

TOMM Special Section on Multimedia Understanding via Multimodal Analytics

Summary and Scope

Information in the real world, more often than not, comes through multiple input channels, in the form of multimodalities. For instance, Web images are frequently associated with tags and textual descriptions; beside textual captions, videos contain visual and acoustic signals; sensory perception includes simultaneous inputs from visual, auditory, motor and haptic pathways. Rather than independent, the multimodalities are correlated in a complex way, which can be temporally synchronized (e.g., video clips and corresponding audio transcripts), spatially related (e.g., object relationships in 3D space), or otherwise semantically connected (e.g., images and their tags). We thus firmly believe that joint learning by discovering and modeling the relatedness among multiple modalities will remarkably strengthen the learning performance.

A great many of pioneering efforts thus far have been dedicated to early or late fusion of mono-modal analysis, such as looking into pure images, texts and videos separately, but ignoring their associated modalities. This, more or less, suffers from curse of dimensionality or non-informative representation. Therefore, novel techniques are highly desired which can reveal the information encoded in different modalities, sew up this information in a non-trivial way, as well as, exploit the combined information to facilitate multimedia content understanding.

We timely see this opportunity and hence organize this special issue to bring together active researchers all over the world to share their recent advances in the area of multimodal learning. We solicit original contributions in three-fold: (1) present state-of-the-art theories and novel application scenarios related to multimodal learning; (2) survey the recent progress in this area; and (3) build benchmark datasets.

The list of possible topics includes, but is not limited to:

- Modality-wise missing data completion
- Deep models for multimodal aggregation
- Multitask learning in Multimodal settings
- Multimedia question answering
- Multimodal feature extraction, selection and fusion
- Multimodal concept detection, object recognition and segmentation
- Novel machine learning for multimodal analysis
- Multimodal data organization, indexing and retrieval
- Latent space learning for multimodal data

- Multimodal approaches to detecting complex activities
- Multimodal approaches to event analysis and modeling
- Temporal or structural modeling for multimodal data
- Scalable processing and scalability issues in multimodal content analysis

Submission Guideline

Originality and impact on the society, as well as the innovative technical aspects of the proposed solutions, will be the major evaluation criteria. Authors should prepare their manuscript according to the TOMM Guide <http://tomm.acm.org/>, and add a comment in the email to the Assistant to the Editor-in-Chief that the submission is intended for the special issue on Learning from Multiple Social Networks. All papers will be reviewed by three external reviewers plus at least one guest editor.

Guest Editors

- Dr. Yan Yan, University of Trento, Italy (tom.yan.555@gmail.com);
- Dr. Liqiang Nie, National University of Singapore, Singapore, (nieliqiang@gmail.com);
- Prof. Rita Cucchiara, University of Modena and Reggio Emilia, Italy, (rita.cucchiara@unimore.it)

Important Date

- Initial papers submission due, ~~Oct. 31, 2016~~ **Extend to Nov. 25. 2016**
- Decision notification, Jan. 15, 2016
- Revision due, Feb. 15, 2017
- Acceptance notification, March 31, 2017
- Camera ready version due, April 15, 2017