Introduction to: Special Issue on Extended Best Papers from ACM Multimedia 2014

This special issue continues the tradition of inviting the best papers from ACM Multimedia to extend their work to a journal article. In 2015, the conference was held in Orlando, FL, USA. A number of new areas were introduced this year. The two articles presented in this special issue came from the Deep Learning for Multimedia area and the Emotional and Social Signals in Multimedia area.

As usual, a rigorous review process was carried out followed by an intense two-day colocated technical program committee meeting. Selecting the final set of best-paper candidates was a very intense process for all concerned, sparking a lot of debate about how important it is to have best paper candidates that are multimodal and take a fresh perspective on new topics.

The following best paper extensions underwent a rigorous review procedure to ensure that the work was sufficiently extended compared to their respective conference paper versions. We thank the anonymous reviewers who helped to ensure the quality of these two extended papers.

The first article, "Emotion Recognition During Speech Using Dynamics of Multiple Regions of Face" by Yelin Kim and Emily Mower-Provost, addresses the challenging task of performing automated facial emotion recognition when someone is speaking simultaneously. In this article, the authors exploit the context of the speech to disambiguate facial behavior that is caused by speech production from true expressions of facial emotion. They investigate an unsupervised method of segmenting the facial movements due to speech, demonstrating an improvement in facial-emotion recognition performance on the IEMOCAP and SAVEE datasets. Importantly, they describe the correspondence of their experimental findings in relation to existing emotion perception studies. This work is particularly valuable in the development of more naturalistic human-centered and emotionally aware multimedia interfaces.

The second article, "Correspondence Autoencoders for Cross-Modal Retrieval" by Fangxiang Feng, Xiaojie Wang, and Ruifan Li, tackles the task of cross-modal retrieval by using correspondence autoencoder which connects the text and image modality. This enables users to issue text queries and have images retrieved using their shared representations. The authors present three distinct architectures for achieving this. A correspondence cross-modal autoencoder reconstructs its input, which may consist of text phrases or images, while using a shared bottleneck layer (with the text and image belonging to the same entity). In the second variant, the full-modal architecture, both inputs must be reconstructed given a single modality. The final deep architecture employs restricted Boltzmann machines. Experimental results show that the described architectures improve upon previously published literature in this domain on the Wikipedia, Pascal, and NUS-WIDE-10k datasets. Moreover, the authors

1551-6857/2015/10-ART24
DOI: http://dx.doi.org/10.1145/2820400

have made their code publicly available, enabling other researchers to improve on the architectures presented in this publication.

We hope you enjoy these extended articles. It will be interesting to see whether these best papers, which were both from new areas, will start new trends in future sessions of the conference.

Hayley Hung
Intelligent Systems Department, Delft University of Technology
The Netherlands

George Toderici
Google Research
Mountain View, USA

Guest Editors