MMSP2018

IEEE 20th International Workshop on Multimedia Signal Processing

August 29 - 31, 2018, Vancouver, Canada

Proposal for Special Session for MMSP2018

Title: *Multimodal Machine Learning: Advances and Applications*

(Short Title: *Multimodal Machine Learning*)

Rationale for Special Session: Advanced machine learning techniques, such as deep neural networks (DNN), have achieved significant success in machine learning for single modalities (e.g., text, images or audio). However, human's perceptual systems including visual, auditory, olfactory, and other somatosensory systems usually work together in human's cognitive and interaction. Human's brain, as a multimodal perceptron, can process multiple modalities simultaneously and exploit the internal correlations of the multimodal signals. For example, human's emotion in a conversation can be understood better if both the human's voice and behaviors are presented. To mimic the way the human brain works, there is growing interests in combining the multimodal signals as the input to the machine learning engines. However, traditional uniform machine learning models cannot handle the heterogeneity of the multimodal signal well. Multimodal machine learning is therefore a vibrant multidisciplinary field of increasing importance and with extraordinary potential.

MMSP is one of the flagship conferences focusing on the multimedia signal processing. As the most prevalent multimodal signal format, how to process the multimedia signals by the multimodal machine learning model is highly desired by the applications, such as augmented reality, affective computing, scene understanding. Multimodal machine learning would therefore be one of the most relevant and interesting topics for the MMSP conference.

What will be presented?

The multimodal machine learning contains plenty of sub research fields, including multimodal data representation and mapping, multimodal feature fusion and collaborative learning, multi-instance learning, multi-kernel learning, CNN and RNN in multimodal machine learning, etc. Besides the theoretical topics, the applications like image/video understanding, event recognition, multimedia information retrieval, natural language processing, and on-device assistant are also the relevant topics of this special session.

Who will attend?

This special session will attract sufficient attention from both academic and industrial communities. According to the records of significant increasing technical papers from the

academic institutions like MIT, CMU, UC Berkeley, and Tsinghua University of related topics, the academic researchers will have high motivation to present their latest progress in this session. On the other hand, the industry leaders, including Amazon, Apple, Google, Qualcomm, and Samsung all invest on the related products like Amazon Alexa, Apple Siri, Google assistant, Qualcomm HVX, and Samsung Bixby. The industry participants would also be interested to join this session to learn the latest technical progress in this field and bring their requirements for solutions.

Programme

The following summarized the skeleton of this session. As the multimodal machine learning is an emerging research fields, we intend to make this session with a balanced "technologies and future" theme. The session is expected to contain visionary presentations, technical papers, and application demo. All the papers, presentations, and demos will be solicited along the lines of

Advances in the technologies and applications of multimodal machine learning.

Possible topics include but not limited to
1. multimodal data representations
2. multimodal feature fusion and collaborative learning
3. multi-instance machine learning
4. image/video content understanding with multimodal signal
5. natural language processing with multimodal signal
6. applications of multimodal machine learning for augmented/virtual reality
7. on-device applications with multimodal machine learning

A number of papers on these topics will be solicited, to ensure high attendance. Other submitted papers to MMSP2018 may also be relevant to this session.

A sample of papers to be invited:

1. TBD
2.

Session Organizers (More will be added)

The task of the organizers will be to ensure a suite of high quality presenters, with topical and cohesive presentations on spatial acoustics principles, visions for how it will manifest in the future, and research for its development.

Dr. Huisheng Wang, is a Staff engineer in Google Research. She received her Ph.D. degree in the Department of Electrical Engineering, University of Southern California, Los Angeles, in 2007. She then joined Google and worked in a number of product areas including YouTube, Google Glass and Wearables, and Google Lens. She received the 2006 EURASIP Journal on

Advances in Signal Processing Best Paper award. Her research interests include computer vision, machine learning, multimedia compression, signal processing and communications.

Dr. Qifei Wang, is a Software engineer in Google Research. He received his Ph.D. degree in Department of Automation, Tsinghua University, Beijing, China, in 2013. He worked as postdoctoral researcher at UC Berkeley between 2013 and 2014. His main research interests including computer vision, machine learning, image/video processing and streaming.

Dr. Bin Shen, is a Software engineer in Google New York. He received Ph.D. degree from the Department of Computer Science, Purdue University, West Lafayette, IN, USA, in 2014. His research interests include image processing, machine learning, and data mining.