MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 8, No. 1, January 2013

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from MMTC Chair

Dear MMTC colleagues:

It is really a great pleasure for me to serve as the Europe vice-chair for this vital ComSoc Committee during the period 2012-2014! As part of my duties, I have contributed to the initial setting of the Interest Groups (IGs) and I am starting to work on the policies for the setting of our workshops and conferences.

Concerning our IGs, I really believe that these represent the core of our networking and scientific activities and I warmly invite all of you to select one or more IG(s) to get involved by contacting the chair(s) so as to take part as key member. The activities of the IGs include, among others, the organization of workshops, sessions and conferences with the involvement of the MMTC, the editing of special issues in major IEEE journals, the setting of invited talks trough conference calls that can be of interest for our community and the rest of the ComSoc members. While these are the major activities, some others can be carried out following the specific IG topics, such as the contribution to standardization activities.

As to the policies for the organization and endorsement of technical events, a special case is represented by our MMCOM workshop. This is held in conjunction with Globecom, focusing every year on a specific hot topic related to the multimedia communications. It was held the first time in 2011 around the theme of green wireless multimedia communications, while in 2012 we had the second edition in Anaheim around the topic of quality of experience. The organization of a workshop focused on a well-defined theme is demonstrating to be successful as this year we had an average of 35 colleagues attending the technical sessions with peaks of 50 attendants, which represent significant numbers for a Globecom workshop. We will be leveraging on this success, by continuing our efforts in this direction so that every year a couple of IGs will be leading the organization around a selected hot topic. The MMCOM will be the only workshop endorsed by MMTC during Globecom. As to ICC and ICME conferences, we will promote the organization of co-located workshops proposed by our IGs. Any initiative coming from the MMTC community through the interest groups, and only these ones, will be endorsed by MMTC. Another different case is the support of conferences and workshops which are not held in conjunction with these conferences. In this case, MMTC provides the endorsement if the organizing committee includes at least one active MMTC member (officers and/or board members) and three MMTC members are in the program committee. The rationale behind these rules is to encourage the active participation of the colleagues to the MMTC community through the relevant IGs.

I would like to thank all the IG chairs and co-chairs for the work they have already done and will be doing for the success of MMTC and hope that any of you will find the proper IG of interest to get involved in our community!

Luigi Atzori
Europe Vice-Chair of Multimedia Communications TC of IEEE ComSoc

EMERGING TOPICS: SPECIAL ISSUE ON MULTIMEDIA AND CLOUD COMPUTING

Multimedia and Cloud Computing

Guest Editors: Zongpeng Li, University of Calgary, zongpeng@ucalgary.ca

Chuan Wu, The University of Hong Kong, cwu@cs.hku.hk

Cloud computing has proliferated as a new computation and resource provisioning paradigm in today's Internet. Featured by its on-demand, scalable resource provision, cloud computing has been exploited for a wide range of multimedia applications, including adaptive and energy-efficient multimedia computing, storage and transmission, both to the Internet users and mobile users. This special issue of E-Letter focuses on the recent progresses of cloud-based multimedia computing, storage and transmission algorithms and systems. We are very glad to introduce five interesting papers in this topic area, from leading research groups around the world, to report their latest solutions and results.

In the first article titled, "*vSkyConf:Cloud-assisted Multi-party Mobile Video Conferencing",* Wu, Wu, Lau from the University of Hong Kong and Li from Hong Kong University of Science and Technology present their multi-party video conferencing system for mobile users, that exploits an IaaS (Infrastructure as a Service) cloud platform for efficient conferencing stream exchange and transcoding. Each mobile user corresponds to a surrogate virtual machine in the IaaS cloud, which sends and receives streams on behalf of the user, in order to achieve much better scalability of a conferencing session. Their design is implemented on Amazon EC2 and the preliminary evaluation results illustrate the high streaming quality perceived by the participants of a conferencing session.

The second article is authored by Zhao, Zhang, Ma and Liu from Simon Fraser University and Yang from Beijing University of Posts and Communications, titled "*Computation Offloading for Cloud-assisted Mobile Video Compression*". The authors propose to offload mesh-based motion estimation, for compression of videos captures by a mobile device, to the cloud, by uploading the reference frames and mesh nodes to the cloud servers. A smart mesh node selection algorithm is proposed, which achieves much lower data upload traffic to the cloud, and thus significantly reduces the transmission energy consumption.

The third article is contributed by Liu and Shu from Huazhong University of Science and Technology, China, and the title is *"e-Time:Energy-Efficient Mobile Cloud Computing for Rich-Media Applications"*. Targeting at energy-efficient data transmission between mobile devices and the cloud, they design a mobile cloud system, eTime, which automatically perceives network conditions and smartly schedules data transmissions for different applications (social network systems, cloud storage of multimedia contents) in an online fashion. Preliminary results on a commercial cloud platform (Sina App Engine) show that eTime can achieve 20-35% energy saving, as compared to a commonly used, random transmission mechanism.

In the fourth article titled *"Adaptive Transrating System for Cloud-Based Live Streaming"*, Lai, Ma and Chao from National Ilan University and National Taiwan University of Science and Technology, Taiwan, investigate an adaptive stream transrating mechanism, and implement it based on the HTTP Live Streaming Protocol. The server side records the present streaming video content and bandwidth condition of the user, and analyzes the bandwidth variance at the past time in order to evaluate and predict probable changes in the bandwidth in the future. The mechanism can provide users with video play quality according to the network conditions, without requiring additional information from the streaming devices.

The fifth article is *"User-Assisted Cloud Storage System: Opportunities and Challenges"*, from Chu, Liu and Leung from Hong Kong Baptist University, Li from University of Calgary, and Lei from Beijing University of Posts and Communications, China. Multimedia storage has been a popular application in cloud platforms. This paper presents opportunities and challenges of a cloud storage architecture that exploits the unused network and storage resources at the users, in order to provide highly available and reliable cloud storage services at lower costs. A successful solution should focus on incentive design to motivate users to contribute resources, the provision of availability and reliability across the storage system, and guarantee of user experience.

We would like to thank all the authors for their contribution. By providing an up-to-date sketch of the current research efforts in multimedia cloud computing, we hope these articles will stimulate further research in the topic area.

Zongpeng Li received his B.E. degree in Computer Science and Technology from Tsinghua University (Beijing) in 1999, his M.S. degree in Computer Science from University of Toronto in 2001, and his Ph.D. degree in Electrical and Computer Engineering from University of Toronto in 2005. Since August 2005, he has been with the Department of Computer Science in the University of Calgary. In 2011-2012, Zongpeng was a visitor at the Institute of Network Coding, Chinese University of Hong Kong. His research interests are in computer networks, particularly in network optimization, multicast algorithm design, network game theory and network coding. Zongpeng was named an Edward S. Rogers Sr. Scholar in 2004, won the Alberta Ingenuity New Faculty Award in 2007, was nominated for the Alfred P. Sloan Research Fellow in 2007, and received the Best Paper Award at PAM 2008 and at HotPOST 2012.

Chuan Wu received her B.E. and M.E. degrees in 2000 and 2002 in Computer Science and Technology from Tsinghua University, China, and her Ph.D. degree in 2008 in Electrical and Computer Engineering from University of Toronto, Canada. Since 2008, she has been an assistant professor in the Department of Computer Science, the University of Hong Kong. Her research interests include cloud computing, mobile computing, peer-to-peer networks and online/mobile social networks. She is a member of IEEE and ACM. She was a recipient of the Best Paper Award at HotPOST 2012.

vSkyConf: Cloud-assisted Multi-party Mobile Video Conferencing

Yu Wu[*], *Chuan Wu*[*], *Bo Li*[†] *and Francis C.M. Lau*[*]
[*]*Department of Computer Science, The University of Hong Kong,*
Email: {ywu,cwu,fcmlau}@cs.hku.hk
[†]*Dept. of Computer Science and Engineering, Hong Kong University of Science and Technology,*
Email: bli@cse.ust.hk

I. Introduction

Online video conferencing has been widely deployed for virtual, face-to-face communication among separate parties, as a greener solution to replace many of the energy-expensive conference travels. Advances in mobile and wireless communication technologies have enabled mobile users to exploit new evolution of phone calls — mobile video conferencing calls — as part of their everyday life, anytime anywhere on the move.

Traditional video conferencing relies on either expensive dedicated multiple control units (MCU), or distributed peer-to-peer (P2P) architectures for signal processing, ingress session transcoding and multiple stream dissemination to end devices. A survey of several representative applications [1][2][3][4][5] unveils the situation that solutions with infrastructure support (S/C) tend to support more concurrent users under expensive user subscription fees, while the P2P-based counterparts are reluctant to allow group video calls, for a fear of compromising call qualities. Most applications stick to flat streaming rates, and even those few providing Dynamic Video Quality (DVQ) can only adapt videos to a limited number of bit rates. Hence, we conclude that high-quality, multi-party mobile video conferencing is still a pending goal to achieve, with key challenges as follows: (1) The workload on each node in a video conferencing session, in terms of both processing and transmission, scales quadratically to the size of the session, which makes it challenging to use mobile devices for multi-party video conferencing. (2) Mobile users are equipped with different devices and downlink speeds; a high-quality solution should enable differentiated call qualities to different users, instead of a homogeneous video broadcast rate enforced by the low-end users.

In this paper, we present *vSkyConf*, a cloud-assisted multi-party mobile video conferencing system over heterogeneous mobile devices.

II. Architecture

A surrogate, *i.e.*, a virtual machine (VM) instance, is created in the IaaS cloud for each mobile user, and responsible for (i) session maintenance, by exchanging control messages with other surrogates; (ii) video dissemination and transcoding, by receiving the video streams its mobile user produces, transcoding it into appropriate format(s), distributing it to the other users' surrogates; (iii) efficient video buffering for its mobile user, for timely, smooth and robust streaming to the corresponding device.

Our design of vSkyConf observes the following principles: (i) decentralized control; (ii) self-evolving routing topology with full adaptivity, in terms of dynamic routing and transcoding decisions; (iii) synchronized playback of multiple streams; (iv) robust, smooth video streaming in case of network jitters and inaccurate route computation.

The key modules on a single surrogate are depicted in Fig.1, which can be divided into two parts: the *control plane* and the *data plane.* The *control plane* is responsible for control signaling between surrogates. The *data plane* is responsible for processing in/out video streams, in terms of both transcoding and forwarding.

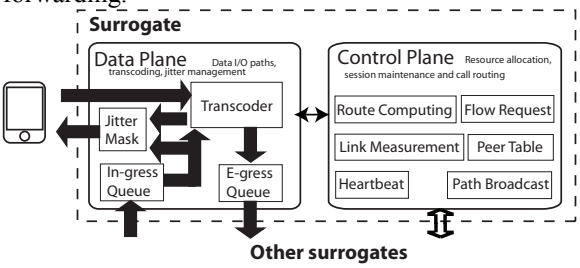


Fig.1 The key modules of a surrogate.

III. Design and Implementation

1. Session Maintenance

When a mobile user logs in to the system via a gateway server, which keeps track of participating users and their surrogates, it is assigned a surrogate VM. The surrogate of the session initiator finds out IP addresses of surrogates of the other users in the session from the gateway server and contacts the participants through their surrogates. Each participant sends periodical "heartbeat" messages to the session initiator, and receives the updated list of IP addresses together with the time-stamped "ack" to calibrate its local "clock". When a mobile user leaves the system, its surrogate VM is released.

2. Routing Computation

Let $G = (S, \varepsilon)$ represent the network of surrogates in a session, where S is the set of surrogates and ε is the set of directed connections among the surrogates. For each $m \in S$, let $\hat{m}$ represent the corresponding mobile user. Suppose C_{ij} is the maximal available bandwidth on link $(ij) \in \varepsilon$, and d_{ij} denotes the link latency. We refer to the stream from a surrogate $m \in S$ as flow m,

with source rate $R_{\hat{m}}^{(m)}$ and maximum acceptable bit rate $R_{\hat{n}}^{(m)}$ at mobile user $\hat{n}$. For ease of practical implementation, we restrict each unicast flow from m to n to be an integral flow along one path with the end-to-end rate $r_n^{(m)}$. Let I_{ij}^{mn} indicate whether the conceptual unicast flow [6] from m to n traverses link $(ij) \in \varepsilon$, $c_{ij}^{(m)}$ denote the actual rate of the multicast flow m on link (ij) and function $\varphi_n(r_1, r_2)$ give the transcoding latency at surrogate n from rate r_1 to r_2 (if $r_1 \le r_2$, $\varphi_j(r_1, r_2) = 0$). If we bound the end-to-end latency (from the time a source surrogate m emits flow m to the time a receiver surrogate n is ready to push the stream to its corresponding mobile user) by $L_n^{(m)}$ whose value is dynamically decided as in Sec.III-4, the optimization problem can be formulated as follows.

$$\max \sum_{m \in S} \sum_{n \in S, n \neq m} U\left(\frac{r_n^{(m)}}{R_{\hat{n}}^{(m)}}\right)$$

subject to:

$$\sum_{i:(i,j)\in\varepsilon} I_{ij}^{mn} - \sum_{k:(j,k)\in\varepsilon} I_{jk}^{mn} = b_j^{mn}, \forall j, m, n \in S, m \neq n$$

$$I_{ij}^{mn} r_n^{(m)} \le c_{ij}^{(m)}, \forall (i,j) \in \varepsilon, m, n \in S, m \neq n$$

$$\sum_{m \in S} c_{ij}^m \le C_{ij}, \forall (i,j) \in \varepsilon$$

$$\sum_{(i,j)\in\varepsilon} I_{ij}^{mn} d_{ij} + \sum_{(i,j)\in\varepsilon} \sum_{k:(j,k)\in\varepsilon} I_{ij}^{mn} I_{jk}^{mn} \varphi_j(c_{ij}^m, c_{jk}^m) + \varphi_n\left(\sum_{j:(j,n)\in\varepsilon} I_{jn}^{mn} C_{jn}^{(m)}, R_{\hat{n}}^{(m)}\right) \le L_n^{(m)},$$

$$\forall m, n \in S, m \neq n$$

$$I_{ij}^{mn} \in \{0,1\}, \forall m, n \in S, m \neq n, (i,j) \in \varepsilon$$

$$0 \le r_n^{(m)} \le R_{\hat{m}}^{(m)}, \forall m, n \in S$$

$$0 \le r_n^{(m)} \le R_{\hat{n}}^{(m)}, \forall m, n \in S$$

where

$$b_j^{mn} = \begin{cases} -1, & j = m \\ 1, & j = n \\ 0, & otherwise \end{cases}$$

3. Distributed Heuristics

We first decide a basic, feasible dissemination topology for each flow m. $\omega_n^{(m)}$ represents the overall latency for flow m from surrogate m to surrogate n. Based on the basic topology, each surrogate carries out dynamic edge and rate adjustments, in order to maximally utilize the available capacity to stream high-quality streams, without violating the latency constraints. The detailed two phased algorithms are depicted as Algorithm 1 and Algorithm 2.

4. Jitter Masking

Synchronization among different streams received at all users is crucial to users' perceived quality of experience. We design an effective buffering mechanism at the surrogates where an end-to-end delay

Algorithm 1 Flow Routing and Rate Allocation

1: Construct shortest-path trees from each surrogate m, $T^{(m)}$;
2: **if** $\exists m, n \in \mathcal{S}, \omega_n^{(m)} > L_n^{(m)}$ **then**
3: No feasible solution exists; **return** ;
4: **end if**
5: N_{ij} := Number of dissemination trees on (i,j);
6: $\forall (a,b) \in T^{(m)}, c_{a,b}^{(m)} := min_{k \in \mathcal{S}, (i,j) \in T^{(m)}} \{R_{\hat{k}}^{(m)}, \frac{C_{ij}}{N_{ij}}\}$;
7: Search for better routing paths, following Alg. 2;

Algorithm 2 Self-Evolving Route/Rate Adjustment at Surrogate n in Flow m

1: **while** $\exists (j,n) \in T^{(m)}, c_{jn}^{(m)} < R_n^{(m)}$ **do**
2: **if** $\exists (i,k) \in T^{(m)}, min\{c_{ik}^m, \bar{C}_{kn}\} > c_{jn}^{(m)}$ **then**
3: $\Lambda := \{n\} \bigcup \{q : (n,q) \in T^m\}$;
4: **if** $\forall p \in \Lambda, \omega_p^{(m)} \le L_p^{(m)}$ **then**
5: $T^{(m)} := T^{(m)} - (j,n) + (k,n)$;
6: **end if**
7: **end if**
8: **end while**

of D is enforced. Let Δ_m indicate delay between mobile device $\hat{m}$ and its surrogate m. For a frame produced at t at the source $\hat{m}$, it will be buffered at surrogate n until $t + \mathcal{L}_n^{(m)}$, where $\mathcal{L}_n^{(m)} = D - \Delta_m - \Delta_n$, in order to guarantee playback of the frame at the mobile device $\hat{n}$ at $t + D$. If there is no jitter in the cloud, we could set the delay bound $L_n^{(m)}$ in the optimization in Sec.III-2, to exactly $\mathcal{L}_n^{(m)}$. However, in a practical system, jitter may occur due to various reasons. The existing measurement work [7] has shown that jitter on a network path approximately follows a normal distribution. Let σ be the standard deviation. We can derive $L_n^{(m)} = \mathcal{L}_n^{(m)} - 3.4\sigma$ to make sure that 99.97% of the video packets can catch their playback deadlines at mobile device $\hat{n}$.

