CrossMark

EDITORIAL

Guest editorial: mobile visual tagging with mobile context

Shuqiang Jiang[1] · Liangliang Cao[2] · Jitao Sang[3] · Jiebo Luo[4] · Ramesh Jain[5]

Recently, the wide popularity of personal mobile devices has greatly changed our daily lives both in terms of human-to-human communication and human-to-computer information access. An important development under this circumstance is that the visual contents are playing more and more dominant roles. We take pictures using mobile phones every day; we can send image/video messages to our friends anywhere and anytime; and online social communities are flushed with images such as Twitter, Facebook and so on. How to satisfactorily utilize and manage such visual information presents a great challenge for visual information understanding technologies.

We are now seeing the rapid improvement of image recognition techniques with the recent development of deep feature learning, cross-media annotation, contextual information complementation, transfer learning and so on. However,

✉ Shuqiang Jiang
  sqjiang@ict.ac.cn

  Liangliang Cao
  liangliang.cao@gmail.com

  Jitao Sang
  jtsang@nlpr.ia.ac.cn

  Jiebo Luo
  jluo@cs.rochester.edu

  Ramesh Jain
  jain@ics.uci.edu

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

[2] Yahoo! Research, New York, USA

[3] Institute of Automation, Chinese Academy of Sciences, Beijing, China

[4] University of Rochester, Rochester, USA

[5] University of California, Irvine, USA

there is still much work to do to satisfy the requirements for different applications. In fact, mobile devices bring a large amount of contextual information, which can be useful clues to facilitate image annotation and tagging. Moreover, mobile context is enriched by application-specific information at two levels: one is internal contextual information which is intrinsically contained in the mobile devices, such as personal profile, stored textual/visual content, camera and other sensor's parameters. The other is external contextual information which can be easily acquired by the mobile devices, such as weather, geo-location and aural information. How to fully utilize these information is an interesting and promising research problem.

This special issue aims to seek innovative papers that exploit novel technologies and solutions from both industry and academia on how to recognize and tag images/videos with mobile contextual information. This special issue includes both papers of direct submission to the call for papers and extension of papers selected from 2015 International Conference on Internet Multimedia Computing and Service (ICIMCS). The extended conference papers have at least 30% improvement compared with the original paper.

The first article is a survey paper entitled "A survey on context-aware mobile visual recognition" by Min et al. This paper focuses on recent advances in context-aware mobile visual recognition and reviews-related work regarding different contextual information, recognition methods, recognition types, and various application scenarios. Various contextual information including the location, time and camera parameters from different sensors of mobile devices are introduced for mobile visual recognition. This paper discusses three types of recognition methods: classification-based methods, retrieval-based methods and tag propagation-based methods. This paper also proposes several open issues that need to be addressed in the future, including designing compact and

discriminative descriptors, effectively integrating content and contextual information and considering users intentions.

The paper entitled "Automatic group activity annotation for mobile videos" by Chaoyang Zhao et al. proposes an approach to annotate group activities for mobile videos. To extract contextual mobile information, this work uses three cues including activity duration time, individual action feature and information shared between person interactions. Then, these appearances and context cues are modeled with a structure learning framework. Furthermore, the group activity labels can be inferred especially for the situation with multiple group activities co-existing.

The paper "Automatic image annotation using fuzzy association rules and decision tree" by Zhixin Li et al. proposes an approach for automatic image annotation by integrating fuzzy association rules and decision tree. This work first obtains fuzzy feature vectors to describe the ambiguity and vagueness of images. Then fuzzy association rules are generated, which can capture correlations between low-level visual features and high level semantic concepts of images. This work uses the method of decision tree to reduce the unnecessary rules for computational efficiency.

The paper "Discovering discriminative patches for free-hand sketch analysis" by Ying Zheng et al. proposes a weak supervised learning method to discover the discriminative patches for different categories of free-hand sketches. First, this work randomly extracts many patches with multiple scales. Then pyramid histogram of oriented gradient is calculated to represent these patches. An iterative detection process is used to implement cluster merging and discriminative ranking, to find the most discriminative patches. Experimental results on public dataset demonstrate the effectiveness of the proposed method.

The paper "Flickr group recommendation with auxiliary information in heterogeneous information networks" by Yueyang Wang et al. proposes a method to combine auxiliary information with implicit user feedback for group recommendation. A nonnegative matrix factorization (NMF) method is used to model user–user similarity via visual features and heterogeneous information networks. By this way, the recommended groups for users are collaboratively combined with user feedback, image visual features, mobile contextual information and common sense knowledge.

The paper "Multi-modal tag localization for mobile video search" by Rui Zhang et al. proposes a multi-modal tag localization framework to learn visual, auditory, and semantic features with deep learning method, which can be used for automatic time-code-level tag generation and query-dependent video thumbnail selection. More specifically, the learned Fast R-CNN model can catch objects in visual frames and filter the irrelevant frames. Besides, the auditory and semantic information are combined together by the learned Word2Vec model. Then, different modalities are fused for mobile video search.

The paper "Semi-supervised image classification via nonnegative least-squares regression" by Wei-Ya Ren et al. proposes a graph construction method called nonnegative least-squares regression (NLSR) to improve the quality of the graph. The nonnegative constraint is included to eliminate subtractive combinations of coefficients and improve the sparsity of the graph. Both small Gaussian noise and sparse corrupted noise are considered to improve the robustness of the NLSR. Furthermore, weighted version of NLSR (WNLSR) is proposed to further eliminate 'bridge' edges.