Large-Scale Geosocial Multimedia

Rongrong Ji
Xiamen University

Yi Yang
University of Queensland

Nicu Sebe
University of Trento

Kiyoharu Aizawa
University of Tokyo

Liangliang Cao
IBM Watson Research Center

With the advance of the Web 2.0 era came an explosive growth of geographical multimedia data shared on social network websites such as Flickr, YouTube, Facebook, and Zooomr. In recent years, we have therefore witnessed two emerging research topics in our multimedia community: social computing and geocomputing. Rather than simply searching for and passively consuming multimedia content, users are now able to create and exchange their own media data (videos, images, music, blogs, and so on) for social interaction using geosocial-related multimedia repositories that closely mirror our physical world. Such engagement has revolutionized our social lives, and it underscores a transformation of the Web as fundamental as its birth. More and more social signals, together with their location context (such as GPS tags, location name identifications, and metadata), are now associated with plain visual signals.

Location-aware media description, modeling, learning, and recommendation in pervasive social media analytics have become a key focus of the recent research in computer vision, multimedia, and signal processing societies. A new breed of multimedia applications that incorporates image/video annotation, visual search, content mining and recommendation, and so on may revolutionize the field. Combined with the popularity of location-aware social multimedia, location context data makes traditionally challenging problems more tractable. For example, for the purposes of mobile location recognition, we can use GPS information to significantly prune large-scale image repositories to filter out irrelevant images with respect to a given query image.

This timely special issue brings together active researchers to share recent progress in this exciting area. We hope this special issue offers a forum for multimedia researchers all over the world to discuss their works and recent advances in recognition and mining of such geosocial multimedia. We seek to present and highlight the latest developments in large-scale multiple evidence-based learning for geosocial multimedia computing. With this issue, we also identify several key challenges and potential innovations regarding the recognition and mining of geosocial multimedia.

Summary of Articles

In "Toward Multiscreen Social TV with Geolocation-Aware Social Sense," Han Hu, Yonggang Wen, Huanbo Luan, Tat-Seng Chua, and Xuelong Li describe a multiscreen social TV system integrated with social media analysis. The authors argue that online social networks have been transforming the information propagation pattern, which posits significant challenges to traditional media outlets, such as TV. To embrace this trend, the article proposes a novel paradigm that combines the TV watching experience with real-time social responses via second-screen technologies. The work consists of two core components: a cloud-clone-based social TV system and a social sense system. Their social TV system proposes a cloud-clone-based video playout functionality. In addition, a proposed social sense system is implemented to crawl the social media data related to TV programs and mine the geolocation-aware public perception. This knowledge is transferred to TV audiences on a secondary device (such as a tablet) to enrich the TV watching experience via a simple operation. The authors built a real system over a private data center at Nanyang Technological University (NTU) in Singapore, and

1070-986X/14/$31.00/© 2014 IEEE Published by the IEEE Computer Society

One of the main driving forces for geographical-aware social multimedia is the prosperity of the user community.

an online video demo is available to help readers understand some of the system's salient features.

Weiqing Min, Bing-Kun Bao, and Changsheng Xu's article, "Multimodal Spatio-Temporal Theme Modeling for Landmark Analysis," focuses on landmark analysis. The authors propose a theme model that enables the differentiation among three kinds of landmark themes: temporal themes (events occurring at a specific moment), local themes (events characterizing at local landmarks), and general themes shared by most landmarks. Based on the model, they propose a framework that includes three modules: data preparation, theme modeling, and theme analysis. Landmark photos are first downloaded and preprocessed. Then, theme modeling is used to identify one of the three types of themes. Lastly, the authors analyze the discovered themes by further evaluating the landmark-specific and time-specific location distribution of themes.

Ling-Yu Duan, Jie Lin, Jie Chen, Tiejun Huang, and Wen Gao review the significant progress in MPEG standardization on Compact Descriptors for Visual Search (CDVS) in their article "Compact Descriptors for Visual Search." Searching for a specific object or scene has a variety of applications in large-scale media collections. With the ever-growing computational power on mobile devices, recent works have attempted to directly extract compact visual descriptors on local devices and send these descriptors as a query to the server. This can significantly reduce the network latency when transmitting queries as well as the computational cost of online or offline feature extraction from the huge amount of concurrent query images or database reference images on the server end. In particular, this will also positively impact privacy because compact descriptors are anonymous, whereas pictures are not. To address the interoperability issue between heterogeneous terminals and servers arising from using visual features to exchange query or reference information, the emerging MPEG CDVS standard has standardized technologies that enable the efficient and interoperable design of visual search applications. This article describes the competitive pipeline of compact feature extraction and visual search process developed by MPEG CDVS Ad-hoc Group.

