Ubiquitous Multimedia: Emerging Research on Multimedia Computing

Yonghong Tian
Peking University

Min Chen
University of Washington Bothell

Leonel Sousa
Universidade de Lisboa, Portugal

Multimedia is ubiquitous in our daily lives. In recent years, many multimedia applications and services have been developed and deployed, including mobile audio/video streaming, mobile shopping, and remote video surveillance, letting people access rich multimedia content anytime, anywhere, and using different access networks and computing devices. It is anticipated that ubiquitous multimedia applications and services will change the way we operate and interact with the world. In this sense, digital multimedia can be viewed as representing a fundamental shift in how we store, transmit, and consume information.

An immediate consequence of ubiquitous multimedia applications and services has been the explosive growth of multimedia data. Today, photos and videos can be easily captured by any handheld device—such as a mobile phone, an iPad, or an iWatch—and then automatically pushed to various online sharing services (such as Flickr and YouTube) and social networks (such as Facebook and WeChat). On average, Facebook receives more than 350 million new photos each day, while 300 hours of video are uploaded to YouTube every minute.[1] Furthermore, Cisco predicts that video will account for 80 percent of all consumer Internet traffic in 2019, up from 64 percent in 2014.[2] Such explosive growth of multimedia data will definitely lead to the emergence of the so-called "big data deluge."[3]

The wide-ranging applications and big data of ubiquitous multimedia present both unprecedented challenges and unique opportunities for multimedia computing research, which were the focus of the 2015 IEEE International Symposium on Multimedia (ISM 2015) held in Miami from 14–16 December 2015. Over the past decade, ISM has established itself as a renowned international forum for researchers and practitioners to exchange ideas, connect with colleagues, and advance the state of the art and practice of multimedia computing, as well as to identify emerging research topics and define the future of this cross-disciplinary field. The ISM 2015 call for papers redefined "multimedia computing" as "one of the computing fields that is generally concerned with presentation, integration, and computation of one or more ubiquitous media, such as text, image, graphics, audio, video, social data, and data collected from various sensors, etc., using computing techniques." Approximately 45 high-quality papers were accepted for ISM 2015, providing novel ideas, new results, and state-of-the-art techniques in the field of ubiquitous multimedia computing. Following this successful event, we aim to provide with this special issue another forum for the researchers of the top symposium papers to further present their research results, potentially increasing the papers' impact on the community.

In this Issue

This special issue is the second successful collaboration between *IEEE MultiMedia* and ISM, which facilitates the publication of the extended versions of the top symposium papers through a fast-track review and publication process. From a total of seven invited submissions, we selected

1070-986X/16/$33.00 © 2016 IEEE

Published by the IEEE Computer Society

four representative articles that investigate the emerging multimedia technology to address the challenges in ubiquitous multimedia data and applications.

An important challenge is how to effectively and efficiently process, analyze, and mine the numerous amounts of multimedia big data.

Next-Generation Video Coding Technology

Although the video compression ratio has doubled in each of the last three decades, it is far behind the growth speed of multimedia data. In addition, this gap is expected to grow even bigger over the next several years, presenting an unprecedented challenge for high-efficiency video coding technology.[4] In the article, "Nonlocal In-Loop Filter: The Way Toward Next-Generation Video Coding?", Siwei Ma, Xinfeng Zhang, Jian Zhang, Chuanmin Jia, Shiqi Wang, and Wen Gao journey through the design philosophy of in-loop filtering, an essential coding tool in H.264/AVC and High Efficiency Video Coding (HEVC), and then present their vision of next-generation (higher-efficiency) video coding technology. Toward this end, they explore the performance of in-loop filters for HEVC with image local and nonlocal correlations.

In their method, a nonlocal similarity-based loop filter (NLSLF) is incorporated into the HEVC standard by simultaneously enforcing the intrinsic local sparsity and the nonlocal self-similarity of each frame in the video sequence. Then, a reconstructed video frame from the previous stage is first divided into overlapped image patches, which are subsequently classified into different groups based on their similarities. Since these image patches in the same group have similar structures, they can be represented sparsely in the unit of a group instead of a block. The compression artifacts can be reduced by hard- and soft-thresholding the singular values of image patches group-by-group, based on the sparse property of similar image patches. Experimental results show that such an in-loop filter design can significantly improve the compression performance of HEVC, providing a new possible direction for improving compression efficiency.