IV. Performance evaluation

We implement a prototype and deploy it in Amazon Elastic Compute Cloud (EC2). Surrogates are provided from "ap-southeast-1a" region for 5 Hong Kong users, "eu-west-1a" for 1 European user, "us-west-1b" and "us-east-1a" for 2 users in west US and east US each, respectively. We emulate dynamic environments by manually injecting jitters up to 150 ms on the links between Hong Kong and Europe via Dummynet [8]. Both uplink and downlink bandwidths of each emulated mobile user are within the range of [1.5, 2] Mbps. We choose $\log(x)$ as the utility function, and the latencies between surrogates are the actual delays between Amazon EC2 instances.

We investigate the conferencing performance at the initiator's surrogate. Fig. 2 illustrates the flow rates for streams from 3 among the other 9 conference participants. We can see that all flows go through a "fast" startup stage, when the basic stream

dissemination topology is being constructed, and then evolve towards their maximal acceptable rates. Fig. 3 presents the load in the jitter buffer for "flow-b" at the initiator's surrogate, where we see that the buffering level varies significantly when "flow-b" takes a path going through the link between "eu-west-1a" and 'ap-southeast-1a", due to jitters along the link. The algorithm then rediects "flow-b" through a better path via the "us-west-1b" region, which leads to a more stable buffering level later on. Fig. 4 shows the corresponding latency of each flow. We observe that the latency of "flow-b" varies more significantly, due to the manually imposed jitters, before redirected to a better path, meeting well the end-to-end latency required (400ms).

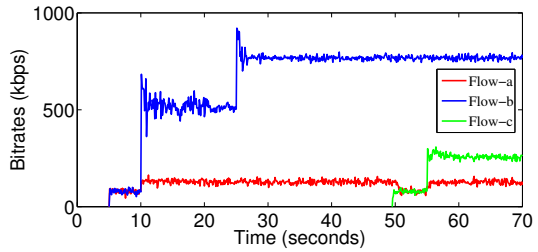


Fig.2 Flow rates at the initiator's surrogate

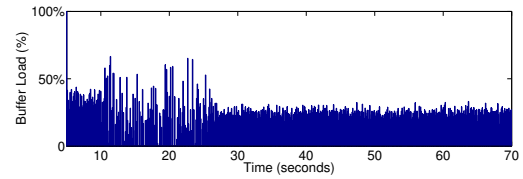


Fig. 3 Load of flow-b's buffer at the initiator's surrogate

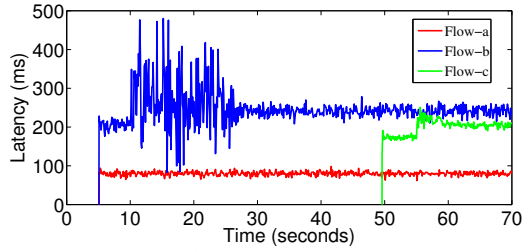


Fig. 4 Flow latencies at the initiator's surrogate

V. Conclusion

We present *vSkyConf*, designed to fundamentally improve the quality and scale of multi-party mobile video conferencing. Leveraging a virtual machine as the exclusive surrogate for each mobile user, *vSkyConf* applies a fully decentralized, efficient algorithm to decide the optimal stream distribution paths as well as the most suitable transcoding spots along the paths, and tailors a buffering mechanism to realize playback synchronization.

References

[1] Skype, http://www.skype.com/
[2] LifeSize, http://www.lifesize.com/
[3] Tango, http://www.tango.me/
[4] Vidyo, http://www.vidyo.com/
[5] Fring, http://www.fring.com/
[6] Z. Li, B. Li, D. Jiang, and L. C. Lau, "On achieving optimal throughput with network coding," in IEEE INFOCOM, 2005.
[7] M. J. Karam and F. A. Tobagi, "Analysis of delay and delay jitter of voice traffic in the internet," Computer Networks, vol. 40, no. 6, pp.711–726, Dec. 2002.
[8]Dummynet, http://info.iet.unipi.it/luigi/dummynet

Yu Wu received his B.E. and M.E. degrees in 2006 and 2009 in Computer Science from Tsinghua University, China. He is currently a Ph.D. candidate in the Department of Computer Science, the University of Hong Kong. His research interests include cloud computing, mobile computing and software defined networking.

Chuan Wu received her B.E. and M.E. degrees in 2000 and 2002 in Computer Science from Tsinghua University, China, and her Ph.D. degree in 2008 in Electrical and Computer Engineering from University of Toronto, Canada. She is currently an assistant professor in the Department of Computer Science, the University of Hong Kong. Her research interests include cloud computing, peer-to-peer networks and online/mobile social network.

Bo Li (IEEE Fellow) is a professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He helds a Cheung Kong Chair Professor in Shanghai Jiao Tong University. His current research interests include: large-scale content distribution in the Internet, cloud computing, datacenter networking, green computing and communications. He received his B. Eng. Degree in Computer Science from Tsinghua University, China, and his Ph.D. degree in Electrical and Computer Engineering from University of Massachusetts at Amherst.

Francis C.M. Lau (SM, IEEE) received his PhD in Computer Science from the University of Waterloo, Canada. He has been a faculty member in the Department of Computer Science, The University of Hong Kong, since 1987, where he served as the department head from 2000 to 2006. His research interests include networking, parallel and distributed computing, algorithms, and application of computing to art.

Computation Offloading for Cloud-Assisted Mobile Video Compression

Yuan Zhao, Lei Zhang, Xiaoqiang Ma, Jiangchuan Liu
Simon Fraser University, Canada
{yza173, lza70, xma10, jcliu}@sfu.ca
Yu Yang, Beijing University of Posts and Telecommunications, China
yangelm@vip.sina.com

1. Introduction

Mobile devices, including smart phones and tablets, are increasingly penetrating into people's everyday life as efficient and convenient tools for communication and entertainment. Despite the fast development of mobile CPU, GPU, memory, and wireless access technologies, mobile applications are still confined by their limited computation capability and limited battery energy [1]. Nowadays, mobile users increasingly use mobile devices to capture video in real-time, expecting to encode the video and then stream it to the Internet in real-time. Both video encoding and streaming require heavy energy consumption. To stream a video to the Internet, the user has to copy the video to a personal computer then compress and upload the video. This however is not convenient and discourages people to share mobile videos on the Internet.

The emergence of cloud computing [2] has been dramatically changing the landscape of modern computer applications and brings new opportunities to mobile devices. The cloud computing enables end users to conveniently access computing resources in a pay-as-you-go manner. This new computing paradigm offers elastic and cost-efficient resource provisioning. Intuitively, offloading computation to the cloud is beneficial whenever a computation intensive task is not affordable by local resources. However, moving the task to the remote cloud introduces a large volume of data transfer, which in turn introduces more energy consumption and data transfer fees.

In this article, we present a novel method of computation offloading in mobile video compression. We identify the key issues in developing new applications that effectively leverage cloud resources for computation-intensive modules. We then analyze cloud-assisted motion estimation for mobile video compression, to illustrate the benefit, implementation, and unique challenges of computation offloading in mobile cloud computing.

2. Issues and Challenges

Assembling local resources and remote clouds organically to make offloading transparent to mobile users requires nontrivial effort. First, the key motivation of offloading must be determined: to save energy, to improve computation performance, or both? This serves as a guideline for the system design of the computation offloading. Second, the potential offloading gain needs to be well evaluated. There is no incentive to resort to clouds for a job that can be easily and efficiently executed locally. Even for computation intensive applications that can be hardly executed locally, moving them to the remote cloud may introduce a large volume of data transfer.

Video encoding is an example of computation intensive mobile applications. A typical video compression consists of motion estimation, transformation, quantization, and entropy coding. Usually a video compression job is performed as a whole process on a single device. Therefore, a smart method should be used to decide what and how to offload the computation to the cloud server. Apparently, offloading the whole video compression task to the cloud is not practical because it is almost the same as uploading the whole raw video data [4]. A profiling shows that motion-estimation is the most computation-intensive module, accounting for almost 90% of the computation. For example, to encode a video of 5 seconds (30 frames per second with resolution of 176 $\times$ 144 and pixel depth of 8 bits) using an H.264 encoder needs almost 10^{10} CPU cycles [3], which means that a 2 GHz CPU is required for real-time encoding. Considering that the newest smart phones and tablets are equipped with high-definition cameras, the CPU workload can be 5-10 times higher than that in the above example. Therefore, motion estimation should be the focus of computation offloading.

3. Mesh-Based Motion Estimation Offloading

Generally, it is not simple to decouple motion estimation from others given the data dependency. For example, the motion estimation of a P frame depends on the data of the previous reference I or P frame. It is necessary to ensure that a minimum amount of data (not all the reference frame data) is uploaded to the cloud and yet estimation can be done accurately, which inspired our design of Cloud-Assisted Motion Estimation (CAME).

CAME employs a mesh-based motion estimation that synergies the mobile device and the cloud. The mesh-

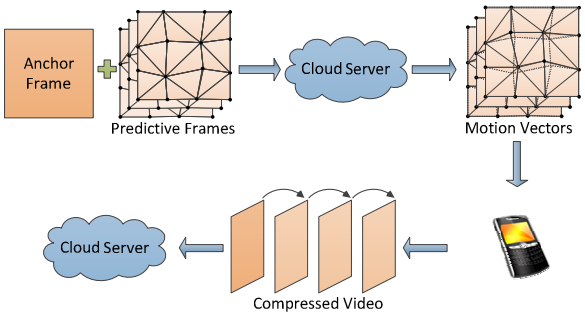
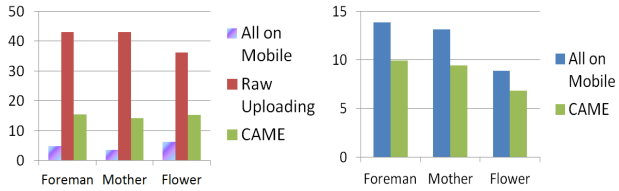


Fig. 1 CAME Architecture



(a) (b)

Fig. 2 (a) Total transmission volume (MB)
(b) Total energy consumption in CPU cycles (billion cycles)

based motion estimation is performed by estimating one motion vector (MV) for each mesh. Mesh nodes are sampled on the reference frame and MVs are calculated from predictive frames, then sub block MVs are calculated using block-based motion estimation inside each mesh. Mesh-based motion estimation provides an opportunity for computation offloading by uploading only the reference frames and mesh nodes to the cloud server, which reduces the data transfer to the minimum.

Unlike standard mesh-based motion estimation, CAME exploits a smart algorithm for mesh node selection. CAME applies a reversed mesh node selection and motion estimation algorithm, in which mesh nodes are sampled on P frames and MVs are calculated from the mesh and the reference frame. CAME gains benefits by uploading only the reference frame and mesh data of P frames instead of uploading the whole video frames. With CAME, a mobile device can upload reference frames and mesh data to the cloud for estimation (mesh node motion estimation), which are of much smaller data volume. It then downloads the estimated Motion Vectors (MVs) from the cloud server and completes the remaining video encoding steps.

CAME architecture is illustrated in Fig. 1, which includes the following four key steps:
a) On the mobile device, the raw video is divided into macro blocks (MBs). For each MB, a reference frame is extracted, together with a mesh for each successive P frame. The device then uploads the reference frame and meshes to the cloud.
b) The cloud server conducts the mesh motion estimation for the uploaded reference frame and meshes, and pushes the generated mesh MVs back to the CAME client on the mobile device.
c) The mobile device, upon receiving the MVs for mesh nodes of each P frame, continues to calculate sub block MVs using block-based motion estimation as well as entry coding.
d) The encoded video is then stored in the device or streamed to internet via wireless interfaces.

Fig. 2(a) compares the total amount of transmitted data for three standard videos Foreman, Mother, and Flower. The baseline here is All on Mobile (AoM), which executes the entire video encoding on mobile devices. The transmission energy consumption is converted into CPU cycles [5] such that the total energy consumption can be quantified. Though Flower's original video size is smallest, the AoM and CAME transmission size is largest because Flower has higher spatial details. It is not surprising that AoM method's data transmission cost is the lowest among all three and the raw uploading has the largest cost. Compared to AoM method, the proposed method introduces more transmission because of the extra data transmission for mesh node uploading and mesh motion vectors downloading. Compared to raw uploading, CAME method still saves approximately 60% on total data transmission. Although CAME consumes more energy on transmission than AoM does, it saves the total energy consumption through offloading the most computation-intensive task motion estimation to cloud servers. CAME spends nearly 40% less energy on computation than the AoM method. Further, Fig. 2(b) confirms the expectation that CAME achieves up to 30% total energy savings on video encoding and transmission compared to AoM.

4. Conclusion

In this article, we presented computation offloading in mobile cloud computing which is a new paradigm combining mobile devices and cloud computing resources to provide seamless rich experiences to mobile users. This new paradigm brings opportunities as well as challenges. Through the case study of cloud-assisted mobile video compression, we illustrated how mobile applications can be enhanced with cloud computing to achieve improved performance or energy saving, and we examined the trade-offs.

References

[1] K. Kumar et al., "A survey of computation offloading for mobile systems", Mobile Networks and Applications, doi: 10.1007/s11036-012-0368-0, 2012,

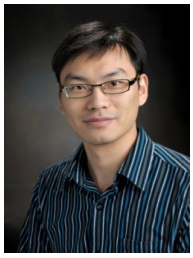
pp. 1-12.

[2] M. Armbrust et al., "A view of cloud computing", Communications of the ACM, vol. 53, no. 4, Apr. 2010, pp. 50-58.

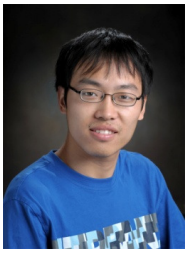
[3] N. Imran, B.-C. Seet and A.C.M. Fong, "A comparative analysis of video codecs for multihop wireless video sensor networks", Multimedia Systems, vol. 12, 2012, pp. 373-389.

[4] Y. Zhao et al., "CAME: Cloud-assisted motion estimation for mobile video compression and transmission", Proc. ACM NOSSDAV '12, 2012.

[5] N. Balasubramanian, A. Balasubramanian and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications", Proc. ACM SIGCOMM '09, 2009, pp. 280-293.



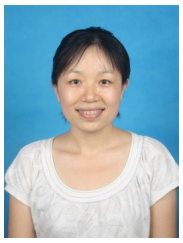
Yuan Zhao received his B. Eng. degree from Beihang University, China. He was a software engineer in IBM China Development Lab. He is now a master student from School of Computing Science, Simon Fraser University, Canada. His areas of research interest are multimedia communications, social networking, and cloud computing.



Lei Zhang received his B. Eng. degree from Huazhong University of Science and Technology, China. He is now a master student from School of Computing Science, Simon Fraser University, Canada. His areas of research interest are multimedia communications, wireless networks, and cloud computing.

Xiaoqiang Ma received the B.Eng degree from Huazhong University of Science and Technology, China., and the M.Sc. degree from Simon Fraser University, Canada. He is now a Ph.D. student from School of Computing Science, Simon Fraser University, Canada. His areas of interest are wireless networks, social networks, and cloud computing.

Jiangchuan Liu received the BEng degree (cum laude) from Tsinghua University, Beijing, China, in 1999, and the PhD degree from The Hong Kong University of Science and Technology in 2003, both in computer science. He is a recipient of Microsoft Research Fellowship (2000), Hong Kong Young Scientist Award (2003), and Canada NSERC DAS Award (2009). He is a co-recipient of the Best Student Paper Award of IWQoS'2008, the Best Paper Award (2009) of IEEE ComSoc Multimedia Communications Technical Committee, and Canada BCNet Broadband Challenge Winner Award 2009. He is currently an Associate Professor in the School of Computing Science, Simon Fraser University, British Columbia, Canada, and was an Assistant Professor in the Department of Computer Science and Engineering at The Chinese University of Hong Kong from 2003 to 2004. His research interests include multimedia systems and networks, wireless ad hoc and sensor networks, and peer-to-peer and overlay networks. He is a Senior Member of IEEE and a member of Sigma Xi. He is an Associate Editor of IEEE Transactions on Multimedia, and an editor of IEEE Communications Surveys and Tutorials. He is TPC Vice Chair for Information Systems of IEEE INFOCOM'2011.



Yu Yang received her Ph.D. degree in Cryptography from Beijing University of Posts and Telecommunications (BUPT). She received a M.S. degree in Computer Software and Theory in 2003 from BUPT. Currently as a lecturer at the School of Computer Science BUPT, her research interests are watermarking, information hiding and information security. She is a member of IET and the co-author of over 20 scientific international SCI-indexed or EI-indexed papers. She has published two textbooks and completed several projects in provincial or ministry levels.

eTime: Energy-Efficient Mobile Cloud Computing for Rich-Media Applications

Fangming Liu (IEEE Member) and Peng Shu
School of Computer Science & Technology
Huazhong University of Science & Technology, Wuhan, China
fmliu@mail.hust.edu.cn

1. Introduction

With an explosive growth of wireless mobile devices nowadays, there is a shift of user preferences from traditional cell phones and laptops to smartphones and tablets. Advances in the portability and capability of mobile devices, together with widespread 3G/4G LTE networks and WiFi accesses, have brought out rich mobile multimedia application experiences to end users. Undoubtedly, the demand for ubiquitous access to a wealth of rich-media content and services will continue to be skyrocketing, ranging from social networking services (SNS), cinematic-quality video streaming [1]-[3] to online cloud storage for various multimedia content [4]. It is reported by Cisco [5] that, the traffic from mobile devices is anticipated to exceed the traffic from wired devices by 2014, and to account for 61% of the total IP traffic by 2016.

