper also provides insights in how semantic concepts can be utilized by users in interactive retrieval, through interfaces that provide affordances of explicit concept selection and search, concept filtering, and relevance feedback. Furthermore, the questions critical for practical applicability of the proposed approach are addressed, such as how many concepts are actually needed, and how accurately they need to be detected and linked through various relationships specified in the ontology structure.

Also addressing the problem of semantic concept detection, the paper "Event Modeling and Pattern Mining in Multimedia Streams" by L. Xie, H. Sundaram, and M. Campbell concentrates on an important category of concepts, namely those describing *events*. Events are seen there as real-world occurrences that unfold over space and time. This paper contains a survey about the problems and solutions in event mining, approached from three aspects: event detection, event modeling components, and current event mining systems. In particular, the authors present a general characterization of multimedia events, motivated by the maxim of Five Ws and one H for reporting real-world events in journalism: when, where, who, what, why, and how. They discuss the causes for semantic variability in real-world descriptions, including multilevel event semantics, implicit semantics and facets, and the influence of context. Furthermore, five main aspects of an event detection system are also addressed, namely: 1) the variants of tasks and event definitions that constrain system design; 2) the media capture setup that collectively defines the available data and necessary domain assumptions; 3) the feature extraction step that converts the captured data into numeric or symbolic forms (e.g., color, texture, regions, and objects); 4) statistical models that convert the feature representations to richer semantic descriptions, including mechanism to incorporate domain knowledge; and 5) applications that make use of event metadata to help in different information-seeking tasks. This paper also gives an excellent introduction to the field of event modeling and pattern mining in multimedia streams with many pointers to further reading material.

In their paper "Active Learning and Relevance Feedback for Interactive Multimedia Retrieval," the authors T. Huang, E. Chang, D. Ellis, C. Daglie, S. Rajaram, M. Mandel, and G. Poliner concentrate on the technology that is considered to lead the way towards the emerging new media retrieval paradigm, namely *interactive* multimedia retrieval. The authors systematically lay down the concepts of how to efficiently nail down the user's retrieval concept, vividly motivate and explain the various approaches on active learning, and give an excellent literature overview for further reading.

The paper "Current and Future Directions in Music Information Retrieval" by M. Casey, R. Veltkamp, M. Goto, M. Leman, M. Slaney, and C. Rodes starts by discussing the steep rise in music downloading over CD sales. This rise has created a major shift in the music industry away from physical media formats and towards online products and services. Music is one of the most popular types of online information and there are now hundreds of music-streaming and download services operating on the World-Wide Web. Some of the music collections available are approaching the scale of ten million tracks and this has posed a major challenge for searching, retrieving, and organizing music content. Research efforts in music information retrieval have involved experts from music perception, cognition, musicology, engineering, and computer science engaged in a truly interdisciplinary activity that has resulted in many proposed algorithmic

and methodological solutions to music search using content-based methods. This paper outlines the problems of content-based music information retrieval and explores the state-of-the-art methods using audio cues (e.g., query by humming, audio fingerprinting, content-based music retrieval) and other cues (e.g., music notation and symbolic representation) and identifies some of the major challenges for the coming years.

How to filter huge collections of multimedia data and deliver the content to the user according to his/her current needs? The research efforts on personalizing content access have grown continuously and have become one of the most important branches of the multimedia information retrieval research field. Still, a renewed focus on the subjective dimension in the multimedia lifecycle, from creation, distribution to delivery, and consumption, is required to address this need beyond what is feasible today. Integration of the subjective aspects of the *media* itself, and in particular its affective, perceptual, and physiological potential (both intended and achieved), together with those of the *users* themselves will allow for personalizing the content access, beyond today's facility. This integration, transforming the traditional multimedia information retrieval indexes to more effectively answer specific user needs, will allow a richer degree of personalization predicated on user intention and mode of interaction, relationship to the producer, and content of the media and their history and lifestyle. In their paper "You Tube and I Find—Personalizing Multimedia Content Access," the authors S. Venkatesh, B. Adams, D. Phung, C. Dorai, R. Farell, L. Agnihotri, and N. Dimitrova identify the challenges in achieving this integration and discuss current approaches to interpreting content creation processes, user modeling and profiling, as well as personalized content selection.

