This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

Towards Multi-Screen Social TV with Geo-Aware Social Sense

Han Hu, Yonggang Wen, Huanbo Luan, Tat-Seng Chua, and Xuelong Li, *Fellow, IEEE,*

***Abstract*—The increasing popularity of social interactions and geo-tagged user generated contents (UGC) has been transforming the television (TV) viewing experience, from a "laid-back" video watching behavior into a "lean-forward" social-engaged experience. This article envisions a multi-screen social TV integrated with social sense via a second screen as a novel paradigm for content consumption. This new application is built upon our cloud-centric media platform, which provides on-demand virtual machines for content platform services, including media distribution, storage, processing, etc. The media platform is also integrated with our big-data social platform (i.e., Social Sense), which crawls and mines social data in the context of contents. Specifically, this new social TV consists of three key subsystems, including interactive TV, social sense, and multi-screen orchestration. Interactive TV implements cloud-based social TV system, offering rich social features; social sense discovers the geo-location-aware public perception and knowledge related to the media content; and multi-screen orchestration provides an intuitive and user-friendly human-computer interface to combine the above two subsystems, fusing the TV viewing experience with social perception. We have built a proof-of-concept demo over a private cloud at NTU. The feature verification and performance comparison demonstrate the feasibility and effectiveness of our innovation in transforming TV viewing experience.**

I. Introduction

In recent years, we have witnessed the explosive growth in social networks and portable devices, which have revolutionized the way information is created, disseminated and consumed. This innovative evolution posits significant challenges to the conventional information propagation carriers, especially for the television (TV) landscape. In the traditional view, TV only provides consumers with limited optional produced programs. What the customers can do, is deciding whether to receive those programs in specific time duration or not. In contrast, social networks together with portable devices offer new media forms (e.g., UGC, instant messaging, online entertainments) and media outlets (e.g., smart phones and tablets). More and more research from both academia and industry is working to integrate these delightful features into TV applications.

New paradigms, such as interactive TV [1] and social TV [2], have been proposed to provide social and interactive features simultaneously with TV viewing experience. Interactive TV combines interaction to screen media, including on-demand delivery of TV programs, as well as some value-added services, like on-line shopping, banking, and so forth. While social TV integrates social networking, like Facebook and Twitter, into the TV environment, allowing remote viewers to interact with each other via the TV set.

With the wide spread use of portable devices, mobile applications [3][4] have been considered to satisfy personalized requirements. Recently, people's video watching behavior has been dramatically changed, from one screen to multi-screens (also known as second screens) either by sequential usage or simultaneous usage [5]. Multi-screen technology provides a parallel companion device, whereby the user can engage in some other activities without interrupting the video watching. Moreover, other solutions attempt to make TV viewing experience ubiquitous and transferable among different devices to realize TV everywhere [6]. Virtualized screen solution [7] has been proposed to achieve synchronized contents among distinctive terminals. In this system, screen rendering is done in the cloud, and delivered as images to the client for interactive display.

All these works have greatly enriched traditional TV viewing experience either by extending interaction or multi-screen fusion features. However, none of them incorporates traditional TV media with the fast growing geo-location social media, which is becoming the most important way that people get information. To lessen the technology gap, we design and implement a multi-screen interactive TV system that enhances the watching experience by providing the social response. In particular, we propose a generic cloud-centric media platform, which encapsulates a set of media storage and processing, big data analytics tools. Leveraging the platform, we build the interactive TV system and the social sense system. The interactive TV system integrates enriched social features, including text/video chatting, video comments, and watching together. While the social sense system leverages a big data analytics paradigm to mine social data associated with the ongoing program to provide deeper insights. Following that, we design a multi-screen fusion scheme that can transfer the ongoing program related social sense to a second screen, using an intuitive human-computer interaction technique. Our prototype system has been implemented on top of a private cloud at NTU and NUS to demonstrate the concept and evaluate its performance. The feature verification and performance evaluation presented in this paper illustrate the intrinsic social features, deep analytics for media contents, and intuitive operations for users.

H. Hu, H. Luan and T.-S. Chua are with the School of Computing, National University of Singapore, Singapore 117417. e-mail: {huh, luanhb, chuats}@comp.nus.edu.sg. H. Hu is visiting NTU for this work.