Rich mobile applications have provided most of the impetus for the prosperity of mobile markets. The time that users spend on mobile applications is increasing fast and has surpassed that of web browsing on PCs in 2011. To overcome resource constraints on mobile devices, some applications have started to leverage cloud computing to extend the capabilities of mobile devices. For example, Apple's Siri takes advantage of computation resources in the cloud to create a speech recognition system with an ability to rival human understanding. Dropbox extends the storage capacity of mobile devices by storing and synchronizing the mobile data in Amazon S3 storage system. These rich applications evoke frequent data transmissions via wireless networks, which constitute a major part of energy consumption on mobile devices.

However, wireless networks are stochastic: not only the availability and network capacity of access points (APs) vary from place to place, but the downlink and uplink bandwidth also fluctuates conditioned on weather, building shields, mobility, flash crowd, and so on. Such stochastic characteristics may incur unpredictable energy consumption in communications between mobile devices and the cloud. In particular, recent measurement studies [6]-[7] show that the energy consumption for transmitting a fixed amount of data is inversely proportional to the available bandwidth. This implies that transmitting data in "good" connectivity could save energy considerably compared to doing so in "bad" connectivity.

Inspired by the above considerations and our preliminary experiments [8], the energy-efficient data transmissions in mobile devices can be achieved by seizing the *good timing* for data transfers. We envision a mobile cloud system framework, named *eTime* [8] as illustrated in Fig. 1, which can automatically perceive network conditions and smartly schedule data transmissions for different popular applications (e.g., SNS, cloud storage for various multimedia content) according to such conditions. Due to the energy saving and potentially more data downloaded, such a framework may lead to prolonged battery life and enhanced user experience in mobile devices.

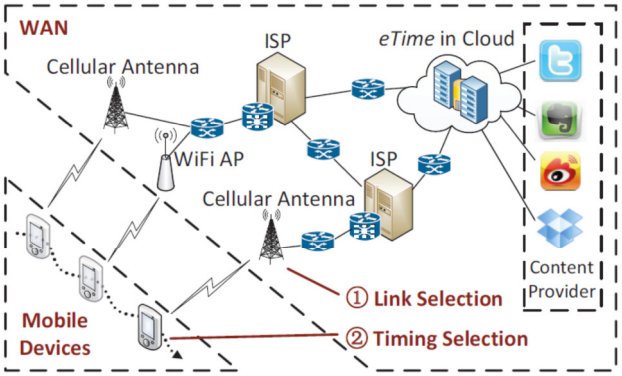


Fig. 1: *eTime* employs cloud computing to provide data management for mobile devices [8].

2. Challenges

Nevertheless, to realize the mobile cloud management system for energy-efficient data transmissions, a number of challenges arise.

First, wireless networks are stochastic and unpredictable — it is hard to tell when the good time is for transmission, even if network conditions are known.

Second, as applications are usually independent of each other and rely on different content providers (e.g., Netflix, iCloud, Facebook), which are deployed on different hosts (e.g., Amazon EC2, Microsoft Azure, private servers), it is challenging to gather and manage data for multiple applications from various source (content) servers jointly.

Third, scheduling may incur delays, if transmission is deferred to a later time. The choices in energy-delay tradeoff may vary with the application type and user context, e.g., users prefer shorter delays when the energy is ample or the response time is strict, while valuing energy savings when they run out of battery.

3. *eTime*: a Mobile Cloud Framework

To address the above challenges, we propose *eTime*, a system framework targeting at ***E***nergy-efficient data ***T***ransm***I***ssions between ***M***obile d***E***vices and the cloud. As shown in Fig. 1, we leverage cloud computing to provide transmission management for various mobile cloud applications while reducing the burden placed on mobile devices. Specifically, *eTime* focuses on realizing energy savings in prefetching-friendly and/or delay-tolerant applications, in which data transfers can be scheduled flexibly without degrading user experiences.

For example, many users rely on Google Maps when visiting a sightseeing spot. Before they open Google Maps on their mobile phones to check the local area, the cloud could pull useful maps from the map provider and push them in advance to the smartphones in an energy-efficient manner. This is especially useful when users visit natural sights, as network connectivity in the wild is not reliable and battery recharge is difficult. On the other hand, in cloud storage applications like iCloud and Dropbox, deferring the synchronization of newly generated mobile data for a short period may not hurt user experience. In such cases, *eTime* can help delay data transfers until good connectivity is observed.

3.1 Online Scheduling

To cope with the intrinsically stochastic nature of wireless networks, we apply theoretic-sound Lyapunov optimization [9] to design an online transmission control framework, which only relies on information of current network bandwidth and data queue backlogs to make scheduling decisions.

3.2 Cloud-based Multi-App Coordination

Rather than running a control algorithm in each application on the mobile device natively and requesting data for different applications separately, *eTime* gathers and stores the data from different applications in the cloud and schedules data transfers using a centralized algorithm. Since a cloud platform has abundant computing resources and is well connected to multiple carriers and ISPs with high-speed links, the cloud-assisted *eTime* promises to be a powerful agent between mobile devices and original content providers.

3.3 Tuning the Energy-Delay Tradeoff

eTime is able to quantitatively control the energy-delay tradeoff in the scheduling. By tuning a single control parameter V in our control algorithm [8] based on Lyapunov optimization, *eTime* can adaptively balance such a tradeoff according to application delay requirements and user contexts.

3.4 Computation Offloading

To save energy in network communication while confining energy consumption in the associated computation overhead, *eTime* offloads most of its workload (e.g., data scheduling and state monitoring) to the cloud, causing the minimum overhead in resource-constrained mobile devices.

4. Implementation and Preliminary Results

We have implemented *eTime* on a commercial cloud platform (Sina App Engine [10]) and HTC smartphones running Android Gingerbread OS. We conduct experiments on social networking service (SNS) applications (e.g., Sina Weibo) which utilize *eTime* for content prefetching.

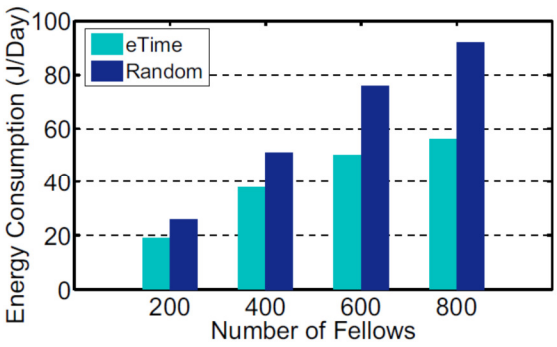


Fig. 2: Energy consumption vs. fellow magnitudes under *eTime* vs. a generally used random strategy [8].

Specifically, we walk around in our campus, carrying the HTC Desire Phone, in which SNS applications download the data packets from the cloud via *eTime* and record the energy consumption per minute. Since the data arrival rate in SNS applications mainly depends on the number of fellows the user has, we fake 4 user accounts, each with a total number of fellows from 200 to 800. As shown in Fig. 2, our preliminary experimental results show that *eTime* can achieve 20%-35% energy saving, in comparison to a generally used random strategy: the user may open applications (which run only on mobile devices) and request data in a random time slot, and the data are then transmitted instantly from content providers to the mobile device under unpredictable wireless connectivity.

5. Conclusion

In this article, we present *eTime*, a mobile cloud framework aiming to achieve energy-efficient data transmissions for mobile devices and rich-media applications. By offloading the overhead of management to the cloud, little burden is placed on the mobile device when saving energy via transmission control. It opens up a new playground for fostering cutting-edge research on mobile cloud computing paradigm, which is beneficial to a variety of multimedia networking applications.

Acknowledgments

The work was supported in part by a grant from The National Natural Science Foundation of China (NSFC) under grant No. 61103176, by a grant from the NSFC under grant No. 61133006, by a grant from the Research Fund of Young Scholars for the Doctoral Program of Higher Education, Ministry of Education, China, under grant No. 20110142120079.

References

[1] F. Liu *et al.*, "Cinematic-Quality VoD in a P2P Storage Cloud: Design, Implementation and Measurements", *IEEE JSAC*, Special Issue on Emerging Technologies in Communications, 2013.

[2] J. Dai, F. Liu, *et al.*, "Collaborative Caching in Wireless Video Streaming Through Resource Auctions", *IEEE JSAC*, special issue on Cooperative Networking Challenges and Applications, 2012.

[3] F. Liu *et al.*, "Novasky: Cinematic-Quality VoD in a P2P Storage Cloud," in *Proc. of IEEE INFOCOM*, Apr. 2011.

[4] F. Liu *et al.*, "FS2You: Peer-Assisted Semi-Persistent Online Hosting at a Large Scale," *IEEE TPDS*, vol. 21, no. 10, pp. 1442-1457, Oct. 2010.

[5] Cisco Visual Networking Index: Forecast and Methodology, 2011-2016.

[6] N. Balasubramanian *et al.*, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications," in *Proc. of IMC*, Nov. 2009.

[7] J. Huang *et al.*, "A Close Examination of Performance and Power Characteristics of 4G LTE Networks," in *Proc. of ACM MobiSys*, 2012.

[8] P. Shu, F. Liu (Correspondence) *et al.*, "eTime: Energy-Efficient Transmission between Cloud and Mobile Devices", in *Proc. of IEEE INFOCOM (Mini-conference),* Apr. 2013.

[9] M. Neely, "Stochastic Network Optimization with Application to Communication and Queueing Systems," *Morgan & Claypool Publishers*, 2010.

[10] Sina App Engine. http://sae.sina.com.cn/

Fangming Liu (S'08-M'11) is currently an Associate Professor in the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China; and he was awarded as the Chutian Scholar of Hubei Province, China. He received his B.Engr. degree in 2005 from the Department of Computer Science and Technology, Tsinghua University, Beijing; and his Ph.D. degree in Computer Science and Engineering from the Hong Kong University of Science and Technology in 2011. He was a recipient of the Best Paper Award from the IEEE GLOBECOM'2011 and a recipient of the Best Paper Award from the IEEE IUCC'2012. From Aug. 2009 to Feb. 2010, he was a visiting scholar at the Department of Electrical and Computer Engineering, University of Toronto, Canada. Since 2012, he has also been a StarTrack Visiting Young Faculty in Microsoft Research Asia (MSRA), Beijing.

His research interests are in the area of cloud computing and datacenter networking, peer-to-peer (P2P) networking, green computing and communications, Internet content distribution and multimedia streaming systems. He is a member of IEEE and IEEE Communications Society (also member of IEEE Technical SubCommittee on Green Communications & Computing), a member of ACM, as well as a member of China Computer Federation (CCF) and CCF Internet Technical Committee. He served as a Guest Editor for IEEE Network Magazine, and served as TPC members for IEEE INFOCOM'2013 and GLOBECOM'2012-2013.

Peng Shu is currently a Master student in the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His research interests focus on cloud computing and wireless mobile applications.

Adaptive Transrating System for Cloud-Based HTTP Live Streaming

Chin-Feng Lai[1], Yi-Wei Ma[2] and Han-Chieh Chao[1]
[1]Institute of Computer Science and Information Engineering, National ILan University, Tsaiwan
[2]Department of Electrical Engineering, National Taiwan University of Science and Technology, Taiwan
cinfonlai@ieee.org, yiweimaa@gmail.com & hcc@niu.edu.tw.

1. Introduction

With the increased on-line streaming video services, such as YouTube, PPS, Google+, etc., the public have gradually changed from viewing off-line videos to viewing videos using network streaming on computers and mobile devices, a trend that leads in many research subjects, such as how to deliver video content with appropriate network protocols, how to effectively reduce delay time for playing videos, how to choose an proper video packaging format, how to provide seamless streaming services in cloud networks, how to balance the operating load at the video server-side [1], how to transcode multimedia video for numerous users instantly [2-3], etc.

The existing multimedia streaming technologies can be divided into Push Streaming and Request Streaming approximately. In order to provide users better streaming quality, adaptive streaming has emerged [4], which means the clients are able to choose the video stream quality by the network bandwidth, current network condition or operational capability of hardware. However, in consideration of computing costs, the servers cannot provide numerous options in video streaming quality at the same time, and if the end device bandwidth condition changes drastically during play, the appropriate video quality cannot be switched immediately, which degrades user experience [5]. In order to solve this problem, dynamically transcoded media content was widely adopted in the past, but occasionally optimum video quality cannot be instantly obtained, as traditional transcoding recodes the entire multimedia video content, which consumes too much time. Some studies have attempted to improve the video format coding mode, hoping to accelerate coding; however, such a practice is not yet helpful to long video content.

Therefore, this study aims at the communication mechanism of the HTTP Live Streaming protocol, as proposed by Apple Inc., uses the characteristics of streaming protocol, records the present streaming video content and bandwidth condition of the user at the server-side, and analyzes the bandwidth variance at some time past in order to evaluate and predict probable changes in the bandwidth in the future. The adaptive transcoding streaming video mechanism is used to provide users with video play quality meeting the environment without the streaming device providing additional streaming information.

2. Adaptive Transrating System

The adaptive transrating mechanism system structure in this study is implemented on HTTP live streaming, and aims to instantly provide picture quality suitable for the user's network environment; in addition to the required media segmenter and M3U8 generator for adaptive streaming, the server-side has a bandwidth recorder, segmenter transrating subsystem, and segment redirecting subsystem. The bandwidth recorder records the network conditions of the client-side connected to this server analyzes the on-line quality of client-side, and checks whether it is in steady state. The segmenter transrating subsystem observes the current status of client-side, according to the analysis results of the bandwidth recorder, in order to determine what strategy to use to calculate the coding rate for the next media segment, and to carry out transrating. Finally, the segment redirecting subsystem leads the HTTP media segment requirement of the client-side in the play list of media segments recording transrating. The schematic diagram is as seen in Figure 1.

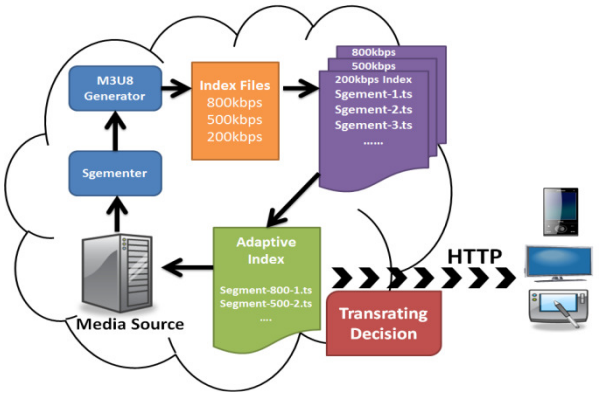


Figure 1. The Architecture of Adaptive Transrating System For Cloud-Based HTTP Live Streaming.

Transrating Decision Mechanism.

In order to effectively evaluate the connection condition between the client-side and server-side, this study uses the tcpdump tool of the Linux system to analyze the TCP packet information between the two sides in order to obtain the client-side information, bandwidth condition, and media segments in streaming,

and implements dynamic transcoding on the tcpdump tool. This study uses tcpdump to store several previous packets in an annular queue, and the number of packets in this annular queue is based on the number of packets to be transmitted per media segment at the present streaming media bit rate. When new packet information enters the annular queue, the total transmission quantity and overall transmission time are updated first, and then, whether the annular queue is full is judged. If it is full, the farthest packet information is deleted and the total transmission quantity and overall transmission time are updated. If it is not full, the present total transmission quantity is divided by the transmission time as the bandwidth information, thus, the server-side can have the latest packet information and master the connection to the client-side.

When the server obtains the bandwidth connection to the client-side, it can analyze the bandwidth condition for further transcoding. However, as the network condition may become unstable at any time, this study defines three modes to handle different network conditions, and uses a mode transition state machine for compensating bandwidth evaluation error. The three modes are Normal Mode, Active Mode, and Conservative Mode.

In Normal Mode, the server-side uses the bandwidth downloading current media segment as the bit rate for transcoding, the computing mode is described as below, n_p is the number of packets, B_i is the bandwidth value measured by using tcpdump to record packets in the annular queue, B_{next} is the target bit rate of the next media segment, and B_{avg} is the average bit rate of this segment download.

$$B_{next} = B_{avg} = \sum_{i=1}^{n_p} B_i \Big/ n_p \quad (1)$$

In Active Mode, the server compares the last bandwidth condition evaluated with the bandwidth condition currently evaluated to determine the trend of bandwidth variation in this period. If the trend is positive, namely, the bandwidth increases with time, the slope of the change trend is multiplied by the remaining time (segment time span minus download time), plus current average bandwidth, as the target bit rate of next transrating. The computing mode is described as below, where α is the slope of change trend, B_{end} is the bit rate of the end of this segment download, and $T_{remained}$ is the remaining time to next download:

$$B_{next} = B_{end} + a * T_{remained}, if\ a > 0 \quad (2)$$

$$B_{next} = B_{avg}, o.w. \quad (3)$$

In Conservative Mode, the server compares the last bandwidth condition evaluated with the bandwidth condition currently evaluated to determine the trend of bandwidth variation in this period. If the trend is negative, namely, the bandwidth decreases with time, the slope of the change trend is multiplied by the remaining time (segment time span minus download time), plus current average bandwidth, as the target bit rate of next transrating. The computing mode is described as below:

$$B_{next} = B_{end} + a * T_{remained}, if\ a < 0 \quad (4)$$

$$B_{next} = B_{avg}, o.w. \quad (5)$$

3. System Implementation Result

In order to implement experimental analysis of the overall system, the coding rate of the test video is 800kbit/s, the resolution is $320*240$, and FPS is 30. In the construction of the testing environment, this video is transrated to 200kbit/s and 500kbit/s, respectively, the media segmenter is used to cut the video to 5-second segments, and then the M3U8 play list generates a subsystem to create the corresponding play list, thus, the original HTTP Live Streaming architecture can provide three kinds of picture quality for the client-side. And the four bandwidth behaviors are as described below:

Stable bandwidth: meaning the network has not drastically changed within a period of time, and the bandwidth is limited to 850kbit/s in this behavior, as represented by Scenario 1 (S1) in the next text.