Benchang Wei, Tao Guan, and Junqing Yu's article "Projected Residual Vector Quantization for ANN Search" describes a method for fast and accurate large-scale approximate nearest-neighbor (ANN) search in high dimensional spaces. The authors observe that most existing quantization methods discard the projection errors generated by the dimension reduction process and the quantization errors generated in quantization process, which inevitably decreases search accuracy. To alleviate the errors generated by both processes, the authors propose projected residual vector quantization (PRVQ) for ANN search by exploiting multiple stage quantization strategy and considering the projection errors in each stage of the quantization process. To improve the performance of PRVQ, they also design three simple optimization strategies: optimizing reduced dimensionality, optimizing each stage quantizer, and optimizing the encoding process. Although the authors verified the PRVQ algorithm's effectiveness by integrating it into a mobile landmark recognition system, it can be broadly used in variety of multimedia applications.

In "View-Based 3D Object Retrieval: Challenges and Approaches," Yue Gao and Qionghai Dai study view-based 3D object retrieval, which has become an emerging and important research topic as a result of the advances in mobile device cameras. This article first introduces 3D object retrieval with multiple views and then discusses the key procedures and challenges: view capture and representative view selection, feature extraction, and object matching using multiple views. The authors also describe the future directions of view-based 3D object retrieval, including the challenges in large-scale data, feature extraction, multiple view matching, multimodal data, and geolocation-based applications. View-based object analysis can be a valuable tool in geographical applications.

Future Directions

One of the main driving forces for geographical-aware social multimedia is the prosperity of the user community, which largely reduces the extensive costs in human labor required to produce or collect media content and metadata. However, the increase in data scale introduces an important research question. As a result of the sparsity and bias of ordinary data, existing approaches usually adopt parametric prediction models. Yet, it is uncertain whether parametric models are suitable for modeling the massive and ever-growing scale of user-generated content (UGC). In such scenarios, there will likely be increased research efforts looking to investigate unparametric models. The most promising application for geosocial multimedia lies in its integration with mobile research. Specifically, the geographical information and social cues can be deployed in numerous applications, such as vision-based city navigation and location-sensitive tourism recommendation.

One potential trend left unexploited lies in the analytics of social statistics from geosocial multimedia. Indeed, photos, videos, or even GPS tags could be regarded as signals reflecting users' social activities. Our inability to recognize user groups from their geosocial multimedia is still an open problem. Solving this problem has many potential applications, such as personalized tourism, friend, and commercial recommendations.

One feasible solution to handling the scalability issue is to distribute computing and services on the cloud. In this case, a key design issue lies in how to develop a parallel computing pipeline and distribute the computing structure. One possibility is to utilize the original geographical information to manage data storage. For instance, in building a near-duplicate visual matching system for location recognition, it is not always necessary to build the indexing model over the entire city. On the contrary, we might subdivide the geotagged photos into subregions by dividing data into geographical groups, for example, using either GPS or base station tags. Such an implementation could provide more precise and efficient indexing models. **MM**

One potential trend left unexploited lies in the analytics of social statistics from geosocial multimedia.

Rongrong Ji is a professor at Xiamen University, China. His research interests include social media analytics and visual search. Ji has a PhD in computer science from the Harbin Institute of Technology. Contact him at rrji@xmu.edu.cn.

Yi Yang is a DECRA fellow at the University of Queensland, Australia. His research interests include machine learning and its application to multimedia content analysis. Yang has a PhD in computer science from Zhejiang University. Contact him at yee.i.yang@gmail.com.

Nicu Sebe is a professor at the University of Trento, Italy. His research interests include multimedia information retrieval and human-computer interaction. Sebe has a PhD in computer science from Leiden University. Contact him at sebe@disi.unitn.it.

Kiyoharu Aizawa is a professor at the University of Tokyo. His research interests include multimedia lifelog. Aizawa has a DrEin electrical engineering from the University of Tokyo. Contact him at aizawa@hal.t.u-tokyo.ac.jp.

Liangliang Cao is a researcher at the IBM TJ Watson Research Center. His research interests include multimedia and computer vision. Cao has a PhD in electrical and computer engineering from the University of Illinois at Urbana–Champaign. Contact him at liangliang.cao@us.ibm.com.