Y. Wen is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798. e-mail: ygwen@ntu.edu.sg.

X. Li is with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China. e-mail: xuelong_li@opt.ac.cn.

Xuelong Li is the corresponding author.

Digital Object Indentifier 10.1109/MMUL.2014.2 1070-986X/$26.00 2014 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

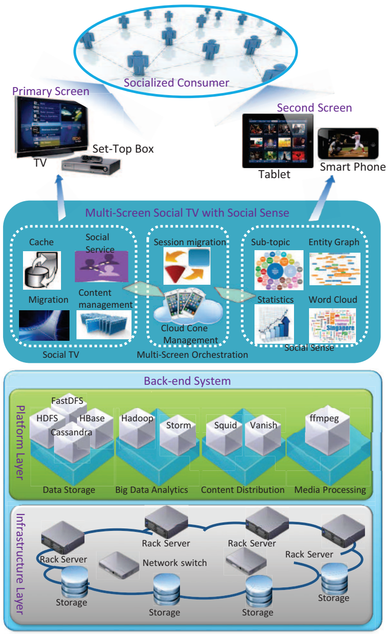


Fig. 1: Architecture of data driven social TV system

The rest of this article is organized as follows. First, we present the overall system architecture for the multi-screen social TV system. Next, we describe the actual implementation on a proof of concept demo, which highlights social TV, social sense, and multi-screen orchestration subsystems. Following that, we provide some representative features, as well as performance evaluation. Finally, we summarize this article.

II. Data-Driven Social TV Architecture

To provide scalability and cost efficiency to the digital TV value chain, we propose a generic layered system architecture for our multi-screen social TV system with social sense, as illustrated in Fig. 1. The anatomy of this proposed architecture consists of three fundamental components, i.e., a back-end system, a multi-screen social TV system with social sense, and consumers, from bottom to up. Each of them is explained as follows:

A. Back-End System

The back-end system leverages two cloud service models, i.e., infrastructure as a service (IaaS), and platform as a service (PaaS), to support upper layer applications.

- In the infrastructure layer, raw ICT resources (e.g., CPU, storage, bandwidth, etc.) are abstracted into a resource pool, and provided in the form of virtual machines (VMs) via virtualization techniques. The capacity of VMs can be dynamically tailored in a fine granularity to adapt to the resource demand, while maintaining the service-level agreement (SLA).
- In the platform layer, we encapsulate a collection of programming tools on top of the infrastructure. In our system's context, we classify essential media specific services into four categories, including content distribution, data storage, media processing, and big data analytics.

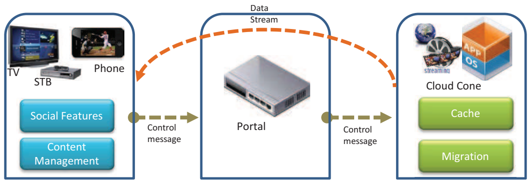


Fig. 2: Work flow of interactive TV

Other possible services can also be integrated into this layer. These offerings greatly facilitate the development and deployment of our upper layer applications.

B. Multi-Screen Social TV with Social Sense

Multi-screen TV system with social sense comprises three complementary subsystems, i.e., a interactive TV subsystem, a social sense subsystem, and a multi-screen orchestration subsystem.

- Interactive TV subsystem implements fundamental TV playback, as well as intrinsic social and interactive features. Our system can accept all possible content sources and integrates with popular social networks, such as Google+, Facebook, to provide online chatting, commenting, group-watching functions, etc.
- Social sense subsystem crawls social media data associated with TV programs, and analyzes them to mine social phenomena, sense, influences, etc. Our system can provide profound analysis results, including the extent of social discussion, sub-topics, trends, and entity sense.
- Multi-screen orchestration incorporates TV viewing experience with the social sense via a second screen. Specifically, social sense related to the ongoing program can be displayed on second-screen, with the minimum human intervention.

C. Consumers

Consumers refer to viewers that use end devices (e.g., TV, laptop, tablet, smartphone) for video consumption. In this paper, we consider the context that audience is watching TV programs, and browsing the related geo-social observation via a second screen at the same time. However, our system architecture is not limited to any specific media outlet that exists or will arise in the future.