Network behavior of gradual bandwidth: meaning the network increases gradually within a period of time. This study uses a media segment time span as the unit, and the flow is increased by 50kbit/s each time, as represented by Scenario 2 (S2) in the next text.

Network behavior of bandwidth decreasing gradually: meaning the network decreases gradually within a period of time. This study uses a media segment time span as the unit, and the flow is decreased by 50kbit/s each time, as represented by Scenario 3 (S3) in the next text.

Unstable bandwidth: meaning the network bandwidth not only increases, but also decreases within a period of time. In this scenario, this study gradually increases the bandwidth to 850kbit/s and maintains it for a period of time, and then gradually decreases it to 300kbit/s, as represented by Scenario 4 (S4) in the next text.

Figure 2 shows the bitrate to segments for scenario 1-4 and Figure 3 shows PSNR comparison for scenario 1-4 in the proposed system.

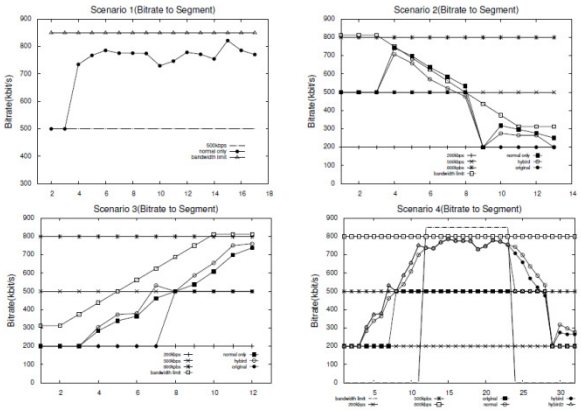


Figure 2. Bitrate to Segment for Scenario 1-4.

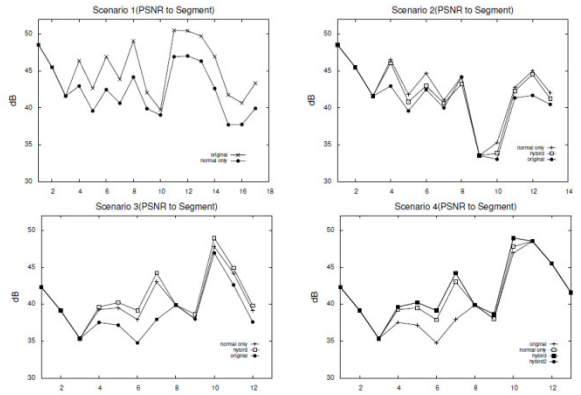


Figure 3. PSNR Comparison for Scenario 1-4.

4. Conclusion

This study designs an adaptive transrating system for cloud-based HTTP live streaming, instantly analyzes the on-line quality between client-side and server-side without changing the HTTP Live Streaming server-side architecture, and provides the optimum picture quality, in order that the client-side leads the HTTP requirement in the transrated media segment using HTTP redirection technology.

References

[1] J. Guo, and L. N. Bhuyan, "Load Balancing in a Cluster-Based Web Server for Multimedia Applications," IEEE Trans. Parallel and Distributed Systems, Vol. 17, No. 11, pp. 1321-1334, Nov. 2006.

[2] A. Mahmood, T. Jinnah, Y. Asfia, and G. A. Shah, "A Hybrid Adaptive Compression Scheme for Multimedia Streaming over Wireless Networks," in Proc. of ICET 4th International Emerging Technologies, pp. 187-192, Oct. 2008.

[3] S. Y. Wu, and C. E. He, "QoS-Aware Dynamic Adaptation for Cooperative Media Streaming in Mobile Environments," IEEE Trans. Parallel and Distributed Systems, Vol. 22, No. 3, pp. 439-450, Mar. 2011.

[4] S.Y. Chang, C.F. Lai, Y.M. Huang, "Dynamic Adjustable Multimedia Streaming Service Architecture over Cloud Computing," Computer Communications Vol. 35, No. 15, pp. 1798-1808, 2012.

[5] A. Begen, T. Akgul, and M. Baugher, "Watching Video over the Web: Part 1: Streaming Protocols," IEEE Internet Computing, Vol. 15, No. 2, pp. 54-63, Mar.-Apr. 2011.

Chin-Feng Lai is an assistant professor at Institute of Computer Science and Information Engineering, National Ilan University. He received Best Study Award from IEEE EUC 2012. He has more than 100 study publications. He is an associate editor-in-chief for Journal of Internet Technology and serves as editor or associate editor for IET Networks, International Journal of Internet Protocol Technology, KSII Transactions on Internet and Information Systems. His research focuses on Mobile Cloud Computing, Cloud-Assisted Multimedia Network, Embedded Systems, etc. He is an IEEE Member since 2007.

Yi-Wei Ma is a Post-Doctoral Fellow in National Cheng Kung University, Taiwan. He received the Ph.D.degree in Department of Engineering Science at National Cheng Kung University, Tainan, Taiwan in 2011. He received the M.S. degree in Computer Science and Information Engineering from National Dong Hwa University, Hualien, Taiwan in 2008. His research interests include internet of things, cloud computing, multimedia p2p streaming, digital home network, embedded system and ubiquitous computing.

Han-Chieh Chao is a joint appointed Full Professor of the Department of Electronic Engineering and Institute of Computer Science & Information Engineering where also serves as the president of National Ilan University, I-Lan, Taiwan, R.O.C. He has been appointed as the Director of the Computer Center for Ministry of Education starting from September 2008 to July 2010.

Dr. Chao is the Editor-in-Chief for IET Networks, Journal of Internet Technology, International Journal of Internet Protocol Technology and International Journal of Ad Hoc and Ubiquitous Computing. Dr. Chao has served as the guest editors for Mobile Networking and Applications (ACM MONET), IEEE JSAC, IEEE Communications Magazine, Computer Communications, IEE Proceedings Communications, the Computer Journal, Telecommunication Systems, Wireless Personal Communications, and Wireless Communications & Mobile Computing. Dr. Chao is an IEEE senior member and a Fellow of IET (IEE). He is a Chartered Fellow of British Computer Society.

User-Assisted Cloud Storage System: Opportunities and Challenges

Xiaowen Chu, Hai Liu, Yiu-Wing Leung
Hong Kong Baptist University
Hong Kong, China
{chxw, hliu, ywleung}@comp.hkbu.edu.hk

Zongpeng Li
University of Calgary
Canada
zongpeng@ucalgary.ca

Min Lei
Beijing University of Posts and Telecommunications, China
byleimin@tom.com

1. Introduction

Cloud storage has recently attracted a substantial amount of attention from both industry and academia. Notable commercial cloud storage services include Amazon S3, Google Drive, Dropbox, Microsoft Skydrive, and Apple's iCloud. Compared with traditional storage systems, cloud storage offers several desirable advantages, including high data availability, high data reliability, and dynamic storage space.

To provide highly available and reliable storage service, cloud storage providers (CSPs) make tremendous investments in storage hardware and network bandwidth. To be profitable, CSPs need to recover these costs from reasonable charges to cloud users. At present, the typical charges include storage ($0.05-0.10 per GB/month), network traffic ($0.05-0.20 per GB), and requests ($0.01 per 1,000 or per 10,000 requests/month, depending on the request type). It is apparent that the major cost comes from storage devices and network bandwidth. When the number of users grows into the millions and the data volume into the exabyte scale, it becomes very critical for CSPs to decrease costs while maintaining the same level of data availability and reliability.

In traditional cloud storage systems, all of the resources are located at data centers. In this paper, we present the opportunities and challenges of a user-assisted cloud storage architecture that aims to reduce the cost of providing highly available and reliable cloud storage services by exploiting the underused storage and network resources of cloud users.

2. Opportunities: User-Assisted Architecture

The envisioned structure of the user-assisted cloud storage system is illustrated in Fig. 1. The CSP invests in several distributed data centers and exploits geo-redundancy to offer continuous service even under extreme cases like natural disasters or power grid failures. The CSPs use the available resources from cloud users as an added asset to augment their own data center resources. This is economically efficient because (1) most users have huge amounts of spare storage space on their PCs and (2) most users pay a fixed monthly fee for their broadband Internet connections. The main motivation behind using the storage space and network bandwidth of cloud users is that it helps CSPs save on both hardware and bandwidth costs. Compared with traditional centralized architecture, user-assisted architecture is more scalable in resource provisioning and hence more cost-effective.

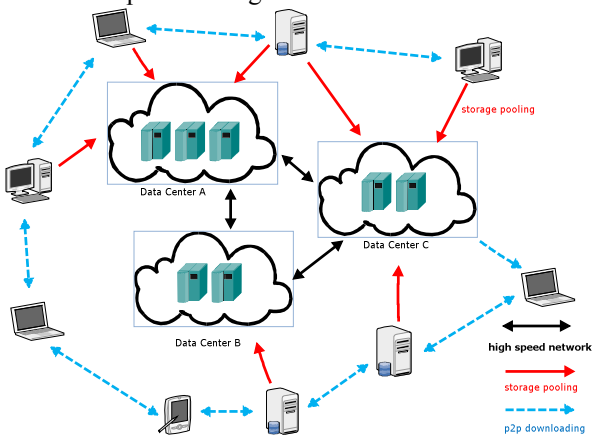


Figure 1 Architecture of user-assisted cloud storage system.

In this user-assisted architecture, CSPs provide data redundancy within their data centers to guarantee high data availability and reliability. The storage services should be resistant to all kinds of failure and system maintenance, such as disk/node/rack failure, power distribution unit failure, and even the failure of an entire data center. CSPs also actively distribute encoded data blocks to cloud users to the extent that each cloud user can download a large portion of his data from other users, hence reducing the network bandwidth costs significantly. Furthermore, a cloud user can better use his network bandwidth by simultaneously uploading/downloading data blocks to/from many other users.

3. Challenges

Although our user-assisted architecture is attractive and promising, it presents several challenges that must be overcome.

Incentive Design: An effective incentive scheme that can motivate cloud users to contribute storage space and outbound bandwidth is vital to the success of our user-assisted architecture. Private BitTorrent communities are a successful example of using a good incentive design to incent end users to contribute as much as possible [1]. These communities deploy a sharing ratio enforcement (SRE) mechanism to overcome the free-riding issue of traditional BitTorrents. SRE forces registered users to keep their

upload-to-download ratios higher than a predefined threshold. A registered user is banned from a private BitTorrent community if his sharing ratio is lower than the threshold. In contrast, a user with a higher sharing ratio is rewarded.

In our user-assisted architecture, end users should be incented to contribute some of their storage space, and more importantly their outbound bandwidth. We must design a new incentive scheme, different from the upload-to-download ratio used in private BitTorrent communities, that can achieve a good balance among the cost of offering the service, the revenue collected from all users, and the reward to the contributing users. A simple design involves returning a portion of the storage and bandwidth cost savings to contributing users in proportion to the effective storage and upload traffic contributed by each user. How to reliably obtain true data storage and uploading traffic statistics without too much overhead is a challenging problem. Collusion attacks present another problem, as some malicious users may collaborate to generate upload traffic and thus receive more reward.

Availability and Reliability: Users are willing to pay for cloud storage because of its high availability and reliability. The cloud service should not be interrupted by any kind of system failure or even natural disasters, and the users' data should never be corrupted. Studies on improving reliability and availability by adding data redundancy have been very active. The simplest form of introducing data redundancy is replication, i.e., storing multiple copies of the original data item, as exemplified by the Google File system [2] and Amazon's S3 [3]. Data replication is simple but inefficient. A more cost-effective way of achieving the same reliability with much less data redundancy is to apply erasure coding [4, 5]. As an example, Windows Azure Storage uses Reed-Solomon (RS) code to reduce the storage redundancy level from 3x to 1.3x-1.5x [5]. Under erasure coding, the original data item is partitioned into k blocks, from which n ($n>k$) encoded blocks are generated for distributed storage. RS code is an example of maximum distance separable (MDS) codes, which possess a regeneration property: any k out of the n encoded blocks can be used to recover the original data item. In a distributed environment, while storage overhead remains a critical concern for system efficiency, the amount of data transfer that is required to replace a lost storage node (i.e., the repair traffic) becomes equally important. With MDS codes, when repairing a lost node u by regenerating its storage onto a new node v, v must first download enough blocks to recover the original data item. Regeneration codes were recently invented based on the concept of network coding to minimize the repair traffic while holding the regeneration property of MDS codes [6-8].

Previous studies have focused on the design of a single coding scheme for the storage system. We argue that different coding schemes should be used at different levels in our user-assisted architecture to handle different types of failures. First, within a storage node with multiple disks, RAID5 or RAID6 are good candidates for handling one or two disk failures. Second, more powerful erasure or regeneration codes should be used across the data center to handle node failures, rack failures, and regular node updates. A good understanding of the pattern of node failures is very important. There exists a tradeoff between the level of data redundancy and the volume of repair traffic, so the network bandwidth between storage nodes should also be considered when designing the coding scheme. Third, to handle the failure of a whole data center, traditional erasure codes may not work well because the repair traffic is usually the same as the original data size (e.g., petascale) whereas the bandwidth between data centers is limited (e.g., gigascale). How to code the data and distribute them among data centers becomes an important optimization problem. Hu et al. recently proposed the application of functional minimum-storage regenerating (F-MSR) coding to handle two cloud failures while reducing repair traffic [9].

User Experience: The user's experience largely depends on how fast he can upload/download files. In traditional cloud storage, the end user transfers data to/from a data center, which may not be able to fully use the available bandwidth of the user's Internet access link. In our user-assisted architecture, because a user can simultaneously transfer data to/from many other users, there is a much greater chance that the available bandwidth is used to the fullest. How to encode and place data to improve overall uploading/downloading performance becomes a challenging research problem. Another factor worth considering is computational overhead. If a user's CPU is always busy handling the coding/decoding process, he may resist using the service. One promising solution is to offload the coding tasks to GPUs, which are widely available on desktop PCs and mobile devices. Our previous studies have shown that network coding can be practically implemented on contemporary GPUs with very high throughput [10, 11].

4. Conclusion

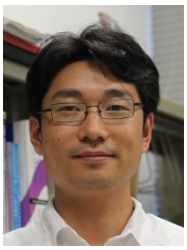
Cloud storage is a promising service that can offer economical and efficient solutions for highly reliable and accessible storage service. We propose using the spare storage and bandwidth resources of end users to save on service providers' storage and bandwidth costs.

How to design an effective incentive scheme, how to design different coding schemes at different levels, and how to enhance the user experience remain major challenges to this proposition.

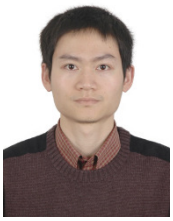
Acknowledgement
This work is supported by Hong Kong GRF grant HKBU 210412.

References

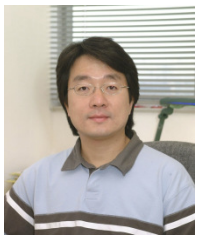
[1] X. Chen, Y. Jiang, and X.-W. Chu, "Measurements, Analysis and Modeling of Private Tracker," in Proc. IEEE P2P 2010.
[2] S. Ghemawat, H. Gobioff, and S. Leung, "The Google file system," in Proc. ACM SOSP 2003.
[3] DeCandia et al., "Dynamo: Amazon's Highly Available Key-value Store," in Proc. ACM SOSP 2007.
[4] R. Rodrigues and B. Liskov, "High availability in DHTs: Erasure Coding vs. Replication," in Proc. IPTPS 2005.
[5] C. Huang et al., "Erasure Coding in Windows Azure Storage," in Proc. USENIX ATC 2012.
[6] A. G. Dimakis et al., "Network Coding for Distributed Storage Systems," IEEE Transactions on Information Theory, vol. 58, no. 9, Sep 2010.
[7] A. G. Dimakis et al., "A Survey on Network Codes for Distributed Storage," IEEE Proceedings, vol. 99, no. 3, March 2011.
[8] K. W. Shum and Y. Hu, "Exact Minimum-Repair-Bandwidth Cooperative Regenerating Codes for Distributed Storage Systems," in Proc. IEEE ISIT 2011.
[9] Y. Hu et al., "NCCloud: Applying Network Coding for the Storage Repair in a Cloud-of-Clouds," in Proc. USENIX FAST 2012.
[10] X.-W. Chu, K. Zhao, and M. Wang, "Massively Parallel Network Coding on GPUs," in Proc. IEEE IPCCC 2008.
[11] X.-W. Chu, K. Zhao, and M. Wang, "Practical Random Linear Network Coding on GPUs," in Proc. IFIP Networking 2009.