Learning over Multimedia Big Data

For ubiquitous multimedia data, another important challenge is how to effectively and efficiently process, analyze, and mine the numerous amounts of multimedia big data. Machine learning is widely recognized as an effective tool to cope with this challenge. However, the very high dimensionality of features for multimedia data significantly increases the complexity of learning algorithms. One approach for simplification is to assume that the data of interest is on an embedded nonlinear manifold within the higher-dimensional space. In practice, when a data set contains multiple classes, and the structures of the classes are different, a single manifold assumption can hardly guarantee the best performance.

To address this problem, Xin Guo, Yun Tie, Lin Qi, and Ling Guan propose a framework of semisupervised dimensionality reduction for multimanifold learning in their article, "A Novel Semi-Supervised Dimensionality Reduction Framework." Technologically, the framework consists of three components: sparse manifold clustering to group unlabeled samples, cluster label predication to calculate the manifold-to-manifold distance, and graph construction to discover both the geometrical and discriminant structure of the data manifold. Experimental results verify the effectiveness of this multimanifold learning framework.

Multimodal Machine Learning

In "Multimodal Ensemble Fusion for Disambiguation and Retrieval," Yang Peng, Xiaofeng Zhou, Daisy Zhe Wang, Ishan Patwa, Dihong Gong, and Chunsheng Victor Fang address the machine-learning problem on multimedia data from another perspective—namely, from that of multimodal machine learning. Toward this end, they first explain why multimodal fusion works by analyzing the correlative and complementary relations among different modalities. By making use of these two properties, multimodal machine learning could achieve higher quality than single-modality approaches.

Following this idea, the authors design a multimodal ensemble fusion model with different ensemble approaches for word sense disambiguation and information retrieval, which combines the results of text-only processing and image-only processing to achieve better quality. Experimental results on the University of Illinois at Urbana–Champaign Image Sense Discrimination (UIUC-ISD) dataset and the Google-MM dataset demonstrated the effectiveness of the proposed model.

Unsupervised Recurring-Pattern Detection

Ubiquitous multimedia computing can also be explored in many new applications and services. In the article, "Planogram Compliance Checking Based on Detection of Recurring Patterns," Song Liu, Wanqing Li, Stephen Davis, Christian Ritz, and Hongda Tian present a recurring-pattern-detection method for automatic planogram compliance checking, which is referred to as the verification process used by the company headquarters to verify whether each chain store follows the planograms created by headquarters to regulate how products should be placed on shelves. In their method, product layout is extracted from an input image by means of unsupervised recurring-pattern detection and matched via graph matching with the expected product layout specified by a planogram to measure the level of compliance. A divide-and-conquer strategy is employed to improve the speed. Specifically, the input image is divided into several regions based on the planogram. Recurring patterns are detected in each region and are merged together to estimate the product layout. Experimental results on real data verify the efficacy of the proposed method.

Future Directions

Nevertheless, many technical challenges are yet to be addressed in the field of ubiquitous multimedia computing. We thus envision several future research directions in this field that are worthy of attention from the multimedia community.

Ultra-High Efficiency Compression

Considering that the volume of multimedia big data approximately doubles every two years,[5] multimedia compression and coding technologies are far from where they need to be. To achieve much higher and even ultra-high coding efficiency, one potential solution is to introduce vision-based mechanisms and models into the coding framework and develop vision-based coding theories and technologies to substitute the traditional signal-processing-based coding framework. Moreover, it is also highly desirable to develop joint compression and coding technology for multiple media data, such as video, audio, and virtual reality data, which is essential to some attractive ubiquitous multimedia applications in unmanned aerial vehicles, self-driving cars, and augmented-reality products.

Brain-Like Multimedia Intelligence

For machine learning research, ubiquitous multimedia indicates both a major challenge and an important opportunity. On the one hand, larger-scale multimedia data available nowadays can lead toward advanced machine learning techniques, such as deep learning.[1] On the other hand, new machine learning models and algorithms are still in urgent need of ways of efficiently processing, analyzing, and mining ubiquitous multimedia data.

For example, in recent years, deep learning has shown overwhelmingly superior performance compared to traditional machine learning methods based on hand-craft features, such as scale-invariant feature transform (SIFT) in image classification. However, there is still much room for improvement when applied to video content analysis—for example, to recognize actions or detect abnormalities. One promising direction is to develop new intelligence algorithms by structurally and functionally simulating a human brain, leading to *brain-like computation*.[6] This new intelligence paradigm will likely open another door for ubiquitous multimedia computing and applications.