III. System Prototype

In this section, we highlight the multi-screen social TV system with social sense in our proposed generic system framework by introducing three key components.

A. Interactive TV

Fig. 2 presents three participatory segments of the interactive TV subsystem: media outlets, portal, and cloud clone. In our system, a cloud clone refers to the virtual machine that manages all the user's related information, including distinct devices, session, authentication, etc. We instantiate a cloud clone for each user when he requests a session. In particular,

Digital Object Indentifier 10.1109/MMUL.2014.2 1070-986X/$26.00 2014 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

a cloud clone serves as his proxy, which is responsible of fetching media content, transcoding them for different outlets, and status synchronization. A typical TV session works as follows: first, audience requests a TV program through media outlets; second, portal accepts the request and creates or migrates a cloud clone for this user; third, the cloud clone parses the request, fetches the demand media content and then sends the content to media outlets via portal. To accomplish this procedure, we customize our solutions in media outlets and cloud clone, the key components of which are detailed in the following.

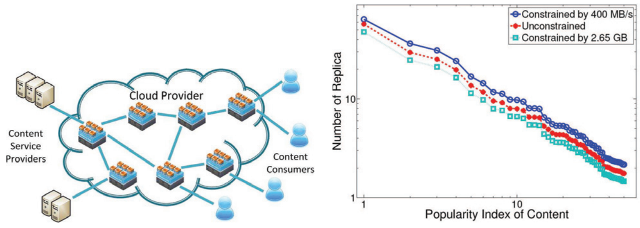
1) Media Outlets: Our system enables audience to watch TV programs either with TV device or other portable devices, such as tablet or smartphone. In particular, we implement two key components, i.e., content management and social feature. Content management module allows users to operate four categories of content resources, including local contents, media sharing center contents, OTT online contents, and the contents stored in the cloud. Social feature module integrates our system with Facebook and implements an internal instant messaging mechanism for text/video chatting and commenting ongoing videos.

2) Cloud Clone: Cloud clone is desired to serve large scale audience from geo-location distributed areas with varying network environments. There are two obstacles to achieve this goal. One is to distribute the rapid growing UGCs with a long tailed nature. This is a challenging issue as request population, content popularity, and resource capacity are all geo-location diverse. The second challenge is caused by the watching behavior in that users will choose different media outlets with distinct bandwidth requirements, thus cloud clone should migrate along the transmission path to save bandwidth consumption. These two problems are tackled as follows:

- *Geo-aware content cache:* Content caching distributes media contents into different cloud edge servers to reduce the distance between users and contents, leading to an improved end user experience. The system architecture is illustrated in Fig. 3(a). However, the decision on choosing the number of replicas and their locations, is essential to the generated operation cost. It involves a fundamental tradeoff. On one hand, more replicas can reduce the distance between user and content, which can translate into networking cost. On the other hand, if too many copies are placed and their locations are not proper, significant storage cost can incur with limited gains on networking cost reduction. This problem can be formulated as a constrained graph optimization problems as follows [8]. Give a topology $G = (V, E)$, where V is the set of edge servers, and E is the set of links between them. The goal of content caching is transformed to computing a subset of n vertices that minimize the mean hop distance between any two nodes for a given content. The numerical verifications (see Fig. 3(b)) in a real deployed network, i.e., US IP backbone network, suggest that the optimal replica number follows a power-law distribution to its popularity.
- *Geo-based cloud clone migration:* In our multi-screen system, cloud clone fetches contents from the afore-mentioned caching system, transcodes them to demanded bitrate, and transfers them to media outlets. The location of cloud clone is within the transmission path as shown in Fig. 4(a). The original bitrate of media content is B_0, while the transcoded bitrate for media outlets is B_t, determined by the type of active end device. When $B_0 < B_t$, placing the cloud clone to the nearest node to user can minimize the transmission cost; otherwise, the furthest node is preferred. However, cloud clone migration incurs extra transmission overhead. The challenge is to find an optimal scheduling policy over time to balance the transmission cost and the migration cost. The problem can be formulated as [9]:

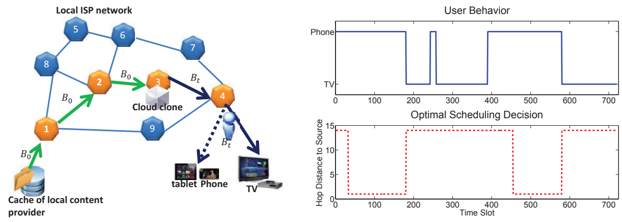
$$\min \sum_{t=0}^{T-1} (C_{tr}(l_k(t)) + C_{mig}(l_k(t+1), l_k(t))),$$

where, $l_k(t)$ is the location of cloud clone at time t, $C_{tr}(l_k(t))$ is the transmission cost, and $C_{mig}(l_k(t+1), l_k(t))$ is the migration cost determined by the location distance at successive time. We utilize Markov Decision Process (MDP) to solve the user behavior driven optimal location strategy. The result, as illustrated in Fig. 4(b), suggests that the cloud clone only can be located at the nearest or the furthest node to user, regardless of the delivery path length.



(a) System architecture (b) Optimal content replica number

Fig. 3: Geo-location aware content cache



(a) Cloud clone migration (b) Optimal migration policy

Fig. 4: User behavior driven cloud clone migration

B. Social Sense

Microblog services provide an essential platform for users to share everyday thoughts, opinions, and experiences. Parts of these UGCs reflect and reveal their interests, concerns and criticisms about TV programs. The aim of geo-location social sense is to associate the public perception with ongoing TV programs. The system can be decomposed into two stages: relevant data collection, and emerging characteristic modeling of detected topics [10].

Digital Object Indentifier 10.1109/MMUL.2014.2 1070-986X/$26.00 2014 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

To mine the social sense from UGCs, the first step is to crawl a relatively complete set of messages associated with the designated media content. This is not a trivial task as most of the live microblog services set limits on the amount and frequency of data that can be crawled. Besides, because of the size limit on microblog messages, many related messages do not contain the expected keywords, while the relevant ratio to specific media content is usually quite low. To tackle this problem, we design four types of crawlers, i.e., fixed keywords, dynamic keywords, known accounts, and key users. Fixed keywords are first manually selected to uniquely identify the media content. Similarly to fixed keywords, known accounts are manually selected to identify a set of media content related users, such as the official account, the director of the ongoing program, etc. Dynamic keywords and key users are extracted from the tweets sets crawled by fix keywords and known accounts.

The second step employs machine learning, text and image analytics [11][12][13] techniques to discover knowledge from the data, such as media context, geo-location, and key users. An online or incremental clustering algorithm is first used to discover topics to guarantee the real-time performance. Then we analyze the emerging topic-related features, including user authority, tweets influence, and emerging keywords. These features are incorporated into a topic learner to identify the emerging topics in a timely manner. The social sense is determined by providing the analysis results in terms of statistics of crawling data, emerging topics, key entities (organization, person, location, and misc), and trending words.

C. Multi-Screen Orchestration

The general idea of multi-screen orchestration is to build a link between TV programs and their social sense. This feature allows consumers to simply shot the quick response (QR) code on the TV screen using the camera of their second screens to obtain the related social sense. Two main components are involved, including cloud clone management and session migration.

Since each user corresponds to a unique cloud clone, the profiles and status of all the application running on media outlets can be orchestrated by cloud clone via the inter-device message bus. As a result, we can easily search, synchronize status of distinct devices belonging to the same user. In order to achieve fast routing and information retrieval, all the cloud clones form a logical ring via the distributed hash table (DHT). In the DHT key space, each cloud clone is uniquely determined by a key. As a result, the route length can be limited at $O(\log n)$, where n is the total number of nodes in the DHT ring.

To start a session migration, users are required to login to the cloud to get authentication. Upon the confirmation, users shot QR code on the TV screen, and send the session migration request to the cloud. Based on this request, the cloud clone will process to recognize the sessions and make confirmations accordingly. After the users receive this confirmation message, they can eventually trigger the session migration, and obtain the social sense data.