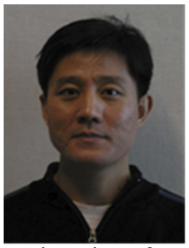
Xiaowen Chu received his B.E. degree in Computer Science from Tsinghua University, Beijing, P. R. China, in 1999, and the Ph.D. degree in the Computer Science from the Hong Kong University of Science and Technology in 2003. He is currently an Associate professor in the Department of Computer Science, Hong Kong Baptist University. His current research interests include Wireless Networks and Parallel and Distributed Computing.



Hai Liu received the BSc and MSc degrees in applied mathematics from South China University of Technology, in 1999 and 2002, respectively. He received the PhD degree in computer science from City University of Hong Kong in 2006. He is currently a research assistant professor with the Department of Computer Science, Hong Kong Baptist University. His research interests include wireless networking, mobile computing, and algorithm design and analysis.



Yiu-Wing Leung received the BSc and PhD degrees from the Chinese University of Hong Kong in 1989 and 1992, respectively. He has been working in the Department of Computer Science of the Hong Kong Baptist University and now he is a full professor. His research interests include two major areas: 1) networking and multimedia which include the design and optimization of wireless networks, optical networks and multimedia systems, and 2) cybernetics and systems engineering which include evolutionary computing and multiobjective programming. He has published more than 70 journal papers in these areas, and most of which were published in various IEEE journals. He is a senior member of the IEEE.



Zongpeng Li received his B.E. degree in Computer Science and Technology from Tsinghua University (Beijing) in 1999, his M.S. degree in Computer Science from University of Toronto in 2001, and his Ph.D. degree in Electrical and Computer Engineering from University of Toronto in 2005. Since August 2005, he has been with the Department of Computer Science in the University of Calgary. In 2011-2012, Zongpeng was a visitor at the Institute of Network Coding, Chinese University of Hong Kong. His research interests are in computer networks, particularly in network optimization, multicast algorithm design, network game theory and network coding.

Min Lei is a lecturer at the School of Computer Science at Beijing University of Posts and Telecommunications (BUPT). Min received his Ph.D. degree in information security from BUPT. Prior to that, he received a M.S. degree in software engineering and theory in 2002 from BUPT and a BEng degree in computer science in 1999 from Nanchang University. Min's research interests are watermarking, information hiding and information security.

INDUSTRIAL COLUMN: SPECIAL ISSUE ON "CROWDSOURCING-BASED MULTIMEDIA SYSTEMS"

Crowdsourcing-based Multimedia Systems

Guest Editor: Cheng-Hsin Hsu, National Tsing Hua University, Taiwan

chsu@cs.nthu.edu.tw

With the increasing prevalence of high-speed networks and surging growth of smartphones, large-scaled crowdsourcing-based multimedia systems with rich and almost real-time contents are becoming reality. Crowdsourcing is a distributed problem-solving and production paradigm. It allows companies and individuals to outsource tasks to general public. Crowdsourcing has been leveraged by various applications, including voting systems, information sharing systems, creative systems, and social games. Voting systems, such as Amazon Mechanical Turk, use online questionnaires to derive answers or opinions from the crowd. Information sharing systems allow the crowd to share knowledge, such as regional noise levels with others. Social games entertain players, e.g., smartphone-based geospatial tagging games require a player to move to specific locations tagged by other players to gain points.

This special issue of E-Letter covers the recent advances on large-scaled multimedia systems that leverage crowdsourcing to efficiently solve various problems that are hard for computers and algorithms. It is our honor to have five exciting articles from the leading research groups to present their latest results and shed some lights on the future research directions in crowdsourcing-based multimedia systems.

In the first article titled, "Crowdsourcing-Based Web Services for Speech and Music", Goto from the National Institute of Advanced Industrial Science and Technology present their successful crowdsoucing-based Web services: PodCastle and Songle. These two Web services leverage on the contributions from general public to improve the performance of automatic speech recognition and music understanding technologies, respectively. Such performance improvements benefit many multimedia applications, including content-based browsing of speech and music.

The second article is contributed by Gottlieb, Choi and Friedland from International Computer Science Institute, Berkley, and Kelm and Sikora from Technical University, Berlin. The article title is "On Pushing the Limits of Mechanical Turk: Qualifying the Crowd for Video Geolocation". In this article, the authors study how to find the best contributors to determine the geolocation of random videos downloaded from the Internet using Amazon Mechanical Turk. Such task is quite different from most problems addressed by crowdsourcing, as determining the geolocation of videos is difficult even for humans. The authors share their techniques on how to pick the highly-skilled contributors for difficult tasks.

The third article is contributed by Chen from Academia Sinica and Chu from National Chung Cheng University, titled "Crowdsourcing for Image Understanding Research". In this article, Chen and Chu present their Web-based storytelling system, called Pomics, which takes users' photos and automatically generate comic book drafts. Pomics provides interactive Web interface for users to revise the drafts. Pomics also collects users' editing behavior, image annotations, and comments, and hence it can be considered as a crowdsourcing system. Three potential image understanding research problems using Pomics are detailed in this article.

Foncubierta-Rodriguez and Muller, from University of Applied Sciences Western Switzerland, authored the fourth article, titled "Crowdsourcing Opportunities in Medical Imaging". This article presents the authors' experience on using crowdsourcing to develop ground truth for medical-related multimedia datasets. For example, the authors consider the problem of classifying a dataset of 3415 medical images into a modality hierarchy of 38 categories using feedbacks from 2470 contributors. They found that these non-experts can quickly build the ground truth at a small degradation on quality, when compared the resulting annotations against those from a small set of experts.

The last article of this special issue is contributed by Ooi from National University of Singapore, Marques from Florida Atlantic University, and Charvillat and Carlier from University of Toulouse. The paper title is "Pushing the Envelope: Solving Hard Multimedia Problems with Crowdsourcing". The authors consider the ultimate multimedia challenge of bridging the sematic gap, which is the difference between the captured raw data and human interpretations. In

particular, the authors postulate that crowdsourcing is much more powerful and is capable to solving many hard problems related to the sematic gap. Two sample problems recently solved by the authors are summarized: (i) identifying interesting regions of videos and (ii) generating text description of an image. The authors also present another two hard multimedia problems that remain open but could possibly be addressed with crowdsourcing.

We thank all the authors for their tremendous contributions. Their valuable experiences on utilizing crowdsourcing for multimedia research shed some lights on the future topics in this emerging research area, which may in turn lead to new challenges in multimedia communications. For example, mobile sensing using smartphones has been studied in the literature, but how to leveraging crowdsourcing for real-time, high-quality multimedia mobile sensing has not been thoroughly investigated. We hope this special issue will stimulate research on the crowdsourcing-based multimedia systems and accompanying multimedia communication problems.

Cheng-Hsin Hsu received the Ph.D. degree from Simon Fraser University, Canada in 2009, the M.Eng. degree from University of Maryland, College Park in 2003, and the M.Sc. and B.Sc. degrees from National Chung Cheng University, Taiwan in 2000 and 1996, respectively. He is an Assistant Professor in Department of Computer Science at National Tsing Hua University, Taiwan, since 2011. He was a Senior Research Scientist at Deutsche Telekom R&D Lab USA, Los Altos, CA between 2009 and 2011. Prior his Ph.D., Cheng-Hsin was with Lucent Technologies and Motorola. His research interests are in the area of multimedia networking and distributed systems. He and his colleagues won the Best Technical Demo Award in ACM MM'08, Best Paper Award in IEEE RTAS'12, and TAOS Best Paper Award in IEEE GLOBECOM'12. He served as the TPC Co-chair of the ACM MoVid'12 and MoVid'13 Workshops, the Proceedings and Web Chair of NOSSDAV'10, and the TPC members of several well-known conferences, including IEEE ICME, IEEE ICDCS, IEEE GLOBECOM, ACM MM, and ACM NOSSDAV.

Crowdsourcing-Based Web Services for Speech and Music

Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST), Japan
m.goto [at] aist.go.jp

Abstract
This column introduces two crowdsourcing-based multimedia systems, *PodCastle* (http://en.podcastle.jp for the English version and http://podcastle.jp for the Japanese version) and *Songle* (http://songle.jp). PodCastle and Songle collect voluntary contributions by anonymous users in order to improve the experiences of users listening to speech and music content available on the web. These multimedia systems, implemented as public web services, use automatic speech-recognition and music-understanding technologies to provide content analysis results, such as full-text speech transcriptions and music scene descriptions, that let users enjoy content-based multimedia retrieval and active browsing of speech and music signals without relying on metadata.

When automatic content analysis is used, however, errors are inevitable. PodCastle and Songle therefore provide an efficient error correction interface that let users easily correct errors by selecting from a list of candidate alternatives. Through these corrections, users gain a real sense of contributing for their own benefit and that of others and can be further motivated to contribute by seeing corrections made by other users.

Our web services promote the popularization and use of speech-recognition and music-understanding technologies by raising user awareness. Users can grasp the nature of those technologies just by seeing results obtained when the technologies applied to speech data and songs available on the web.

1. Introduction

Our goal is to provide end users with public web services based on speech recognition, music understanding, signal processing, machine learning, and crowdsourcing so that they can experience the benefits of state-of-the-art research-level technologies. Since the amount of speech and music data available on the web is always increasing, there are growing needs for the retrieval of this data. Unlike text data, however, the speech and music data itself cannot be used as an index for information retrieval. Although metadata or social tags are often put on speech and music, annotations such as categories or topics tend to be broad and insufficient for useful content-based information retrieval [1]. Furthermore, even if users can find their favorite content, listening to it takes time. Content-based active browsing that allows random access to a desired part of the content and facilitates deeper understanding of the content is important for improving the experiences of users listening to speech and music. We therefore developed two web services for speech and music, PodCastle (Figures 1 and 2) and Songle (Figures 3 and 4).

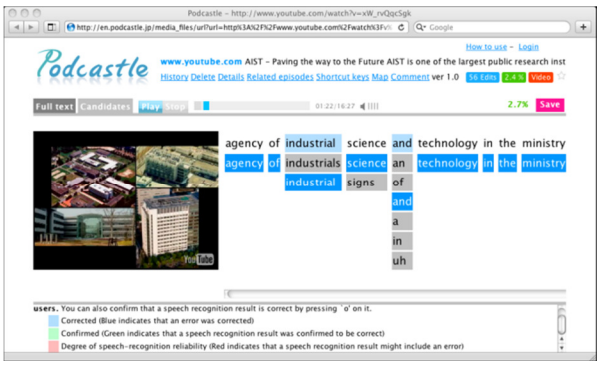


Figure 1. Screen snapshot of PodCastle's interface for correcting speech recognition errors. Competitive candidate alternatives are presented under the recognition results. A user corrected two errors in this excerpt by selecting from the

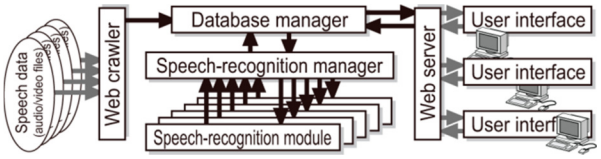


Figure 2. Implementation overview of PodCastle.

2. PodCastle

PodCastle (http://en.podcastle.jp for the English version and http://podcastle.jp for the Japanese version) [3–8, 10, 11] is a spoken document retrieval service that uses automatic speech recognition (ASR) technologies to provide full-text searching of the speech data in podcasts, individual audio or movie files on the web, and the video clips on the video sharing services (*YouTube*, *Nico Nico Douga*, and *Ustream.tv*). PodCastle enables users to find English and Japanese speech data including a search term, read full texts of their recognition results, and easily correct recognition errors by simply selecting from a list of candidate alternatives displayed on an error correction interface. The resulting corrections are used to improve the speech retrieval and recognition performance, and

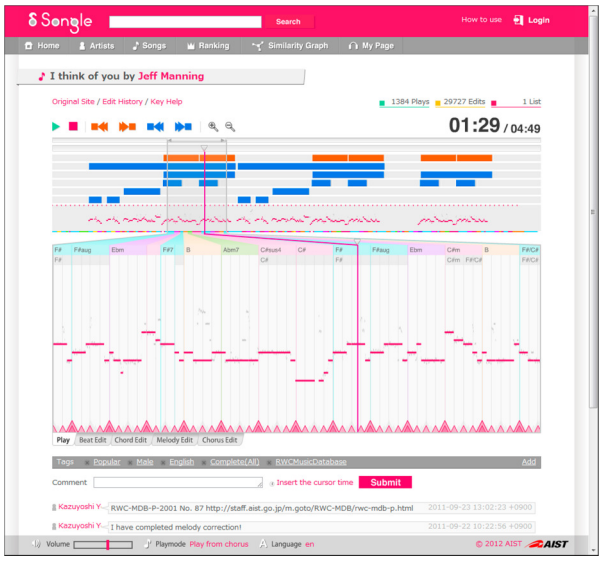


Figure 3. Screen snapshot of Songle's main interface for music playback with the visualization of automatically estimated music scene

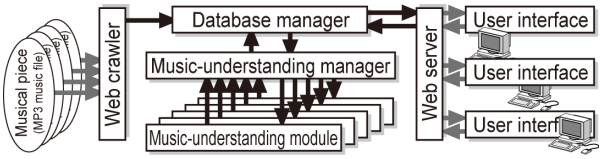


Figure 4. Implementation overview of Songle.

users can actively browse speech data by jumping to any word in the recognition results during playback. In our experience with its use over the past six years (since December 2006), over five hundred ninety thousand recognition errors were corrected by anonymous users and we confirmed that PodCastle's speech recognition performance was improved by those corrections.

3. Songle

Following the success of PodCastle, we launched Songle (http://songle.jp) [7–9], an active music listening service that enriches music listening experiences by using music-understanding technologies based on signal processing. Songle serves as a showcase, demonstrating how people can benefit from music-understanding technologies, by enabling people to experience active music listening interfaces [2] on the web. Songle facilitates deeper understanding of music by visualizing automatically estimated music scene descriptions such as music structure, hierarchical beat structure, melody line, and chords (Figure 3). Users can actively browse music data by jumping to a chorus or repeated section during playback and can use a content-based retrieval function to find music with similar vocal timbres. Songle also features an efficient error correction interface that encourages people to help improve Songle by correcting estimation errors (Figure 5).

4. Conclusion

PodCastle and Songle made academic contributions by demonstrating a new research approach to speech recognition and music understanding based on signal processing; this approach aims to improve the speech-recognition and music-understanding performances as well as the usage rates while benefiting from the cooperation of anonymous end users. This approach is designed to set into motion a *positive spiral* where (1) we enable users to experience a service based on speech recognition or music understanding to let them better understand its performance, (2) users contribute to improving performance, and (3) the improved performance leads to a better user experience, which encourages further use of the service at step (1) of this spiral. This is a *social correction* framework, where users can improve the performance by sharing their correction results over a web service. The game-based approach of Human Computation or GWAPs (games with a purpose) [13] like the ESP Game [14] often lacks step (3) and depends on the feeling of fun. In this framework, users gain a real sense of contributing for their own benefit and that of others and can be further motivated to contribute by seeing corrections made by other users. In this way, we can use the *wisdom of the crowd* or *crowdsourcing* [12] to achieve a better user experience.

Acknowledgments

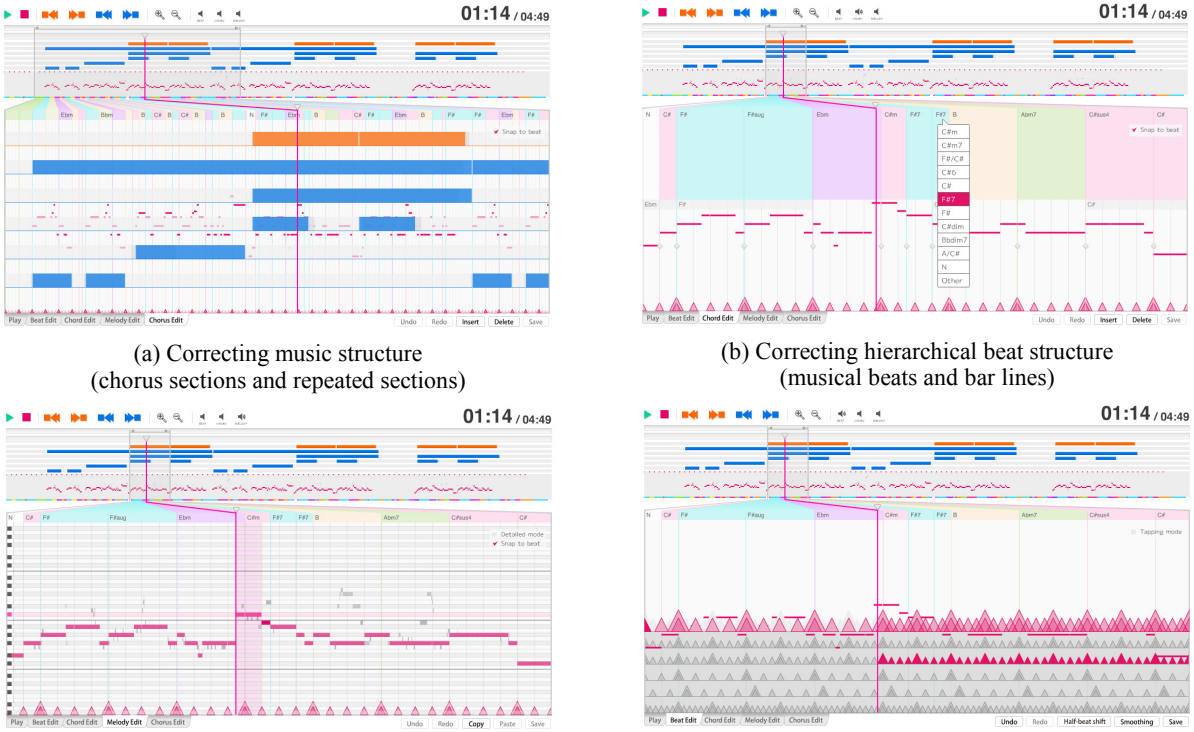
We thank Jun Ogata who collaborates with me for PodCastle, and Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano who collaborate with me for Songle. We also thank Youhei Sawada, Shunichi Arai, Kouichirou Eto, and Ryutaro Kamitsu for their web service implementation of PodCastle, Utah Kawasaki for the web service implementation of Songle, and Minoru Sakurai for the web design of PodCastle and Songle. We thank anonymous users of PodCastle and Songle for correcting errors. This work was supported in part by CREST, JST.

