As the capturing and processing of digital media data becomes increasingly ubiquitous and effortless, multimedia content has penetrated almost every aspect of our life and work, such as education, communication, and even life assistance, and has brought unprecedented challenges as well as opportunities for developing next generation computing and communication platforms.

Multimedia applications are centered on content analysis and delivery and have progressed a long way beyond that nowadays. Multimedia systems rely on the support of technologies from multiple disciplines such as operating systems, networking, signal processing, and system engineering. However, multimedia application systems are not well supported by current operating/computing systems/architectures. Aside from some commonly known issues such as lack of efficient resource management facilities and no support for real time rendering or presentation, a more challenging issue lies in the fact that, with the ever-changing requirements from the user and system perspectives, constant reevaluation and adaptation of multimedia methodologies is needed. Adaptive HTTP video streaming is one such good example that involves the adaptation of existing data streaming technologies to suit the dynamically changing needs of devices and different network conditions by encoding the video in multiple target bitrates. The video quality, largely determined by the bitrate, can be changed dynamically to match the varying bandwidth and thus has fewer problems with buffer underflow or data loss. As another example, in many applications, there is a need to efficiently adapt JPEG images to satisfy given constraints such as file size and quality, from ubiquitous media access for mobile devices to multimedia messaging. However, none of these tasks are trivial, and they often command expensive computations and non-negligible resources. Expertise from multiple fields, such as optimization, high performance computing, and even data mining, is much needed to make a practical system when such complex computing tasks are involved.

Another pressing issue, while being understudied for the past decade, is the interoperability of multimedia systems. Without properly defined APIs and component-based design, it is hard to enable the communication between two different multimedia systems that are built on different platforms and are based on different architectures, and thus it is impossible to standardize the development of interoperable multimedia applications for different systems.
Very few works exist that contribute to this field of research, and most of them try to address this issue by developing middleware based approaches, which requires another dimension of expertise from software engineering.

As interactive multimedia clearly becomes an emerging field of fundamental research and has social, educational, and economical importance, it commands the combination of multiple disciplines. Applications in this kind include interactive TVs, multimedia training/education, video/image composing, etc. Expertise in the fields of human-computer interface, psychology, education, and system engineering are all needed, in addition to domain expertise.

The goal of this special issue is to show some of the state-of-the-art research in the ever-growing field of multimedia data engineering that demands multidisciplinary efforts. This issue contains works from experts in the multimedia, signal processing, data mining, networking, human-computer interface, and system engineering research communities, discussing the current research on complex multimedia systems, and presenting theoretic framework and practical implementations.

ABOUT THIS ISSUE

The articles in this special issue are a selection of the full papers accepted into the 2011 IEEE International Symposium on Multimedia (ISM 2011), held at Dana Point, California, USA, on December 5-7, 2011. ISM is an international forum for researchers to exchange information regarding advances in the state of the art and practice of multimedia computing, as well as to identify the emerging research topics and define the future of multimedia computing. The technical program of ISM 2011 consists of invited talks, full and short paper presentations, workshops, technical demos, and panel discussions. The authors of selected papers were asked to revise their original ISM 2011 presentation and extend it by at least 30% new content. Each submission was rigorously reviewed by a panel of referees selected from the conference program committee and a selected few external reviewers.

In “Reducing Processing Demands for Multi-Rate Video Encoding: Implementation and Evaluation,” Espeland et al. discuss the idea of reusing intra- and inter- predictions in the video encoder to avoid expensive computational cost associated with encoding for multiple target bitrates of the same video object. Multiple instances of encoding are run in parallel, and the redundant steps can be largely avoided by the reuse of intermediate results. The authors use Google’s VP8 encode to implement this idea, and their experimental results show that the computational cost has been significantly reduced at the same rates and close qualities compared to the VP8 reference implementation.

The second article titled “Generating Window of Sign Languages on ITU J.200-Based Middlewares” is authored by Lacet Silva Ferreira et al. from Digital Video Applications Lab at Federal University of Paraiba, Brazil. The authors propose a solution to provide support for sign language in middlewares of Digital TV (DTV) systems compatible with ITU J.200 specifications which aim to promote harmonization between different DTV application environments. Their solution allows sign language content to be signed by 3D-Avatars when human interpreters are not available. Two major improvements over existing architectures are proposed and implemented in this work. The first improvement is the definition of features and APIs necessary to develop interactive DTV application to support sign language contents in different DTV systems, so as to standardize the development of interoperable applications for different DTV systems. Secondly, the proposed solution also includes a signaling protocol for transmission of sign language content from the TV station to the DTV receiver. A case study was conducted to evaluate their solution using Ginga. Their results show low delay and less data transmission for the proposed solution. Real tests with Brazilian deaf show that the users could understand the contents signed by 3D Avatars reasonably well.

The article by Steven Pigeon and Stéphane
Coulombe, titled “K-Means Based Prediction of Transcoded JPEG File Size and Structural Similarity” deals with the problem of JPEG image adaptation under given constraints of file size and resolution. They improve over their previous work, which can predict the near-optimal file size and perceived quality of an image under simultaneous changes in compression ratios and scaling factors. Since no decompression is needed and predictors use only the information directly accessible from the image header, the cost is kept low. In this work, improvements have been made to take better advantage of all the readily available information about the image such as resolution and bits per pixel, which help train a more accurate predictor. In addition, the proposed new approach lifts some of the restrictions of their previous work, particularly the scalability problem with a larger set of parameters and the need for uniform quantization of parameters. This is achieved by using K-Means to cluster transcoding operations represented as high-dimensional vectors, so as to formulate predictions. The experimental results show that the new approach significantly outperforms previous methods in accuracy and has a potential to be combined with some other existing works to form more efficient and more accurate transcoding systems.

Hiroko Mitarai and Atsuo Yoshitaka from Japan Advanced Institute of Science and Technology contribute an article entitled “Emocap: Video Shooting Support System for Non-Expert Users”. In this work, they present a system which enables shooting assistance for better expression of affective information for non-professional users. First, a model is trained to find the proper shooting parameter settings (e.g., face size, camera angles, and zoom speed) for expressing each of the different respective atmospheres according to film grammar, such as tension/excitement, liberation, and loneliness. Based on this model and the user-input affective parameters, the user receives interactive assistance as he captures a scene, enabling him to shoot scenes more effectively and with more expressive visual power, without receiving intensive professional trainings. Their system evaluation shows that most users involved in the test have seen improvements on their captured scenes, through making better decisions in their camerawork following the guidance provided by the system.

ACKNOWLEDGMENT

The editor wishes to thank all of the authors, reviewers and publishers. Without their excellent work, and the devotion of their valuable time, this special issue would not have been possible.

Chengcui Zhang
Guest Editor
IJMDEM