The advantages of this scheme are multi-fold. First, our scheme supports session migration without recalling and typing password. User only needs to register all their terminals and bind them together in the beginning. Second, users often have difficulties in understanding new applications. Our scheme offers users a simple and intuitive procedure of how it works and how to user it, which should help to improve the learning-curve in practice. Finally, our scheme provides a generic link between TV and mobile terminal, which is compatible to other applications.

IV. Feature Verification and Performance Evaluation

In this section, we first demonstrate the salient features of our system. Following that, a subjective evaluation is conducted to illustrate the operation superiority of our multi-screen orchestration scenario.

A. Testbed Implementation

We build our system on top of a modular data center at NTU, which consists of 10 racks. Each rack contains up to 30 HP servers and 2 Gigabit Cisco switches. The data center can provide ICT capacity of 25 TB disk, 1200 GB memory, and 600 CPUs. We utilize CloudStack on CentOS 6.3, the host OS, to construct the infrastructure layer, which virtualizes physical servers into a collection of virtual machines. Based on our infrastructure layer, we encapsulate a set of data storage, indexing, and processing tools, including Hadoop suite, FastDFS, Cassandra, SOLR, Storm, and ffmpeg, to form the platform layer. The social sense platform locates at NUS.

B. Feature Verification

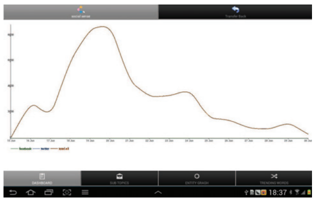
1) Social sense: Social sense aims to discover the knowledge about the media contents in the context of social media. Such information is determined by identifying who are saying what about the media contents, and where they are. Our system provides four aspects of social sense, including topics, degree of interest, entity graph, and word cloud, as shown in Fig. 5. Degree of interest refers to the potential interest and popularity of the subject or topics, as reflected by the message count. From Fig. 5(a) and Fig. 5(b), we can easily see the emerging topics, and the corresponding degree of interest. Entity graph (see Fig. 5(c)) presents who are talking about the media contents, in terms of person, organization, location and misc. Keyword cloud (see Fig. 5(d)) shows the set of keywords extracted from crawled messages and their weight.

An example content is a famous Singapore film, “Ah boys to men”. Users can see all the topics about this film from Fig. 5(a), including “Song theme my brother”, “time Singapore film”, etc. He can click one of the topic “time Singapore film” in the topic cloud to see all the related messages, as shown in Fig. 6(a). Moreover, all the relevant persons, organizations, and locations about this film are graphed in Fig. 6(b), which can be accessed through the link in entity graph. The interested reader is referred to the attached video demo.

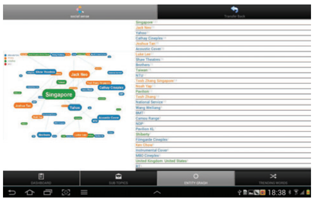
Digital Object Indentifier 10.1109/MMUL.2014.2 1070-986X/$26.00 2014 IEEE

This article has been accepted for publication in IEEE MultiMedia but has not yet been fully edited.
Some content may change prior to final publication.

(a) Topics



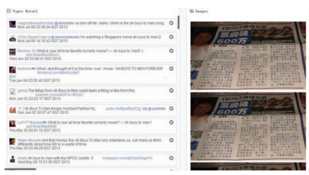
(b) Degree of interest



(c) Entity graph

(d) Keyword cloud

Fig. 5: Social sense features

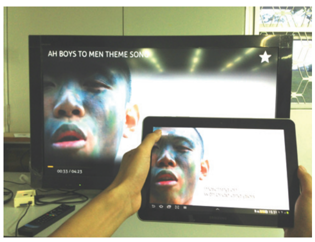


(a) Tweets related to topic "time Singapore film"

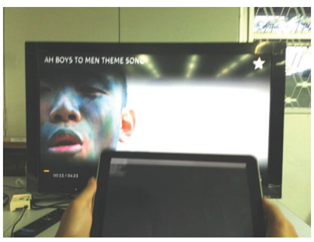
(b) Entity related to the film, including person, organization, location, and misc

Fig. 6: Social sense of a famous Singapore local film, "Ah Boys to Men"