References

[1] M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. *Content-based music information retrieval: Current directions and future challenges*. Proceedings of the IEEE, 96(4):668-696, 2008.

[2] M. Goto. *Active music listening interfaces based on signal processing*. In Proc. of IEEE ICASSP 2007, 2007.

[3] M. Goto and J. Ogata. **[Invited talk]** *PodCastle: A spoken document retrieval service improved by anonymous user contributions*. In Proc. of PACLIC 24,



(a) Correcting music structure
(chorus sections and repeated sections)

(b) Correcting hierarchical beat structure
(musical beats and bar lines)

(c) Correcting melody line (F0 of the vocal melody)

(d) Correcting chords (root note and chord type)

Figure 5. Screen snapshots of Songle's error correction interface for correcting music scene descriptions.

pages 3-11, 2010.

[4] M. Goto and J. Ogata. **[Invited talk]** *PodCastle: A spoken document retrieval service improved by user contributions*. In Proc. of KJDB 2010, 2010.

[5] M. Goto and J. Ogata. *PodCastle: Recent advances of a spoken document retrieval service improved by anonymous user contributions*. In Proc. of Interspeech 2011, 2011.

[6] M. Goto, J. Ogata, and K. Eto. *PodCastle: A Web 2.0 approach to speech recognition research*. In Proc. of Interspeech 2007, 2007.

[7] M. Goto, J. Ogata, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. **[Keynote talk]** *PodCastle and Songle: Crowdsourcing-based web services for spoken content retrieval and active music listening*. In Proc. of ACM CrowdMM 2012, pages 1-2, 2012.

[8] M. Goto, J. Ogata, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. *PodCastle and Songle: Crowdsourcing-based web services for retrieval and browsing of speech and music content*. In Proc. of CrowdSearch 2012, pages 36-41, 2012.

[9] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. *Songle: A web service for active music listening improved by user contributions*. In Proc. of ISMIR 2011, pages 311-316, 2011.

[10] J. Ogata and M. Goto. *PodCastle: Collaborative training of acoustic models on the basis of wisdom of crowds for podcast transcription*. In Proc. of Interspeech 2009, pages 1491-1494, 2009.

[11] J. Ogata, M. Goto, and K. Eto. *Automatic transcription for a Web 2.0 service to search podcasts*. In Proc. of Interspeech 2007, 2007.

[12] G. Parent and M. Eskenazi. *Speaking to the Crowd: Looking at past achievements in using crowdsourcing for speech and predicting future challenges*. In Proc. of Interspeech 2011, 2011.

[13] L. von Ahn. *Games with a purpose*. IEEE Computer Magazine, 39(6):92-94, June 2006.

[14] L. von Ahn and L. Dabbish. *Labeling images with a computer game*. In Proc. of ACM CHI 2004, pages 319-326, 2004.

Masataka Goto received the Doctor of Engineering degree from Waseda University in 1998. He is currently a Prime Senior Researcher and the Leader of the Media Interaction Group at the National Institute of Advanced Industrial Science and Technology (AIST), Japan. He serves concurrently as a Visiting Professor at the Institute of Statistical Mathematics, an Associate Professor (Cooperative Graduate School Program) in the Graduate School of Systems and Information Engineering, University of Tsukuba, and a Project Manager of the Exploratory IT Human Resources Project run by the Information Technology Promotion Agency (IPA), Japan. Over the past 20 years, Masataka Goto has published more than 190 papers in refereed

journals and international conferences and has received 31 awards, including several best paper awards, best presentation awards, and the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology (Young Scientists' Prize).

On Pushing the Limits of Mechanical Turk: Qualifying the Crowd for Video Geolocation

Luke Gottlieb, Jaeyoung Choi, Gerald Friedland
International Computer Science Institute
{luke, jaeyoung, fractor}@icsi.berkeley.edu

Pascal Kelm, Thomas Sikora
Communication System Group
Technische Universität Berlin
{kelm, sikora}@nue.tu-berlin.de

1. Introduction

This work was first appeared in Gottlieb et al. [1]. In this article we summarize the methods we took for finding skilled Mechanical Turk participants for our annotation task, which will be to determine the geo-location of random videos from the web. The task itself is unlike the standard setup for a Mechanical Turk task, in that it is difficult for both humans and machines, whereas a standard Mechanical Turk task is usually easy for humans and difficult or impossible for machines. There are several notable challenges to finding skilled workers for this task: First, we must find what we termed "honest operators", i.e., people who will seriously attempt to do the task and not just click quickly through it to collect the bounty. Second, we need to develop meaningful qualification test set(s) that are challenging enough to allow us to qualify people for the real task, but were also solvable by individuals regardless of their culture or location, although English language understanding was required for instructions.

2. Qualification Task Setup

For this experiment we used the dataset of the Placing Task of the MediaEval benchmark. The MediaEval Placing Task 2010 data set consists of Creative Common-licensed Flickr videos. The metadata for each video includes user-annotated title, tags, description, and also information about the user who uploaded the videos. According to [2], videos were selected both to provide a broad coverage of users, and also because they were geotagged with a high accuracy at the "street level". Accuracy shows the zoom level the user used when placing the photo on the map.

The setup of this task had two important parts, the selection of videos and the design and deployment of the Mechanical Turk user interface. The task of video selection was relatively straightforward, although during the process of selecting videos we had to make several important decisions on the types of videos which were useful in the qualification task. First, we randomized the complete list of videos, then our annotator viewed a subset of that list, and attempted to determine the location that the video presented. The annotator was allowed to use video and audio information, but not meta-tags, and was instructed to spend no more than 5 minutes per video. From this we collected the initial 40 videos which we used in our initial approach. Our discoveries there led us to take a subset of those videos, which we used in our revised approach. In condensing the videos we tried to reduce the requirement for information from worker's previous experience as much as possible, e.g., in the initial set there were videos of people in Machu Picchu, which our annotator immediately recognized, however there were no clues to reveal this location that would be usable to someone who had not heard or seen this location previously.

The next step of out setup was the development of a user interface for the qualification task. We went through several rounds of internal testing and feedback to enhance the usability of the tool. One of our more important discoveries was how the addition of a tutorial greatly aided the workers.
At the end of the workers task, we asked participants to leave comments about the task. We updated the interface to reflect the feedback about the usability, and will be using the information that we received to make further improvements to the interface as we move forward with this project.

3. Evaluation

To evaluate the performance of the online workers, the geodesic distance between the ground truth coordinates and those of the outputs from participants were compared. To take into account the geographic nature of the evaluation, the Haversine distance was used. This measure is calculated thus:

$$d = 2 \cdot r \cdot \arcsin(\sqrt{h})$$

$$h = \sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\psi_2 - \psi_1}{2}\right)$$

where d is the distance between points 1 and 2 represented as latitude ($\varphi 1$, $\varphi 2$) and longitude ($\psi 1$, $\psi 2$) and r is the radius of the Earth (in this case, the WGS-84 standard value of 6,378.137 km was used).

4. Initial Approach

In our initial approach to the qualification task we created four randomly selected subsets of our 40 videos. We then asked internal volunteers to attempt the task, as a "baseline baseline," to give us some expectation of how well a Mechanical Turk worker might be able to perform, so that we could set a qualification threshold

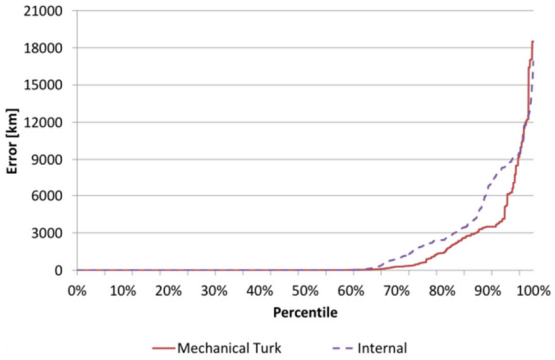


Figure 1. Initial comparison of internal workers and Mechanical Turk workers with 40 videos.

for the actual task. After several rounds of internal tests and then experimenting on Mechanical Turk, we discovered that by making random sets we had not taken into account that the videos to be classified would be of varying degrees of difficulty, thus while we could compare the performance on a per video basis, we could not provide a threshold for qualification in the general sense.

Figure 1 shows the performance of our internal testers and the Mechanical Turk workers. While some of the Turkers did relatively well, there were many who did not seem to understand the task, and apparently guessed at random. We also monitored the time it took for the worker to make a guess, and in our performance analysis we eliminated the outliers who were clearly attempting to speed through the task for the bounty without making a legitimate attempt. We also rejected submissions where all the videos had wildly inaccurate answers, as some of the selections were quite easy, and served as a gold standard for the training sets.

5. Revised Approach

For our second round of qualification attempts we revised our approach based upon the results of the first attempt. We created a tutorial page that provides a walkthrough showing how to locate a video that most of the workers, both internal and external, did quit-poorly on. We presented the relevant frame that workers could use to determine the location the video was filmed. A Google image search of "City Market Hall" brings up a large number of images, and following some of those links will lead you to pages for the city of Roanoke, Virginia. A Further search of Google maps will give the exact latitude and longitude of the building in question. We abandoned the use of 4 randomized sets, and hand selected 10 videos for a new qualification test. Doing this allowed us to compare our qualification results on a set basis and have a more precise threshhold for qualification.

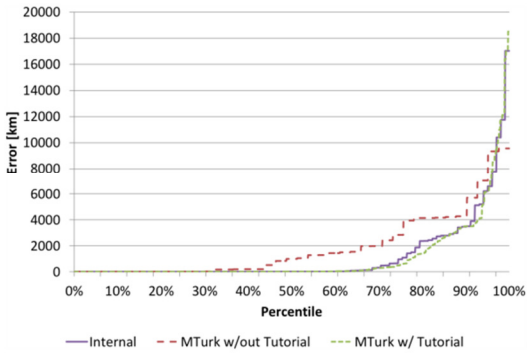


Figure 2. Performance results for revised approach.

Figure 2 presents the comparison of the performance results for our internal testers, the initial test results, and the tutorial. We can see then, that while the internal testers still perform better than the Turkers, the addition of the tutorial narrows the performance margin. Ultimately the purpose of this experiment was the qualification of annotators for our ongoing annotation task of uncategorized videos. As we are attempting to produce a baseline for comparison to our automated geolocation systems, we require very high accuracy, thus we set the threshold for qualification at 80% accuracy. To be scored as getting a correct answer, the video had to be geolocated with 10km of its posted location. This requirement meant that 16% of participants were qualified to do the actual task. After eliminating the individuals who tried to get the bounty without seriously attempting the task, our acceptance threshold went up to 19%.

4. Conclusion

Since we have very large pool of workers on Mechanical Turk, we are able to qualify only those who are able to achieve very high accuracy (19% of our workers had a at least 80% accuracy, which was our threshold for qualification) so that during the task of estimating the location of unknown videos, we can be reasonably certain that they will complete this difficult task with a high degree of accuracy. The bottom line is, it is possible to use crowd sourcing for a very difficult task but one has to be highly careful in the selection of the crowd.

We believe that the techniques that we have developed here are applicable to other tasks that require highly skilled crowd-sourced workers. First, it is important to generate a high quality baseline to use for the qualification task. This baseline need not be very large, but it should be created by people who have a good understanding of what the project is. Second, you need to conduct several rounds of internal testing using traditional workers. This will allow you to have feedback on what is confusing about the project in

order to develop a tutorial for the crowd workers, and have a baseline of how motivated workers perform. Third, deploy the task to the actual crowd based workers, seek feedback from them, and compare results to the baseline and motivated workers. Use this information to improve the tutorial, and then repeat and refine until you have a sufficient number of workers who reach your qualification threshold.

5. Acknowledgments

This research is supported in part by NSF EAGER grant IIS-1128599 and KFAS Doctoral Study Abroad Fellowship. Human subjects experiments authorized under IRB approval CPHS 2011-06-3325. We acknowledge Howard Lei for his aid in this work.

References

[1] L. Gottlieb, J. Choi, P. Kelm, T. Sikora, and G. Friedland. Pushing the Limits of Mechanical Turk: Qualifying the Crowd for Video Geo-Location. To appear in the proceedings of the ACM Workshop on Crowdsourcing for Multimedia (CrowdMM 2012), held in conjunction with ACM Multimedia 2012, Nara, Japan, October 2012.

[2] M. Larson, M. Soleymani, P. Serdyukov, S. Rudinac, C. Wartena, V. Murdock, G. Friedland, R. Ordelman, and G. J. Jones. Automatic Tagging and Geo-Tagging in Video Collections and Communities. In ACM International Conference on Multimedia Retrieval (ICMR 2011), pages 51:1–51:8, April 2011.

Luke Gottlieb received a Bachelor of Arts Degree in Linguistics from The University of California, Berkeley in 2004. He has worked at the International Computer Science Institute since 2004 as a Research Assistant. His research interests include speech processing, multimedia information retrieval, geolocation and privacy. In 2009 his work, along with Gerald Friedland and Adam Janin on Joke-o-Mat which won ACM MultiMedia's grand challenge.

Jaeyoung Choi is a research assistant at the International Computer Science Institute, a private research lab affiliated with the University of California, Berkeley, where he works in a multimedia group, focusing on merging visual, acoustic and natural language processing techniques for large-scale multimedia retrieval and online privacy issues arising from the retrieval technology. He holds a B.S. in Computer Science from KAIST and a M.S. in Computer Science from University of California, Berkeley.

Gerald Friedland is heading the audio and multimedia group at the International Computer Science Institute, a private non-profit research lab affiliated with the University of California, Berkeley, where he leads research mostly focusing on acoustic and multimodal methods for large scale video analysis but also on privacy and crowdsourcing. Dr. Friedland has published more than 100 peer-reviewed articles in conferences, journals, and books and is and Associate Editor for ACM Transactions on Multimedia Computing, Communications, and Applications. Most recently, he lead the team that won the ACM Multimedia Grand Challenge in 2009. Dr. Friedland received his doctorate (summa cum laude) and master's degree in computer science from Freie Universitaet Berlin, Germany, in 2002 and 2006, respectively.

Pascal Kelm is a PhD student at the Communication Systems Group (Technische Universitaet of Berlin) and received his Dipl.-Ing. degree in electrical engineering from TU Berlin, in 2009. His research expertises include multimedia analysis and retrieval, machine learning and multimodal fusion techniques for automatic geotagging in social media. He has actively participated in several EU funded multimedia oriented research projects including Petamedia, OpenSEM, VideoSense and is organizing the Placing task in the MediaEval benchmarking initiative.

Thomas Sikora is professor and director of the Communication Systems Group (in German: Fachgebiet Nachrichtenübertra-gung) at Technische Universitaet Berlin, Germany. He received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Bremen University, Bremen, Germany, in 1985 and 1989, respectively. In 1990, he joined Siemens Ltd. and Monash University, Melbourne, Australia, as a Project Leader responsible for video compression research activities in the Australian Universal Broadband Video Codec consortium. Between 1994 and 2001, he was the Director of the Interactive Media Department, Heinrich Hertz Institute (HHI) Berlin GmbH, Germany. Dr. Sikora is co-founder of 2SK Media Technologies and Vis-a-Pix GmbH, two Berlin-based start-up companies involved in research and development of audio and video signal processing and compression technology.