While the papers introduced above address the technical aspects and challenges of multimedia information retrieval, the paper "Application Potential of Multimedia Information Retrieval" by M. Kankanhalli and Y. Rui presents a survey of the existing state of impact of the research in the field of multimedia information retrieval in real-world applications. The paper also analyzes the current research trends in the field that may have an influence on future applications. Furthermore, the authors detail the future possibilities and bottlenecks in applying the multimedia information retrieval research results in the main target application areas, such as consumers (e.g., personal video recorders, web information retrieval), public safety (e.g., automated smart surveillance systems), and the professional world (e.g., automated meeting capture and summarization). In particular, recommendations will be made to the research community regarding the challenges that need to be met in order to make the knowledge transfer towards the applications more efficient and effective.

While retrieval refers to the process by which a user, human or machine, identifies the content it needs, *delivery* refers to the adaptive transport and consumption of the identified content in a particular context or usage environment. Both the retrieval and delivery processes may require content and context metadata. In their paper "Multimedia Retrieval and Delivery: The Role of Metadata Standards," the authors F. Pereira, T. Sikora, and A. Vetro argue that maximum quality of experience depends not only on the content itself (and thus content metadata) but also on the consumption conditions (thus context metadata). To enable efficient management and exchange of multimedia information in many application domains, interoperability between coded data and metadata is required. The necessary level of interoperability can only be achieved through standardization. This paper reviews the existing multimedia standards related to information retrieval and adaptive delivery of multimedia content, emphasizing the need for such standards and showing how these standards can help the development, dissemination, and valorization of research results in the field of multimedia information retrieval. Moreover, the paper also discusses limitations of the current standards and anticipates what future standardization activities would be needed.

IV. WHAT NEXT?

All the exciting achievements in the multimedia information retrieval field in the past years, and as reported in this issue, have made the field mature enough to enter a new development phase—the phase in which multimedia information retrieval technology should start to be adopted in practical solutions. As is the case in the broader field of information retrieval, machine-based solutions will always involve tradeoffs in precision versus recall or detection rate versus false alarm rate. Users' expectations for multimedia information retrieval is generally high, and it will continue to be hard for users to understand why computers cannot "see" and understand images the way humans can. However, the results outlined in the special issue show that a variety of multimedia retrieval techniques are providing value for users and significant hurdles in development of multimedia information retrieval technology have been passed. As research continues, there will be a need to improve dependability of multimedia information retrieval systems. For example, in the case of consumer-oriented multimedia retrieval solutions (e.g., a personal video recorder, a mobile video retrieval system, a music search framework, a web search engine), the service they provide in terms of the paradigm *content I like, anytime and anyplace* will have to be at least as dependable as the button turning a TV set or a mobile device on and off. In other domains, such as automated surveillance, where semantic inference solutions may be employed to automatically analyze surveillance data and alert the authorities in case of

a threatening situation, the dependability plays even a more critical role.

An increase in dependability could be achieved in several ways. For instance, a better understanding of the processes underlying semantic concept detection could help forecast, prevent, or at least correct possible semantic inference errors. Furthermore, next to optimizing the performance of an individual (local) semantic inference system, the theory of using redundancy for building reliable structures from less reliable components could be applied to integrate a local multimedia information retrieval solution, e.g., a content classification system, into a network characterized by distributed and collaborative intelligence. This could be either a social network of users tagging the content and seamlessly recommending the content to each other through collaborative filtering and/or a network of semantic inference devices that learn from each other and complement each other. ■

Acknowledgment

The Guest Editors would like to express their gratitude to the contributors to this special issue for their dedication to create exceptional papers covering the most important research problems in the field of multimedia information retrieval. Special thanks go to Prof. J. Biemond from the Delft University of Technology, The Netherlands, for his support and valuable advice in the process of creating this special issue and to Ms. J. Sun for her technical and administrative assistance.

ABOUT THE GUEST EDITORS

Alan Hanjalic received the Diplom-Ingenieur (Dipl.-Ing.) degree from the Friedrich-Alexander University, Erlangen, Germany, in 1995, and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 1999, both in electrical engineering.

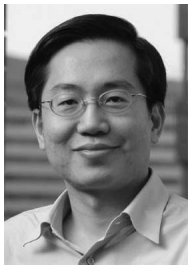
He is an Associate Professor of Multimedia Technology at the Department of Mediamatics, Delft University of Technology. He was a Visiting Scientist and Research Fellow at Hewlett–Packard Labs, Palo Alto, CA, British Telecom Labs, Ipswich, U.K., Philips Research Labs, Briarcliff Manor, NY, and Microsoft Research Asia, Beijing, China. His research interests and expertise include the broad area of multimedia information retrieval, with a focus on multimedia content analysis and personalized content delivery. He authored and coauthored more than 70 publications, among which the books entitled *Image and Video Databases: Restoration, Watermarking and Retrieval* (Elsevier, 2000) and *Content-Based Analysis of Digital Video* (Kluwer, 2004). He holds a number of EU/U.S. patents.