(a) Screen shot



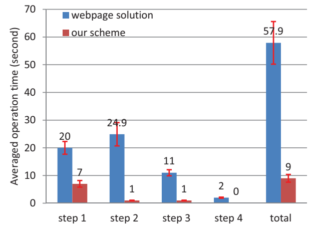
(b) Interest indication



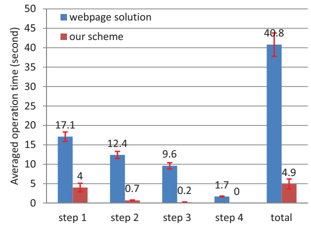
(c) Gesture triggering

(d) Social sense acquisition

Fig. 7: Different steps to complete a session migration



(a) First Time



(b) Tenth Time

Fig. 8: Averaged operation time

2) Multi-screen orchestration: Fig. 7 shows the intuitive multi-screen orchestration process, which consists of four steps. First, user holds his mobile phone/tablet in front of the TV, to obtain the authentication details through the QR code (Fig. 7(a)). He then presses the "sense" icon to select the social sense information (Fig. 7(b)). Once the cloud is ready for social sense transmission, he triggers the process by performing a flipping-in gesture (Fig. 7(c)). After those simple steps, the social sense information is displayed on his mobile phone/tablet (Fig. 7(d)).

C. Subjective Evaluation

TABLE I: Operation procedure comparison for webpage solution and our scheme

	Webpage solution	Our scheme
step 1	go to the webpage	screen shot
step 2	log in	interest indication
step 3	search the content	gesture triggering
step 4	choose the social sense	social sense transmission

As described in the last section, our multi-screen orchestration scheme possesses several benefits, like user-friendly, simple & intuitive, etc. In this research, we conducted a user study to compare our scheme with the traditional webpage solution in terms of retrieval time. Less retrieval time indicates a smoother operation procedure resulting in better user experience, a shorter learning curve, and the operation success ratio. In the webpage solution, participants access the social sense webpage, login to the accounts using passwords, search the media content, and choose the expected social sense information. The comparison of the operation procedures of the two solutions is shown in Table I. We recruited 15 participants from NTU students with major in computer science. Seven participants were female, and eight were male. The experiment procedure was as follows: we first asked participants to acquire social sense information using the webpage solution, and recorded the operation time for each step separately. We then requested them to repeat this procedure for ten times. Following that, the same flow is conducted for our multi-screen orchestration scheme.

Fig. 8 shows the averaged retrieval time and their standard variation for the two solutions for the first and tenth times of access. In the beginning, all the participants are unacquainted with the two solutions. We compared the operation time to show the learning curve. As shown in Fig. 8(a), in total,

Digital Object Indentifier 10.1109/MMUL.2014.2 1070-986X/$26.00 2014 IEEE

the first operation time is 57.9 seconds for the webpage solution, and 9 seconds for our scheme. It means that our scheme is more user-friendly, allowing user climbing a deep learning curve faster. After ten times, all the participants get familiar with the operation procedure. The operation time is 40.8 seconds for the webpage solution, and 4.9 seconds for our scheme, as shown in Fig. 8(b). Clearly, our scheme outperforms the webpage solution with shorter operation time. In particular, we can see that the majority time is consumed in the step (1) and step (2) for both solutions. However, our scheme exploits the QR code scanning to replace the typing keyboard operations, resulting in less operation time.

V. Summary

In this paper, we proposed a cloud based multi-screen social TV system, that enriches the TV viewing experience with more insightful geo-location aware social perception. We first provided an overview on our system framework. We then introduced subsystems of our system, including interactive TV, social sense, and multi-screen orchestration. Interactive TV subsystem addresses the geo-location related content caching and cloud clone migration to deliver media content to viewers with guaranteed quality of experience (QoE); and the social sense subsystem discovers geo-aware social perception, which will be displayed with the ongoing program synchronously via the multi-screen orchestration subsystem. A proof-of-concept demo was built on top of a private cloud at NTU. Feature verification and user experience study were provided to demonstrate the superiority of the proposed system.