Crowdsourcing for Image Understanding Research

Kuan-Ta Chen[1] and Wei-Ta Chu[2]
[1]Institute of Information Science, Academia Sinica
[2]Dept. of Computer Science and Information Engineering, National Chung Cheng University
swc@iis.sinica.edu.tw, wtchu@cs.ccu.edu.tw

1. Introduction

Recently, people are getting used to rely on photo browsing, photo/video slideshow, and illustrated text to share their lives in pictures. However, these presentation schemes may require high production efforts (video slideshow and illustrated text), lack rich and clear expression (photo browsing and photo slideshow), impose limited user control (photo and video slideshow), and have poor ubiquitousness (except for illustrated text, all the other media forms cannot be printed on a paper for reading anytime, anywhere). Comics, on the other hand, are considered to be a potential medium for visual storytelling because of rich expressivity, high interactivity, and medium portability. The rich expressivity is due to the fact that relative importance of photographs and the story's progression shown by photographs can be expressed with various frame sizes and shapes (e.g., rectangles, quasi-squares, and trapezoids). A picture depicting a critical moment can be given larger space, and an action sequence may be framed in matching trapezoids. Comics often use zig-zag reading lines to guide the reader. This style is more interesting and more efficient than slideshows presented in a linear manner. Thus, comics *can be conveyed by any display medium* without information loss. Moreover, readers have *full control over their reading pace and target* as no temporal restrictions are imposed. In terms of the reader's literacy level, available time, and patience, comics require *less effort* than illustrated text because a significant part of the information is conveyed in pictorial form.

Nevertheless, comic creation is not an easy task, especially for the storyboarding and layout planning phases. Storyboarding requires creators to select the most informative and representative photos; layout planning involves arranging photos (i.e., frames) on a fixed rectangular page, where a frame's size and shape are related to its importance and visual content. In addition, editing a comic's two-dimensional layout is much more challenging than editing one-dimensional content, such as a slideshow.

To address these challenges, an online system, called Pomics [1], is developed to simplify the comic storytelling process. Pomics provides an automated pictorial storytelling module that can generate comic book drafts fully automatically and also provides friendly interactive interface for users to revise the drafts at their wills. Meanwhile, Pomics is able to collect users' editing behaviors, image annotations, and comments, and thus can implicitly behave as a crowdsourcing platform. In the rest of this article, we will describe the proof-of-concept implementation of the Pomics system and how it can contribute to the image understanding research community.

2. The Pomics system

To realize the proposed storytelling system, we have developed a proof-of-concept implementation that includes two phases: photo scoring and comic editing [1]. When a set of photos is loaded, the system automatically assigns a score to each photo according to the following rules. A photo is deemed representative if it 1) contains people, 2) contains more than one person, 3) is one of a series of shots, 4) shows a new location, and 5) is with acceptable exposure. Successive shots and location changes are determined from time and exposure information in the EXIF records. A color histogram of the pixels is used to judge whether the exposure setting is reasonable.

When the user is satisfied with the computed scores, he/she can set a desired number of comic pages, *k*, and press the "Generate" button to generate a comic. No matter how many photos the user provides, the system creates *k* comic pages using the most representative pictures. The user can then refine the comics in the Pomics Editor by adding onomatopoeias, revising the dialogues and narratives, altering panel borders, rotating pictures, and even replacing the pictures if the chosen pictures are not preferred. As an example, in Figure 1, we show two sample pages that were generated automatically from a set of 60 travel photos. The pages were not edited manually, so the quality could certainly be improved by some degree of fine-tuning.

[1] http://www.pomics.net

Figure 1. Sample auto-generated comics.

3. Implications on Image Understanding Research

From the viewpoint of users, Pomics is a comics-based storytelling system; at the same time, Pomics is able to implicitly collect a variety of semantics information about pictures, such as user's aesthetics preferences among photos, which part(s) of an image is more significant, an image's semantic meanings, and the emotions associated to an image (based on the emoticons used). In the following, we briefly describe implications of Pomics user editing actions on various research topics.

Image selection and summarization.

Users are allowed to add or drop photos at their wills to shape the auto-generated comics. The selected photos may present representative scenes, important events and important people, human faces with specific emotions, and so on. To the best of our knowledge, there is no public dataset available for image summarization research. We plan to collect datasets on how users (subjectively) select representative photos from an image pool in order to compile a compact presentation as a story.

Image aesthetics analysis.

Users are allowed to assign each image an aesthetic score (from 1 to 10) to denote the aesthetic rating of this image. Figure 2 shows some pictures and their corresponding aesthetic scores provided by a Pomics user. Comparing with existing image aesthetics datasets [2] and websites like *dpchallenge*[2], images collected through Pomics normally consist of more human faces, specific events, and possess subtle preferences for specific subjects, e.g., a cute child. Therefore, more cues beyond content-based features (such as the color histogram) and compositional features (such as the rule of thirds) would be covered in the dataset of Pomics.

Image saliency analysis.

Pomics allows users to adjust each comic frame's location and size, and displace the photo inside a frame, so that the most interesting (or important) region of the photo can be appropriately shown. This behavior implicitly provides information about the region of interest (ROI) of a picture, and through accumulating many users' interests, the most salient part of a photo can be defined. Figure 3 provides two examples. The original photos (left) are used a number of times by different Pomics users and their visible areas (i.e., not masked by a comic frame) are aggregated and shown on the right side of Figure 3. Regions with less shadow indicate those parts more favorable and therefore considered more salient.

From the aforementioned examples, the dataset collected through Pomics can be utilized in various research topics, and often more faithfully reflects real situations in daily life. This shows the strength of Pomics as a crowdsourcing platform that implicitly collects human's wisdom without intervention.

Figure 2. Sample images associated with aesthetic scores.


Figure 3. Left: original photos; right: accumulated masks (higher intensity means higher saliency).

4. Conclusion

In this article, we present an online system Pomics that provides highly interactive and enjoyable editing functions to accomplish photo storytelling, with the presentation of comics. Implicitly acting as a crowdsourcing platform, photos and editing behavior collected by this system can be utilized to advance image understand research from various perspectives, such as image summarization, saliency, and semantics analysis.

[2] http://www.dpchallenge.com/

References

[1] Ming-Hui Wen, Ruck Thawonmas, and Kuan-Ta Chen, "Pomics: A Computer-aided Storytelling System with Automatic Picture-to-Comics Composition", Proceedings of TAAI 2012, Nov 2012

[2] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *CVPR*, pp. 2408-2415, 2012.

Kuan-Ta Chen is an Associate Research Fellow at the Institute of Information Science and the Research Center for Information Technology Innovation (joint appointment) of Academia Sinica. Dr. Chen received his Ph.D. in Electrical Engineering from National Taiwan University in 2006, and received his B.S. and M.S. in Computer Science from National Tsing-Hua University in 1998 and 2000, respectively. His research interests include QoE management, multimedia systems and networking, online gaming, and crowdsourcing. He has been an Associate Editor of IEEE Transactions on Multimedia since 2011. He is a member of ACM, IEEE, IICM, and CCISA.

Wei-Ta Chu received B.S. and M.S. in Computer Science from National Chi Nan University, Taiwan, in 2000 and 2002, and received Ph.D. in Computer Science from National Taiwan University, Taiwan, in 2006. His research interests include digital content analysis, multimedia indexing, digital signal process, and pattern recognition. He was awarded the Young Faculty Awards presented by National Chung Cheng University in 2011. He won the Best Full Technical Paper Award in ACM Multimedia 2006. He was a visiting scholar at Digital Video & Multimedia Laboratory, Columbia University, from July to August 2008.

Crowdsourcing Opportunities in Medical Imaging

Antonio Foncubierta-Rodríguez and Henning Müller
University of Applied Sciences Western Switzerland (HES-SO)
{antonio.foncubierta, henning.mueller}@hevs.ch

1. Introduction

The production of medical images in clinical practice has been growing quickly over the past decades [1] [2]. Medical images amount for 30% of the global storage in 2010 according to [3] and mammography in the US alone to over 2.5 Petabytes in 2009. In addition to medical images stored in clinical Picture Archiving and Communication Systems (PACS), medical images are frequently used in the biomedical literature, carrying much information in connection with the associated text. Despite the valuable amount of information stored, images are seldom used more than once in clinical practice and image content is very rarely analyzed to facilitate reuse. Improving accessibility to medical images both in clinical and research environments can provide clinicians, trainees and researchers with additional and valuable tools in their daily work. Computer assisted indexing, classification and retrieval can successfully improve the accessibility of relevant images from the medical literature and reuse of clinical images for clinical decision support. However, these tools often require large training sets to deliver a convincing performance. For optimal, generalizable results, the size of these sets needs to be large and representative of the actual class distribution found in real-world data. Ground truth generation of large datasets is a tedious, repetitive task that is costly and time-consuming, and might require specialists to perform the annotation.

Crowdsourcing has received much attention in the past years, and has been proposed several times to solve challenging problems by using the so-called wisdom of the crowd [4]. In this paper, a medical image crowdsourcing-based ground truth generation method that reduces the manual interaction as much as possible is discussed, together with several ideas for crowdsourcing in medical imaging.

2. Crowdsourcing-Based Ground Truth Generation

Although crowdsourcing has previously been used for obtaining definite solutions to challenging scientific problems [4], it has also been proposed as a way of obtaining reliable, comprehensive ground truth in a time-effective way with limited costs. One of the first examples of ground truth generation was the ESP Game (Edd Smith's People), developed by von Ahn [5] and later acquired by Google to improve their image search engine. The game consisted of presenting two

Figure 2 LabelMe interface for image labeling and annotation.

players with an image and requiring them to assign labels or tags to it. If the two players agreed on a label, the label would be assigned to the image. The same author developed later a game-based crowdsourcing platform, called Game With A Purpose (GWAP[3]), that proposed other image-related games, including classification, manual segmentation, labeling, etc. The term GWAP has also become popular in the game community [6]. Similar games have also been deployed, such as LabelMe[4] [7], which makes it possible to use a labeling and segmentation framework on user-uploaded datasets within crowdsourcing platforms such as Amazon Mechanical Turk, as shown in Figure 1.

When no game-like application is developed, crowdsourcing platforms like Amazon Mechanical Turk, Crowdflower, Zoombucks, etc. offer contributors the possibility to earn money for their work. The amounts paid are decided by the task creators according to the amount of units produced, correct units produced or even bonuses for extraordinary work (in terms of quality or quantity). With Amazon Mechanical Turk being the most popular platform, most of the ground-truth work is based on their infrastructure [8]. However, the requirement of a US-based credit card might be blocking some research groups from using crowdsourcing. When the tasks are better suited for a

[3] http://www.gwap.com/
[4] http://labelme.csail.mit.edu/

group of people with specific knowledge, crowdsourcing contributors can be filtered-out within a selection phase to control high quality of responses before the actual task begins.

3. Opportunities of Crowdsourcing for the Medical Image Domain

Although medical images are an important area of interest for the information retrieval community, there has been little effort in using crowdsourcing to develop ground truth for medical-related datasets. Special knowledge and the requirement to get high quality annotations also make crowdsourcing difficult in this domain.

In [9] we propose that ground truth generation can be quickly performed by non-experts after an initial training phase and with good explications. The study consisted of classifying a dataset of 3415 images of the biomedical literature into a modality hierarchy of 38 categories [10] as shown in Figure 2. The intent of classifying these images was to use them as ground truth for the medical task in the ImageCLEF [5] benchmarking event. In order to evaluate the quality and speed of the annotations, three user groups where compared: a) a medical doctor manually classified all 3415 images, providing a gold standard for the crowdsourcing experiments, b) a set of 18 known experts with experience in medical imaging classified the dataset and c) 2470 contributors from open crowdsourcing. The study also included an iterative approach, where images from a smaller training set were used to train an automatic system that was later evaluated using crowdsourcing, asking the contributors to accept or reject the automatic classification.

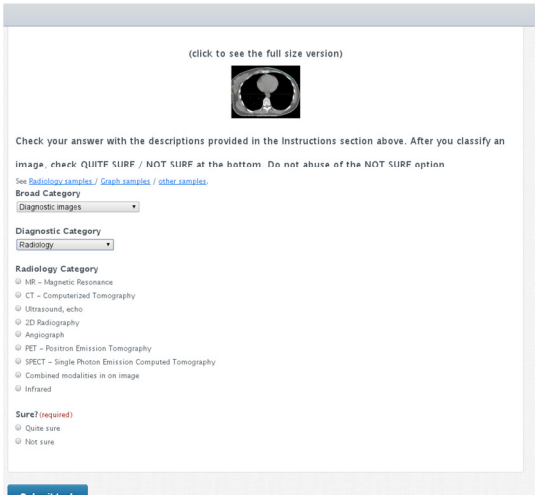


Figure 3 Screenshot of the crowdsourcing interface for modality classification described in [9].

The study concluded that non-experts can quickly build the ground truth at a limited cost of quality, which strongly varies among categories when comparing the annotations of the crowd to those from a set of known experts. The study resulted in several outcomes

- strict quality control and thus a good gold standard is necessary to obtain meaningful results (in our case 50% of the proposed images had a known ground truth and judges below 80% quality were automatically removed);
- very good explications and a tutorial are necessary to make the experiment work. Users need to know precisely what is expected from them;
- to increase the speed of annotations, many users are required, but users that contribute more to the system become more familiar and therefore provide better annotations, so frequent users need to be favored;
- complex tasks still take much time and demotivate users; simplifying tasks produces better (higher inter-rater agreement) and faster results (yes/no questions were answered twice as fast as a simple classification);
- a multi-step approach where system output can be validated or refused by *crowdsourcers* could make it scalable to big data.

One of the advantages of crowdsourcing is that it allows a very large number of users to annotate data (in our case more than 2400 persons tested the system in two weeks). For instance, image retrieval often relies on the concept of visual similarity, which is related to many aspects and can be defined in various ways (subjectiveness). A large-scale experiment with *crowdsourcers* allows inferring a visual similarity model from several users' understanding and creating a solid ground truth for visual similarity evaluation. Other initiatives like the VISCERAL project[6] propose the creation of a silver corpus, where the results with high variance can be further analyzed manually, potentially using crowd sourcing or specialist annotators.

4. Conclusion

Crowdsourcing is a promising tool for scientific research, specifically in multimedia environments. However, there is the risk of abandoning machine learning techniques in favor of crowdsourcing in some cases. This should be carefully analyzed in order to obtain the best from both worlds.

Medical imaging has been underusing crowdsourcing

[5] http://www.imageclef.org/

[6] http://www.visceral.eu/

techniques, but this might change in the future when models for strict quality control and high-quality tutorials for the tasks are delivering convincing results.

It is difficult to predict the background of the contributors, as well as the outcome of their work in crowdsourcing environments. There has been much effort in getting rid of untrusted or *incorrect* annotations, however, rather than blocking their participation in the system, stronger methods to learn from them need to be defined, because a consistent trend of wrongly classified images from a certain group of contributors is showing a pattern of how things can be perceived, and can be the key to redefining concepts that have been understood as static for a long time.

Crowdsourcing can become an important tool in many domains but it is not a one shot thing but an iterative process. Analyzing first results, leaning from them and then redefining tasks are important. Quality control is equally essential for obtain good outcomes to develop many of tomorrow computer-based tools based on human intelligence.

References

[1] K. P. Andriole, J. M. Wolfe and R. Khorasani, "Optimizing Analysis, Visualization and Navigation of Large Image Data Sets: One 5000-Section CT Scan can ruin your whole day," Radiology, vol. 259, no. 2, pp. 346-362, 2011.

[2] H. D. Tagare, C. Jaffe and J. Duncan, "Medical Image Databases: A Content--Based Retrieval Approach," JAMIA, vol. 4, no. 3, pp. 184-198, 1997.

[3] "Riding the wave: How Europe can gain from the rising tide of scientific data," Submission to the European Comission. 2010. [Online]. Available: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf.

[4] F. Khatib, F. DiMaio, F. C. Group, F. V. C. Group, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, M. Jaskolski and D. Baker, "Crystal structure of a monomeric retroviral protease solved by protein folding game players," Nature Structural & Molecular Biology, no. 18, pp. 1175-1177, 2011.

[5] L. von Ahn and L. Dabbish, "Labelling images with a computer game," in *SIGCHI conference on Human factors in Computing Systems (CHI'04)*, New York, USA, 2004.

[6] B. Steinmayr, C. Wieser, F. Kneißl and F. Bry, "Karido: A GWAP for Telling Artworks Apart," in The 16th International Conference on Computer Games (CGAMES2011), Kentucky, USA, 2011.

[7] B. C. Russel, A. Torralba, K. P. Murphy and W. T. Freeman, "LabelMe: a database and web-based tool for image annotation," International Journal of Computer Vision, vol. 77, no. 1-3, pp. 157-173, 2008.

[8] O. Alonso and S. Mizzaro, "Can we get rid of TREC assessors? Using Mechanical Turk for relevance assessment," in SIGIR 2009 Workshop on The Future of IR Evaluation, Boston, USA, 2009.

[9] A. Foncubierta-Rodríguez and H. Müller, "Ground Truth Generation in Medical Imaging: A Crowdsourcing-based Iterative Approach," in ACM multimedia 2012 workshop on Crowdsourcing for multimedia (CrowdMM'12), Nara, Japan, 2012.

[10] H. Müller, J. Kalpathy-Cramer, D. Demner-Fushman and S. Antani, "Creating a classification of image types in the medical literature for visual categorization," in SPIE Medical Imaging, San Diego, USA, 2012.

Antonio Foncubierta-Rodríguez received the M.Eng. degree in telecommunication engineering at the University of Seville, Spain in 2009. In 2007, he worked part-time as a researcher for the Department of Communications and Signal Processing in the University of Seville. His research was related to video compression and transmission over mobile networks, leading to a master's thesis. In 2008 he worked in a project on medical image retrieval for the University Hospitals Virgen del Rocío in Seville. Currently, as a PhD Student at the University of Geneva, he is a research assistant at University of Applied Sciences Western Switzerland in Sierre, where he works on several Swiss national and EU projects.

