Guest Editorial: Special Issue on Large Scale Multimedia Semantic Indexing

Nowadays visual big data has become the reality. Image and video information comprise a large portion of the traffic on Internet and wireless network. Huge volume of visual data is still accumulating in many industries, including entertainment, surveillance, and telecommunications, to name a few. The ability to extract accurate and actionable information from this data at large scale is a key component of successful big data activity. Among the large spectrum of challenges that the researchers in this field face, one of the most significant is the well-known semantic gap between human-understandable high-level semantics and machine-generated low-level features. Unlike the self-describing data as in other domains (such as textual data), visual data relies on sophisticated feature representations and algorithms for mining the semantics. Though recent years have witnessed plenty of research effort on driving for higher accuracy and improved efficiency of multimedia semantic indexing algorithms, there are still many open problems in this research field.

Moreover, considering the explosive increase of the volume and variation of the multimedia database, we envisage that the avalanche of visual data will inspire more data-driven semantic modeling algorithms and quickly advance the boundary of the techniques. On one hand, the availability of visual big data enables more robust estimation of the parameters in highly complex semantic models and makes semantics extraction much easier. Even the near point search or statistical correlation analysis will reveal useful facet of the visual semantics. On the other hand, many conventional algorithms are not obviously applicable to visual big data. It poses great challenges in tailoring machine learning algorithms in large scale, utilizing modern distributed cluster architecture (*e.g.*, Hadoop or Storm) towards improved algorithmic efficiency, and proposing new problem formulations based on the properties of visual big data.

This special issue presents an interesting collection of papers that propose new image/video tagging or classification techniques, scalable data indexing algorithms and some new methods under special problem settings (*e.g.*, cross-modal domain adaptation). Each paper included in this special issue went through a highly selective reviewing procedure of CVIU. The selected works are supposed to make significant contributions in improving either the *effectiveness* or *scalability* of modern multimedia semantic indexing systems.

The special issue begins with four papers on the latest large-scale similarity-based visual search problem. Zhang *et al.* propose the algorithm of *fitted spectral hashing* which is inspired by classic binary hashing scheme *spectral hashing*. The idea is to remove the over-restrictive assumptions on data distributions in *spectral hashing* by using Sigmoid function for data mapping. The work by Cheng *et al.* addresses the problem of learning compact hash code from multiple modalities. A novel semi-supervised multi-graph hashing framework is proposed, where the influence of each modality is adaptively modulated. Wang *et al.* rely on a generalized likelihood ratio analysis to handle the quantization of out-of-sample data, such that the proposed algorithm avoids the drawback of classic K-means quantization and outperforms several representative data hashing schemes for similarity-based search. Xie *et al.* demonstrate the effectiveness of performing affinity propagation on image graph in solving the problem of near-duplicate image search.

The special issue also includes several works in image tagging and image classification respectively. Among the three image tagging papers, Lin *et al.* attack the problem of tag completion (namely predict the missing textual tags of images that are not manually provided) and the other two works address the problem of refining existing image tags or predicting the tags of a new image. Several modern modeling tools are utilized such as *dual view linear sparse reconstructions* and *regularized latent Dirichlet allocation*. The following are three works on image classification. Zhu *et al.* develop a collaborative branch selection scheme to mitigate the error accumulation problem in hierarchical classification. Tang *et al.* formulate the visual dictionary learning as a graph partitioning problem. The idea in the work of Yan *et al.* is to combine both the global and local information towards an improved visual feature selection algorithm.

The reminder of the special issue collects the work by Habibian and Snoek, which studies the methodology of constructing an effective concept dictionary for semantic event detection in videos, and the work by Pereira and Vasconcelos, which proposes to use auxiliary text to regularize cross-modal domain adaptation.

We believe that this special issue serves as an excellent reference for recent advances in large-scale multimedia indexing techniques, and will stimulate more significant progress in this exciting research field.

We thank the generous help from CVIU journal managers and all the reviewers.

Guest Editors
Yadong Mu
AT&T Labs Research, U.S.A.
E-mail: myd@research.att.com

Yi Yang
The University of Queensland, Australia
E-mail: yiyang@cs.cmu.edu

Liangliang Cao,
IBM T.J. Watson Research Lab, U.S.A.
Email: liangliang.cao@us.ibm.com

Shuicheng Yan,
National University of Singapore, Singapore
E-mail: eleyans@nus.edu.sg

Qi Tian
University of Texas at San Antonio, U.S.A.
E-mail: qitian@cs.utsa.edu