NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S S E M I N A R

Title:	**SEMINAR THEME: FUTURE TRENDS IN SOCIAL (VISUAL) MEDIA ANALYSIS Talk 1: Scalable Mobile Visual Recognition - Opportunities and Challenges Talk 2: Attractiveness Computing in Multimedia Talk 3: Object Level Feature Extraction for Large-scale Image Retrieval Talk 4: Video Understanding Meets Deep Learning Talk 5: Satisfaction Prediction for Web Search Users**

Speaker:	Winston H. HSU Professor, National Taiwan University, Taiwan
	Toshihiko YAMASAKI Associate Professor, University of Tokyo, Japan
	GAO Ke Associate Professor, Institute of Computing Technology, Chinese Academy of Sciences, China
	MEI Tao Lead Researcher, Microsoft Research, China
	ZHANG Hanwang Research Fellow, National University of Singapore, Singapore

Date/Time:	4 December 2015, Friday, 01:45 PM to 05:45 PM
Venue:	Executive Classroom, COM2-04-02
Chaired by:	Dr Chua Tat Seng, KITHCT Chair Professor, School of Computing (chuats@comp.nus.edu.sg)

Talk 1: Scalable Mobile Visual Recognition - Opportunities and Challenges
Time: 14:00

Abstract:
The rapid development of technologies has made image/video analytics services possible on mobile devices such as smartphones and tablets; and strong needs for mobile visual search and recognition have emerged. So does the emergence for Internet of Things - potentially visual signals will communicate intensively between the embedded devices and the servers?

While many real applications requires a large-scale recognition system, the same

technologies that support server-based scalable visual recognition may not be feasible on mobile (or embedded) devices due to resource constraints. Although the client-server framework ensures the scalability, but the real-time response subjects to the limitation on network bandwidth. Therefore, the main challenge for mobile visual recognition system should be the recognition bitrate, which is the amount of data transmission under the same recognition performance.

In this talk, we will first review how the thousand-scale recognition systems have been developing and introduce a few strategies that we having been working on for mobile visual recognition - (1) label space compression for human attribute detection, (2) the brand-new paradigm in "compression for recognition," and (3) a few attempts for constructing mobile-based visual recognition systems. Then we will review the state-of-the-art (e.g., scalable recognition in deep convolutional neural network), demonstrate the exciting findings in the evaluations on the large-scale image dataset, and highlight certain opportunities for further investigations.

Biodata:
Winston is an active researcher dedicated to novel algorithms and systems for ultra-large-scale image/video retrieval, social media mining, and visual recognition. He is keen in advanced researches towards business deliverables via academia-industry collaborations and co-founding startups. He received several technical awards and best paper awards in multimedia research community.

Talk 2: Attractiveness Computing in Multimedia
Time: 14:30

Abstract:
Taking photos and videos and sharing them on social networks has become very popular and common, it does not necessarily mean that users are getting better at taking photos and videos, annotating them, and generating new contents from them than before. In fact, many users, including myself, are not satisfied with the quality of the media data they have created. And this is one of the reasons why data cleansing is needed before big multimedia data analysis.

We have been investigating three types navigation systems to tackle this issue: (1) navigation of user activity design: navigating users where to go and what to do for taking good photos and for having good experiences based on their personal preferences, (2) navigation of multimedia item generation: navigating users to take good photos and videos and annotate with good text tags, and (3) navigation of multimedia content generation: navigating users to aggregate multimedia items such as images, video clips, texts, etc., and to generate high quality multimedia content such as videos and presentations.

We expect that these navigation systems would increase the users' satisfaction, as well as improve the quality of the big multimedia data uploaded to the Internet. In this paper, we would like to review some of our representative work.

Biodata:
Toshihiko received the B.S. degree, the M.S. degree, and the Ph.D. degree from The

University of Tokyo in 1999, 2001, and 2004, respectively. He is currently an Associate Professor at Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo. He was a JSPS Fellow for Research Abroad and a visiting scientist at Cornell University from Feb. 2011 to Feb. 2013.

Toshihiko's current research interests include multimedia big data analysis, pattern recognition, and machine learning, and so on. His publication includes three book chapters, more than 50 journal papers, more than 150 international conference papers, more than 420 domestic conference papers. He has received around 30 awards including Geomm 2013 Best Paper Award, ICMLA 2012 Best Special Session Paper Award, IEICE Young Researcher Award, so on.