Henning Müller studied medical informatics at the University of Heidelberg from 1992–1997. After a diploma in telemedicine he worked at Daimler-Benz research and technology North America in

Portland, OR. In 2002, he received the Ph.D. degree on content-based image retrieval at the University of Geneva with a research stay at Monash University in Melbourne, Australia, in 2001. Since 2002 he has been working at the medical faculty of the University of Geneva and since 2007 he has been professor at the HES-SO. Henning is currently coordinator of the EU-project Khresmoi and scientific coordinator of the VISCERAL project. He has published over 300 research articles and organized the benchmarking event ImageCLEF for ten years. content analysis with crowdsourcing.

Pushing the Envelope: Solving Hard Multimedia Problems with Crowdsourcing

Wei-Tsang Ooi
National University of Singapore
ooiwt@comp.nus.edu.sg

Oge Marques
Florida Atlantic University, USA
omarques@fau.edu

Vincent Charvillat Axel Carlier
University of Toulouse, France
vincent.charvillat@enseeiht.fr
axel.carlier@enseeiht.fr

1. Introduction

The solution to many contemporary problems in multimedia research involves discovering ways to bridge – or at least narrow – the *semantic gap* (the difference between the data that can be captured from raw pixels or sound samples and the high-level interpretation assigned by humans to the associated images, videos, or music clips). The difficulty in bridging the gap has led the multimedia research community to break the problem down into smaller sub-problems, such as image segmentation, image tagging, speech-to-text conversion, and natural language processing. Alas, in almost every field of multimedia research, the performance achieved by state-of-the-art algorithms is far inferior to humans performing comparable tasks. Moreover, most of the research focuses on specific domain or applications. Finding a general solution remains an open problem.

More recently, *crowdsourcing*, in which inputs from large numbers of human participants are pooled to serve as a basis for statistical analysis and inference, has arisen as a promising approach to address the sub-problems, in an effort to bridge the semantic gap. In this paper, we argue that – despite the success achieved by several of these early works – crowdsourcing is a much more powerful weapon and we should use it to directly solve the significantly harder *primary* problems, instead of its sub-problems. The rest of this article provides a few examples to support this argument, including two from our own research work.

2. Example 1: Identifying Interesting Regions in a Video

Our first example concerns the following primary problem: *Given a video clip, which region in the video would a user be interested to view at a given time?* This problem is extremely challenging if we attempt to tackle it using content analysis techniques. The notion of "interesting" region is not even computationally well defined.

To answer the question "what the user would like to look at?" one would have to at least model the content of the video, which is extremely hard. For the sake of exposition, let's consider the problem in the specific domain of lecture videos — typically consisting of a person lecturing in front of a lecture hall, with a blackboard or projected slide in the background. Understanding the content of such video can be broken down into several sub-problems: (i) extracting the text from the slides or the blackboard, (ii) separating the voice of the lecturer from the background noise, (iii) transcript the speech into text, (iv) perform natural language understanding to understand the context of the lecture, (v) separating the lecturer from the background in the video, (vi) identifying the gesture of the lecturer. Once the context of the lecture and the relationship between the speech and the text are known, one can then make a guess about the region of video that is of interest to the user. For instance, when the lecture says "According to the Fermat's Theorem" (from (ii), (iii) and (iv)) and pointing to the direction on the board (from (v) and (vi)), one can guess that at that moment in the video, the scribbled Fermat's Theorem on the board (from (i)) would be the region of interest to the users.

The computer vision and multimedia research community, to a varied degree of success, has extensively studied each of the sub-problems above. Some problems, such as transcribing a professor's handwriting on the board to text, remain a challenge. Even if each of the sub-problems above is satisfactorily solved, the solution to infer the interesting regions is restricted to a specific domain.

One could be tempted to tackle these sub-problems via crowdsourcing. While we acknowledge that solving some of these sub-problems would be useful in some applications (e.g., speech to text is useful for indexing and retrieval), we argue that we can skip the sub-problems, and use crowdsourcing to directly address the original question, "which region would the user like to view in the video."

We presented an approach to identify the interesting regions that user would like to view in a video through crowdsourcing [1][2]. Our idea is to utilize a novel interface for watching video, that supports zoom and pan operations. The interface allows user to explicitly zoom into regions they are interested in to view in more details. The zoom action provides explicit feedback to the system that a particular region is of

interest to a user. By aggregating the feedback from multiple users, we are able to determine a set of regions that user are likely to be interested in naturally.

3. Example 2: Describing an Image

The second example concerns the following primary problem: *describe a given image in text.* This problem is extremely challenging to solve, since it requires understanding the content and context of the image in order to generate the appropriate text.

The term "description" could be broad, and similar to the first example, is not computationally well defined. Here, we narrow down the problem into describing the objects in the image, their actions, and relationships to each other. One can break the problem down into several sub-problems: (i) segment the image into objects, (ii) identify *what* is each object in the image, (iii) identify the *role* of each object, and (iv) identify the *relationships* among the objects. It is reasonable to assume that additional meta-data from the camera – containing information about *when* and *where* the photo was taken – is available. Even with this assumption, being able to automatically generate a description such as "Obama hugs Michelle" (from the most-liked image of all time from Facebook at the time of writing) is extremely hard.

Researchers have had great success with image segmentation techniques (Step (i)) and are actively pursuing the sub-problem of annotating images (Step (ii)-(iv)), including many recent efforts that are based on crowdsourcing.

Again, we argue that one could attempt to solve the primary problem with crowdsourcing, instead of solving the sub-problems individually. Unlike our first example, where pooling information about what existing users look at naturally tells us what other users should look at, however, this second example is non-trivial to solve with crowdsourcing. A naïve way would be to ask the crowd to describe an image in natural language, or ask them to explicitly label the objects, actions, and relations in the image. Each of this approach has its own drawback: the former would require natural language understanding, which itself is a hard problem, while the latter requires incentives for users to perform the labor-intensive tasks of labeling.

We recently presented an approach towards answering the problem, focusing on identifying the relevant objects in an image and their spatial relationship through a game with a purpose (GWAP). The game, Ask'nSeek, is a two-player Web-based guessing game that asks users to guess the location of a hidden region within an image with the help of semantic and topological clues [3]. The information collected from game logs is combined with results from content analysis algorithms and used to feed a machine learning algorithm that outputs the outline of the most relevant regions within the image and their names (Figure 1). The approach solves two computer vision problems – object detection and labeling – in a single game and, as a bonus, allows learning of spatial relations (e.g., sky is above the man) within the image.

Figure 1. Examples of object detection and labeling results obtained with the game-based approach described in [3]: (left) four objects /regions were detected and their bounding boxes were labeled as 'woman', 'sky', 'motorbikes', and 'man'; (right) two objects ('cat' and 'dog') were detected and labeled.

4. Other Examples

Besides the two examples taken from our own research work, there are many other examples of primary multimedia problems that require bridging of the semantic gaps and are well known to be hard. We sample two such problems in this section that remains open but we believe could benefit from crowdsourcing.

Lyrics Transcription. Lyrics transcription involves transcribing the lyrics of a given songs to text, and is well recognized as a hard problem. Traditional approaches typically address three sub-problems: (i) separating the vocal from the instrument, (ii) segmenting the vocal into words, and (iii) transcribing each word.

Video-Song Matching. The problem of automatic soundtrack generation involves finding the most appropriate song from a collection to use as a soundtrack of a video clip (e.g., a home video), such that the song's content fits the scene of the video. One could view this problem as a combination of several sub-problems: (i) find the interesting regions of the video, (ii) transcript the scene, (iii) find the song with lyrics that best fit the description of the scene. Even if we assume the lyrics of the songs are available, the problem is still extremely difficult.

There is no known work that uses crowdsourcing to address the two problems above. We, however, believe that the technique is powerful enough to solve such

hard problems, with a well thought out system or GWAP. A system similar to reCAPTCHA could perhaps be useful for lyrics transcription. A song-guessing game may help with matching between video clips and songs.

5. Discussion

In the above, we argued that crowdsourcing is a powerful tool and we should push its envelope and use it to directly address hard multimedia problems that are challenging to solve using traditional content analysis approach, instead of using crowdsourcing on the sub-problems. Doing crowdsourcing correctly and effectively, however, is non-trivial.

Our experience with crowdsourcing leads to the following insights to designing useful crowdsourcing systems. The tasks or games should be designed such that *the input collected from the users is as simple as possible, but carries as much meaningful information as possible*. One key aspect in crowdsourcing is to cope with outlying data, and it is easier to filter out diverging contributions if the input is simple (for example, a click on a video to zoom in, a click on an image to play the game of Ask'nSeek). These inputs, however, should contain meaningful information, i.e., information hard to gather computationally, but easy to get by a human, to establishing relations between (different) media, semantics, and user interest.

We also would like to highlight that, despite the power that crowdsourcing yields, we should not ignore the traditional approach of content analysis. We argue that *content analysis should be used to augment crowdsourcing,* to reduce the number of participants needed to obtain meaningful results. To obtain meaningful relations from a limited number of user inputs in our work [2][3], we tried to plug the results from a limited number of contributions with content analysis outputs. In order for this combination to make sense, it is interesting to visualize the contributions from users as constraints that can help improve content analysis algorithms. For example, the relations obtained from Ask'nSeek can help filtering false positives from the classical approaches of object detection, as well as usage-based attention maps can complement saliency maps from the content analysis. In that sense we think that semi-supervising the content analysis by usage analysis could be a promising way of research.

6. Conclusion

In this paper we have postulated that it is time for the multimedia research community to take early successful attempts to apply crowdsourcing (including micro-tasks and games) to a new level, moving from bite-sized building blocks (e.g., image tagging or object detection) to grander, more encompassing primary problems, e.g., video-song matching, scene understanding and context-aware object recognition. We shared how crowdsourcing can be used to address two examples of such problems, drawing from our own research work. We also highlighted insights from our work on how to make crowdsourcing effective when addressing such hard problems.

References

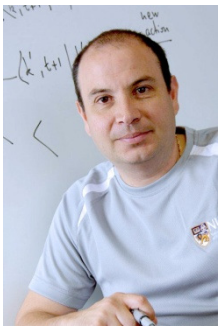
[1] Axel Carlier, Vincent Charvillat, Wei Tsang Ooi, Romulus Grigoras, and Geraldine Morin. 2010. Crowdsourced automatic zoom and scroll for video retargeting. In *Proceedings of the international conference on Multimedia* (MM '10). ACM, New York, NY, USA, 201-210

[2] Axel Carlier, Guntur Ravindra, Vincent Charvillat, and Wei Tsang Ooi. 2011. Combining content-based analysis and crowdsourcing to improve user interaction with zoomable video. In *Proceedings of the 19th ACM international conference on Multimedia* (MM '11). ACM, New York, NY, USA, 43-52.

[3] Axel Carlier, Oge Marques, Vincent Charvillat: Ask'nSeek: A New Game for Object Detection and Labeling. ECCV Workshop on Web-Scale Vision and Social Media, Florence, Italy, 249-258.

Wei Tsang Ooi received his B. Sc. (Hon.) degree from National University of Singapore in 1996, and Ph. D. in Computer Science from Cornell University in 2001. He spent a year as postdoc at Berkeley Multimedia Research Center in U.C. Berkeley, before re-joining NUS in 2002, where he is currently an Associate Professor in the Department of Computer Science. Wei Tsang's research focuses on interactive multimedia systems, including zoomable videos and networked graphics.

Oge Marques is Associate Professor of Engineering and Computer Science at Florida Atlantic University (FAU). He received his Ph.D. in Computer Engineering from FAU in 2001. His current research interests include the use of serious games and crowdsourcing to advance human and computer vision. He is a senior member of both the ACM and the IEEE.



Vincent CHARVILLAT received the Eng. degree in Computer Science and Applied Mathematics from ENSEEIHT, Toulouse France and the M.Sc. in Computer Science from the National Polytechnic Institute of Toulouse, both in 1994. He received the Ph.D. degree in Computer Science from the National Polytechnic Institute of Toulouse in 1997. He joined the Computer Science and Applied Mathematics department of ENSEEIHT in 1998 as an assistant professor. He obtained the habilitation degree in Computer Science in 2008 and is currently a full professor at the University of Toulouse, IRIT research lab, ENSEEIHT Eng. School. Vincent CHARVILLAT is the head of VORTEX research team at ENSEEIHT (Visual Objects: from Reality To EXpression). His main research interests are visual processing and multimedia applications.

Axel Carlier received his Master's Degree from INPT in 2011 and is currently a Ph. D. student, working under the supervision of Vincent Charvillat on the combination of content analysis with crowdsourcing.

CALL FOR PAPERS

IEEE EUROCON 2013 - International Conference on Computer as a Tool

IEEE EUROCON Special Session: From QoS to QoE and Back

Purpose of this special session:

In recent years, there has been a marked increase in the use of multimedia applications over the Internet. Nowadays video streaming accounts for a large fraction of all bandwidth used, and Internet Telephony and IP-based teleconferencing systems have become part of our daily lives.

The way users perceive the quality of these services is critical both for the users themselves and for the content and service providers. In turn, the quality of these services depends strongly on the underlying transport network and its performance. It is therefore very important to understand how the network-level QoS (Quality of Service) is related to the services' Perceptual Quality and QoE (Quality of Experience), and conversely, how can the latter be exploited in order to improve the performance of the networks and services that run on them.

This Special Session will focus on the relationship between QoS, Perceptual Quality and QoE in general. We welcome submissions on the following topics:

- QoS to QoE mappings
- QoE-driven Network Management
- Cross-layer optimization
- Network QoS and QoE modeling
- Integration of QoS and QoE models
- Standardization issues

Author instructions:

All submissions to this Special Session should follow the regular submission procedure and will be subjected to peer review coordinated by the Special Session Chairs and TPC Chairs. The paper format is the same as for the general conference. The papers must be submitted through the ConfTool, under the respective Special Session track. See paper format for more details.

Important dates:

- Manuscript submission: **February 1st, 2013**
- Notification of acceptance: **March 15th, 2013**
- Camera ready papers: **April 15th, 2013**

Organizers:

- Martin Varela (VTT Technical Research Centre of Finland, Finland) (Martin.Varela@vtt.fi)
- Periklis Chatzimisios (Alexander TEI of Thessaloniki, Greece) (pchatzimisios@ieee.org)

CALL FOR PAPERS
COMMUNICATION SOFTWARE, SERVICES, AND MULTIMEDIAAPPLICATION SYMPOSIUM

Symposium Co-Chairs

Vincent Wong, University of British Columbia, Canada
vincentw@ece.ubc.ca

Liang Zhou, Nanjing University of Posts and Telecommunications, China
liang.zhou@ieee.org

Scope and Topics of Interest

The Communications Software, Services and Multimedia Application Symposium will provide an international technical forum for discussing and presenting recent research results on any aspects of software, services, and multimedia communications. It aims at bringing together experts from industry and academia to exchange ideas and present results on advancing the state-of-the-art and overcoming research on the challenging issues related to the software design, system deployment of services, and multimedia applications over heterogeneous networks. Papers may present theories, techniques, applications, or practical experiences related to that. Topics of interest for this Symposium include, but are not limited to:

Multimedia Applications and Services

- Multimedia delivery and streaming over wired and wireless networks
- Cross-layer optimization for multimedia service support
- Multicast, broadcast and IPTV
- Multimedia computing systems and human-machine interaction
- Interactive media and immersive environments
- Multimedia content analysis and search
- Multimedia databases and digital libraries
- Converged application/communication servers and services
- Multimedia security and privacy
- Multimedia analysis and social media

Network and Service Management and Provisioning

- Multimedia QoS provisioning
- Multimedia streaming over mobile social networks and service overlay networks
- Service creation, delivery, management
- Virtual home environment and network management
- Charging, pricing, business models
- Security and privacy in network and service management
- Cooperative networking for streaming media content

Next Generation Services and Service Platforms

- Location-based services
- Social networking communication services
- Mobile services and service platforms
- Home network service platforms
- VoP2P and P2P-SIP services

Software and Protocol Technologies for Advanced Service Support

- Ubiquitous computing services and applications
- Networked autonomous systems
- Communications software in vehicular communications
- Web services and distributed software technology
- Software for distributed systems and applications, including smart grid and cloud computing services
- Peer-to-Peer technologies for communication services
- Context awareness and personalization

Submission Guidelines

Prospective authors are invited to submit original technical papers by the deadline of **15 March 2013** for publication in the IEEE Globecom 2013 Conference Proceedings and for presentation at the conference. Submissions will be accepted through EDAS. All submissions must be written in English and be at most six (6) printed pages in length, including figures. For full details, please visit the following website:

http://www.ieee-globecom.org/2013/submguide.html

MMTC OFFICERS

CHAIR

Jianwei Huang
The Chinese University of Hong Kong
China

STEERING COMMITTEE CHAIR

Pascal Frossard
EPFL, Switzerland

VICE CHAIRS

Kai Yang
Bell Labs, Alcatel-Lucent
USA

Chonggang Wang
InterDigital Communications
USA

Yonggang Wen
Nanyang Technological University
Singapore

Luigi Atzori
University of Cagliari
Italy

SECRETARY

Liang Zhou
Nanjing University of Posts and Telecommunications
China

E-LETTER BOARD MEMBERS