Acknowledgment

This research is partially supported by Singapore MOE tier-1 fund (RG 11/31) and NTU SUG, the Singapore National Research Foundation under its International Research Center @ Singapore Funding Initiative and administered by the IDM Programme Office, by the National Basic Research Program of China (973 Program) (Grant No. 2012CB316400), by the National Natural Science Foundation of China (Grant No. 61125106), and by Shaanxi Key Innovation Team of Science and Technology (Grant No. 2012KCT-04).

References

[1] M. Ursu, M. Thomas, I. Kegel, D. Williams, M. Tuomola, I. Lindstedt, T. Wright, A. Leurdijk, V. Zsombori, J. Sussner *et al.*, "Interactive tv narratives: Opportunities, progress, and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 4, no. 4, pp. 25:1–25:39, 2008.

[2] P. Cesar and D. Geerts, "Past, present, and future of social tv: A categorization," in *2011 IEEE Consumer Communications and Networking Conference (CCNC)*. IEEE, 2011, pp. 347–351.

[3] R. Ji, L.-Y. Duan, J. Chen, H. Yao, J. Yuan, Y. Rui, and W. Gao, "Location discriminative vocabulary coding for mobile landmark search," *International Journal of Computer Vision*, vol. 96, no. 3, pp. 290–314, 2012.

[4] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua, and R. Jain, "W2go: a travel guidance system by automatic landmark ranking," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 123–132.

[5] Google, Ipsos, and Sterling, "The new multi-screen world: Understanding cross-platform consumer behavior," [Online], http://services.google.com/fh/files/misc/multiscreenworld_final.pdf, August 2012.

[6] R. Martin, A. L. Santos, M. Shafran, H. Holtzman, and M.-J. Montpetit, "neXtream: a multi-device, social approach to video content consumption," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*. IEEE, 2010, pp. 1–5.

[7] Y. Lu, S. Li, and H. Shen, "Virtualized screen: A third element for cloud mobile convergence," *IEEE Multimedia Magazine*, vol. 18, no. 2, pp. 4–11, 2011.

[8] Y. Jin, Y. Wen, K. Guan, D. Kilper, and H. Xie, "Toward monetary cost effective content placement in cloud centric media network," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[9] Y. Jin, Y. Wen, and H. Hu, "Minimizing monetary cost via cloud clone migration in multi-screen cloud social tv system," in *Proceedings of IEEE Global Communications Conference (Globecom)*. IEEE, 2013, accepted for publication.

[10] Y. Chen, H. Amiri, Z. Li, and T.-S. Chua, "Emerging topic detection for organizations from microblogs," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013, pp. 43–52.

[11] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu, "Visual-textual joint relevance learning for tag-based social image search," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 363–376, 2013.

[12] R. Ji, H. Yao, W. Liu, X. Sun, and Q. Tian, "Task-dependent visual-codebook compression," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 2282–2293, 2012.

[13] R. Ji, L.-Y. Duan, J. Chen, L. Xie, H. Yao, and W. Gao, "Learning to distribute vocabulary indexing for scalable visual search," *Multimedia, IEEE Transactions on*, vol. 15, no. 1, pp. 153–166, 2013.

Han Hu received the B.E. and Ph.D. degrees from the Department of Automation, University of Science and Technology of China, Hefei, China, in 2007 and 2012 respectively. He is currently a research fellow with the School of Computing, National University of Singapore. His research interests include social media distribution and analysis.

Yonggang Wen is an Assistant Professor in the School of Computer Engineering at Nanyang Technological University (NTU) in Singapore. He received his Ph.D degree in electrical engineering and computer science from Massachusetts Institute of Technology (MIT) in 2008. His research interests are cloud computing, content networking and green networks.

Huanbo Luan received the B.E. degree from Shandong University, Shandong, China, in 2003, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2008. He is currently a Senior Research Fellow with the School of Computing, National University of Singapore. His current research interests include multimedia analysis, mining, and retrieval.

Tat-Seng Chua is a Full Professor with the School of Computing, National University of Singapore. In 1998−2000, he was the Acting and Founding Dean of the School of Computing. His main research interests include multimedia information processing, in particular on the extraction, retrieval, and questioning-answering of text and video information.

Xuelong Li (M'02-SM'07-F'12) is a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China.