Dr. Hanjalic is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA. Next to his role in organizing this Special Issue of the PROCEEDINGS OF THE IEEE, he was also a Guest Editor of the *International Journal of Image and Graphics* (World Scientific), Special Issue on Content-based Image and Video Retrieval, July 2001, a Guest Editor of the *Journal of Visual Communication and Image Representation* (Elsevier), Special Issue on Emerging Techniques for Multimedia Content Sharing, Search and Understanding, December 2008, a Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, Special Issue on Integration of Content and Context for Multimedia Management, January 2009, and the Lead Guest Editor of the *EURASIP Journal on Image and Video Processing*, Special Issue on Dependable Semantic Inference, June 2009. Dr. Hanjalic was a Keynote Speaker at the Pacific-Rim Conference on Multimedia, Hong-Kong, December 2007. He has also served as a Co-Chair of the IS&T/SPIE Conference on Multimedia Content Analysis, Management and Retrieval 2006, Co-Chair of the IS&T/SPIE Conference on Multimedia Content Access: Algorithms and Systems 2007, Technical Program Co-Chair of the ACM Multimedia 2007 conference, Workshops Co-Chair of the ACM Multimedia 2006 conference, and as Area/Track Co-Chair of the MMM 2007, IEEE ICME 2007, PCM 2007 and WWW 2008 conferences. Dr. Hanjalic currently serves as a Technical Program Co-Chair of the ACM CIVR 2008 conference, and will serve as a General Co-Chair of the ACM Multimedia 2009 conference. Dr. Hanjalic was the initiator and organizer of the first Dutch Symposium on Multimedia Retrieval, Eindhoven, The Netherlands, January 2002, and of the Symposium "Multimedia at your Fingertips," Delft, The Netherlands, November 2006, addressing the possibilities and bottlenecks for boosting the multimedia knowledge transfer in the Netherlands towards new products and services. He serves on Technical Program Committees of numerous conferences, and as a reviewer for more than 20 journals in the broad fields of computer science and multimedia. He is an Advisor/Reviewer of the Swiss National Science Foundation (SNSF), Austrian Science Fund (FWF) and Flemish Science Foundation (IWT) in the fields of multimedia and information technology and systems.

Rainer Lienhart received the Ph.D. degree in computer science from the University of Mannheim, Germany, in 1998, where he was a member of the Movie Content Analysis Project (MoCA).

He is a Full Professor in the Computer Science Department, University of Augsburg, Germany, leading the lab for multimedia computing. His group is focusing on all aspects of very large-scale image, video, and audio mining algorithms. From August 1998 to July 2004, he was a Staff Researcher at Intel's Microprocessor Research Lab, Santa Clara, CA, where he worked on transforming a network of heterogeneous distributed computing platforms into an array of audio/video sensors and actuators capable of performing complex DSP tasks such as distributed beamforming, audio rendering, audio/visual tracking, and camera array processing. In particular, this requires putting distributed heterogeneous computing platforms with audio–visual sensors into a common time and space coordinate system. At the same time, he was also continuing his work on media mining, where he is well known for his work in video content analysis with contributions in text detection/recognition, commercial detection, face detection, shot and scene detection, and automatic video abstraction. He has published over 50 papers in major conferences and journals and filed more than 20 patents.

Dr. Lienhart as been a committee member of the ACM Multimedia, IEEE International Conference on Multimedia Systems, IEEE International Conference on Multimedia Expo, SPIE Storage and Retrieval of Media Databases, IEEE Workshop on Content-Based Access of Image and Video libraries, and the International Eurographics Workshop on Multimedia. He is a Reviewer for IEEE TRANSACTIONS ON PATTERN RECOGNITION AND MACHINE LEARNING, IEEE TRANSACTIONS ON MULTIMEDIA, *Journal of Computer Vision*, and *Image Understanding and ACM Multimedia Systems Journal*.



Wei-Ying Ma received the B.S. degree in electrical engineering from the National Tsing Hua University, Taiwan, R.O.C., in 1990, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 1994 and 1997, respectively.