Benchmark Data for Ubiquitous Multimedia Computing

The dataset is a core component of research and development in all scientific fields. More recently, the Yahoo Flickr Creative Commons 100 Million (YFCC100M) dataset has become the largest public multimedia collection ever released, with 100 million media objects.[1] This dataset enables large-scale unsupervised learning, semisupervised learning, and learning with noisy data to address questions across many fields—from computer vision to social computing. However, YFCC100M consists of photos and videos only from Flickr, making it remarkably different from many ubiquitous multimedia data, such as surveillance video. Even so, the availability of such large-scale benchmark data might shift the way in which we cope with the long-standing challenges in ubiquitous multimedia computing, leading to important breakthroughs.

Attractive Applications and Services

The challenges and opportunities highlighted in this field will further foster some interesting developments in ubiquitous multimedia applications and services. Some examples include unmanned aerial vehicles and augmented-reality applications. Looking into the future, ubiquitous multimedia applications will bring new opportunities and driving forces to the research in the related fields.

From the history of information technology, we can see that typically a major breakthrough will happen every 10 to 15 years, which will in turn foster new applications, markets, business models and industrial fields. Clearly, ubiquitous multimedia technology should take the responsibility of leading such a change in the next few years. **MM**

Acknowledgments

We thank all the authors and reviewers for their efforts on this special issue under a very tight schedule. In addition, we thank EIC Yong Rui and AEIC Wenjun Zeng for giving us the opportunity to organize this special issue. Yonghong Tian is partially supported by grants from the National Natural Science Foundation of China under contract No. 61390515 and No. 61425025.

References

1. B. Thomee et al., "YFCC100M: The New Data in Multimedia Research," *Comm. ACM*, vol. 59, no. 2, pp. 64–73.
2. *Cisco Visual Networking Index: Forecast and Methodology*, white paper, Cisco, May 2015; www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white-paper_c11-481360.html.
3. "Dealing with Data: Challenge and Opportunities," *Science*, vol. 331, no. 6018, 2011, pp. 692–693.
4. W. Gao et al., "IEEE 1857 Standard Empowering Smart Video Surveillance Systems," *IEEE Intelligent Systems*, vol. 29, no. 5, 2014, pp. 30–39.
5. J. Gantz and D. Reinsel, *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, tech. report, Int'l Data Corp. (IDC), 2012; www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf.
6. J. Hawkins and S. Blakeslee, *On Intelligence*, Henry Holt and Company, 2004.

Yonghong Tian is a full professor with the Cooperative Medianet Innovation Center, School of Electronics Engineering and Computer Science, Peking University, Beijing, China. His research interests include machine learning, computer vision, and multimedia big data. Tian received his PhD in computer science from the Institute of Computing Technology, Chinese Academy of Sciences. He is currently an Associate Editor of *IEEE Transactions on Multimedia* and the *International Journal of Multimedia Data Engineering and Management* (IJMDEM), and a Young Associate Editor of *Frontiers of Computer Science*. He is a senior member of IEEE and a member of ACM. Contact him at yhtian@pku.edu.cn.

Min Chen is an assistant professor in the Computing and Software Systems Division, School of STEM, at the University of Washington Bothell. Her research interests include distributed multimedia database systems, data mining and their applications on real-life problems, and interdisciplinary projects. Chen received her PhD in computing and information sciences from Florida International University, Miami. She is the treasurer for the IEEE Computer Society Technical Committee on Multimedia Computing, and she's on the IEEE ICME Steering Committee. She was the lead program chair for 2015 IEEE International Symposium on Multimedia. Contact her at minchen2@uw.edu.

Leonel Sousa is a full professor in the Electrical and Computer Engineering Department at Instituto Superior Tecnico (IST), Universidade de Lisboa (UL), Lisbon, Portugal. He is also a senior researcher with the R&D Instituto de Engenharia de Sistemas e Computadores (INESC-ID). His research interests include VLSI and computer architectures, high performance computing, and signal processing systems. Sousa received his PhD in electrical and computer engineering from IST, UL, Portugal. He is currently an Associate Editor of *IEEE Transactions on Multimedia*, *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Access*, *Springer Journal of Real-Time Image Processing (JRTIP)*, and *IET Electronics Letters*, and Editor in Chief of the *Eurasip Journal on Embedded Systems (JES)*. He is Fellow of the Institution of Engineering and Technology (IET), a Distinguished Scientist of ACM, and a senior member of IEEE. Contact him at leonel.seabra@gmail.com.