Talk 3: Object Level Feature Extraction for Large-scale Image Retrieval
Time: 15:00

Abstract:
A basic problem of large-scale content-based image retrieval is to extract image features effectively and efficiently from suitable 'visual units'. The ideal feature representation should be robust, distinctive and concise. Recent years have witnessed the change of visual units from pixels, superpixels, local salient structures, to object level regions. This new trend is motivated by the increasing achievements of generic object proposal, object boundary detection and bottom-up saliency analysis.

In this talk, I will give a brief review of different visual unit detection and description strategies in large-scale image retrieval, and introduce several important cues measuring object characteristics proposed these years. Based on this, I will share our recent research progress in object level feature extraction.

Biodata:
Gao Ke received the M.S. degree and Ph.D. degree from Institute of Computing Technology, Chinese Academy of Sciences. She is currently an Associate Professor in Chinese Academy of Sciences. Her research interest includes content-based image/video retrieval, mobile visual search--especially focus on local feature extraction and indexing. She has published over 30 papers in the areas of her research interests, including CVPR, ACM MM and IEEE Trans on MM. She received the Beijing Science and Technology Award in 2014, the Academic Stars Award in ICT-CAS in 2012, and won the first prize in CBCD task in TRECVID in 2009.

Tea Break

Talk 4: Video Understanding Meets Deep Learning
Time: 16:00

Abstract:
The recent advances in deep learning have boosted the research on video analysis. For

example, convolutional neural networks have demonstrated the superiority on modeling high-level visual concepts, while recurrent neural networks have been proven to be good at modeling mid-level temporal dynamics in the video data.

We will present a few recent advances for understanding video content using deep learning techniques. Specifically, this talk will focus on: 1) translating video to sentence with joint embedding and translation, which achieves the best to-date performance in this nascent vision task, 2) first-person video highlight extraction with a pairwise deep ranking model, and 3) action recognition with a multi-granular spatiotemporal architecture which achieved rank 2 in CVPR THUMOS 2015 video classification challenge.

Biodata:
Tao is a Lead Researcher with Microsoft Research, Beijing, China. His current research interests include multimedia analysis and retrieval, and computer vision (video processing, analysis, and understanding). In particular, he is interested in applying the techniques from these areas to a broad range of multimedia applications, such as video analytics, personal media, Web search, social and mobile multimedia applications. He has shipped his inventions and technologies to Microsoft products, such as Bing, Office, MSN, One Drive, etc.

Receiving the B.E. degree in automation and the PhD degree in pattern recognition and intelligent systems from the University of Science and Technology of China in Hefei, China, Tao has authored or co-authored over 150 papers in journals and conferences and holds 13 US-granted patents. He was the recipient of several paper awards from prestigious multimedia journals and conferences, including the IEEE T-CSVT Best Paper Award in 2014, the IEEE TMM Prize Paper Award in 2013, and the Best Paper Awards at ACM Multimedia in 2009 and 2007, etc. Tao was also the recipient of Microsoft Gold Star Award in 2010, and Microsoft Technology Transfer Awards in 2010 and 2012.

Talk 5: Satisfaction Prediction for Web Search Users
Time: 16:30

Abstract:
Traditional feature learning requires a considerable amount of well-annotated data (e.g., ImageNet, NUSWIDE and CCV), whose construction per se is expensive and time-consuming. Unfortunately, these data hardly keep up with the ever-evolving trends in multimedia applications, such as the target domain shift and novel semantic concepts.

In this talk, I will share our recent research progress in learning features from collective intelligence, which is natively collected from the inexhaustible Web user-generated contents and behaviors like Facebook "like", Google "click" and Pinterest "pin". In fact, our research is a more aggressive and practical implementation of weakly-supervised and unsupervised learning. We will explore several interesting tasks on how to discover meaningful semantics from user behaviors and try to find the underlying rationales.

Biodata:
Hanwang is a research fellow of NExT Research Centre, a joint center between National University of Singapore and Tsinghua University to develop technologies for live media

search. His research interest includes multimedia and computer vision -- developing techniques for efficient search and recognition in visual contents. He received the Best Demo runner-up award in ACM MM 2012 and the Best Student Paper award in ACM MM 2013. He is the winner of Best PhD Thesis Award of School of Computing, NUS in 2014.

Closing

Refreshment