He is a Principal Researcher and a Research Area Manager at Microsoft Research Asia, Beijing, China. He leads a team of talented, passionate researchers to advance the state-of-the-art in Web search and data mining. Under his leadership, the team is pushing the boundaries of current search technologies by leveraging machine learning and knowledge discovery techniques to deliver the next generation of Web search. This includes projects in information analysis, organization, retrieval, and visualization. Over the past few years, the team has generated an impressive record of conference publications, including key papers at SIGIR, WWW, and ACM Multimedia. The research has included direct transfers to products, including key technologies delivered to Windows Live image search, product search, mobile search, and academic search. Before joining Microsoft in 2001, he was with Hewlett–Packard Labs, Palo Alto, CA, where he worked in the field of multimedia adaptation and distributed media services infrastructure. From 1994 to 1997, he was engaged in the Alexandria Digital Library project at the University of California, Santa Barbara. During this time, he developed one of the first web-based image-retrieval systems, Netra, which has been widely cited by other researchers and is regarded as one of the most influential image retrieval systems. He has published more than 180 papers in fields such as Web search, information retrieval, content-based image retrieval, data mining, adaptive content delivery, and mobile browsing.

Dr. Ma currently serves on the editorial boards of both the *ACM Transactions on Information System* (TOIS) and *ACM/Springer Multimedia Systems Journal*. He is the Program Cochair of the 17th International World Wide Web Conference (WWW2008), Program Cochair of the Pacific Rim Conference on Multimedia (PCM) 2007, and General Cochair of the Asia Information Retrieval Symposium (AIRS) 2007. He was General Cochair of the International Multimedia Modeling (MMM) Conference 2005 and International Conference on Image and Video Retrieval (CIVR) 2005. He has served on the organizing and program committees of other international conferences, including ACM Multimedia, SIGIR, CIKM, KDD, and WWW.

John R. Smith received the M. Phil. and Ph.D. degrees in electrical engineering from Columbia University, New York, NY, in 1994 and 1997, respectively.

He is a Senior Manager of the Intelligent Information Management Department, IBM T. J. Watson Research Center, Yorktown Heights, NY. He leads a research department addressing technical challenges in multimedia information management and database systems. From 2001 to 2004, he served as Chair of ISO/IEC JTC1/SC29 WG11 Moving Picture Experts Group (MPEG) Multimedia Description Schemes group with responsibilities in the development of MPEG-7 and MPEG-21 standards. He also served as Co-Project Editor for the following parts of MPEG-7 standard: "MPEG-7 Multimedia Description Schemes," "MPEG-7 Conformance," "MPEG-7 Extraction and Use," and "MPEG-7 Schema Definition." He also serves on the Advisory Committee for NIST TREC Video Retrieval Evaluation.

Dr. Smith received the Wall Street Journal 2004 Innovation Award for the MARVEL Multimedia Analysis and Retrieval System, Nov. 2004, IEEE Multimedia Prize Paper Award given to the best paper in the field of Multimedia published in the IEEE TRANSACTIONS ON MULTIMEDIA (over a four-year period), 2003, Eliahu I. Jury Award for outstanding achievement as a graduate student in the areas of systems communication or signal processing, 1997, IBM Corporate Environmental Affairs Excellence Award, 2003, IBM Research Division Award for Content Adaptation Framework, 2001, and IBM Outstanding Innovation Award for SPIRE: Spatial Image Content-Based Retrieval Engine, 2000. He was Co-General Chair of the International Conference on Image and Video Retrieval, 2006, Co-General Chair, ACM Multimedia Information Retrieval (MIR) Workshop, 2005, Co-Program Chair, IEEE International Conference on Multimedia and Exposition, 2005, Co-Program Chair, ACM Multimedia 2003, Nov. 2003, Co-Chair, SPIE Internet Multimedia Management Systems, 2001 to present, Co-Chair, IEEE Workshop on Content-Based Access for Image and Video Libraries (CBAIVL), 2000 to 2001, General Chair, Watson Workshop on Multimedia, June 2003, and Program Area Chair, IEEE International Conference on Multimedia and Exposition, 2002 to 2004. He also served as a Program Committee member and gave tutorials in numerous ACM, IEEE and SPIE/IS&T conferences and workshops. He is an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA, Associate Editor-in-Chief and Standards Editor, *IEEE Multimedia Magazine*, Editorial Board Member for *ACM Computers in Entertainment*, Board Member for *Kluwer International Journal of Multimedia Tools and Applications*, and was a Co-Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, Special Issue on "MPEG-21 Multimedia Framework," April 2005, *Journal of Pattern Recognition*, Special Issue on "Image Understanding for Digital Photographs," Elsevier, 2004, *Journal of Visual Communications and Image Representation*, Special Issue on "Multimedia Database Management Systems," 2004, *International Journal of Imaging Systems and Technology*, Special issue on "Multimedia Content Description and Video Compression," 2003, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, Special Issue on "Conceptual and Dynamical Aspects of Multimedia Content Description," 2002.