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

R-LETTER

Vol. 5, No. 5, October 2014

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from the Review Board Directors **2**
Towards Cloud-based Mobile Multimedia Applications **3**
A review for "Adaptive Mobile Cloud Computing to Enable Rich Mobile Multimedia Applications" (Edited by Bruno Macchiavello) 3
A Framework for the Quality of Service Assessment of Cloud Gaming Systems **5**
A review for "On the Quality of Service of Cloud Gaming Systems" (Edited by Roger Zimmermann) 5
Challenges and a Green Solution for Cloud Gaming **8**
A review for "Cloud Gaming: A Green Solution to Massive Multiplayer Online Games" (Edited by Yan Zhang) 8
Towards Context-Aware Multimedia Content Delivery in Cooperative Satellite and Terrestrial Networks **11**
A review for "Prospects and Challenges of Context-Aware Multimedia Content Delivery in Cooperative Satellite and Terrestrial Networks" (Edited by Koichi Adachi) 11
Using Depth Data to obtain a High-level Description of the Hand **13**
A review for "Parsing the Hand in Depth Images" (Edited by Carl James Debono) ... 13
Exploration of Naturalness in Image Categorization **15**
A review for "From Large Scale Image Categorization to Entry-Level Categories" (Edited by Jun Zhou) 15
Paper Nomination Policy **17**
MMTC R-Letter Editorial Board **18**
Multimedia Communications Technical Committee Officers **18**

Message from the Review Board Directors

Welcome to the October issue of the IEEE Communications Society Multimedia Communications Technical Committee (MMTC) R-Letter. This issue is brought to you by review board members who independently nominated research papers published within IEEE MMTC sponsored publications and conferences. Interestingly, for this issue we received a few nominations related to **cloud computing**, specifically within **mobile systems**, **gaming**, and **green solutions**, which are reviewed in the first part of this R-Letter. The second part covers reviews related to **multimedia delivery in satellite and terrestrial networks**, **depth data analysis**, and **image categorization**.

We hope you that this issue stimulates your research in the area of multimedia communication and an overview of all reviews are provided in the following:

The **first paper**, published in the *IEEE Transactions on Multimedia* and *edited by Bruno Macchiavello*, describes means for adaptive cloud computing to enable rich mobile multimedia computing and, thus, provides a first step towards cloud-based mobile multimedia applications.

The **second paper**, also published in the *IEEE Transactions on Multimedia* and *edited by Roger Zimmermann*, presents a framework for the Quality of Service (QoS) assessment of cloud gaming systems.

The **third paper** is *edited by Yan Zhang* and has been published within the *IEEE Wireless Communication Magazine*. It highlights challenges for cloud computing and suggests a green solution for cloud gaming.

The **forth paper**, published in the *IEEE Communications Magazin* and *edited by Koichi Adachi*, describes means towards context-aware multimedia content delivery in cooperative satellite and terrestrial networks.

The **fifth paper**, published in *IEEE Transactions on Multimedia* and *edited by Carl James Debono*, describes a method using depth data to obtain a high-level description of the hand.

Finally, the **sixth paper**, *edited by Jun Zhou* and published within the *Proceedings of the IEEE International Conference on Computer Vision*, explores the naturalness in image categorization.

We would like to thank all the review board members for their time and efforts. In particular, we would like to thank Gene Cheung and Guillaume Lavoue who resigned and we would like to welcome Bruno Macchiavello as a new member of the review board.

IEEE ComSoc MMTC R-Letter
Director: Christian Timmerer
Alpen-Adria-Universität Klagenfurt, Austria
Email: christian.timmerer@itec.aau.at
Co-Director: Weiyi Zhang
AT&T Research, USA
Email: wzhang@ieee.org
Co-Director: Yan Zhang, Simula, Norway
Email: yanzhang@simula.no

Towards Cloud-based Mobile Multimedia Applications

A review for "Adaptive Mobile Cloud Computing to Enable Rich Mobile Multimedia Applications"
(Edited by Bruno Macchiavello)

S. Wang, S. Dey, "Adaptive Mobile Cloud Computing to Enable Rich Mobile Multimedia Applications," IEEE Transactions on Multimedia, vol.15, no.4, pp.870-883, June 2013.

Nowadays, mobile devices, like smartphones, are increasingly becoming essential tools for human communication, mostly due to the fact that they provide computing capability irrespective of user movement. This has led to a consolidated market for development of different applications (e.g. computer games, location and map information, etc.) that aim mobile devices. With this continuous increase of mobile applications, new challenges are emerging regarding security, bandwidth, storage, battery life, etc.

The use of mobile cloud computing seems a promising solution for several of these new challenges. Mobile cloud computing refers to an infrastructure where both the data storage and the data processing can happen outside the mobile device. Studies are predicting a significant increase of revenue from mobile enterprise cloud-base applications and services in the upcoming years [1,2]. Cloud storage, not only solves the problem of insufficient local storage capability, but gives the ability to access the data from any device, platform and network. However, the possibility of not only storing but also processing the data outside the mobile device can enable a much richer multimedia experience compare to native applications.

In this paper the authors present and discuss different possible types of Cloud Media Mobile (CMM) applications and analyze some challenges that need to be addressed in order to make those application viable, including response time, user experience, computing cost, bandwidth and scalability for a large number of users. Moreover, they propose an approach for cloud-based image rendering, used during Cloud Mobile Gaming. As observed by the authors, despite the current progress in the capabilities of mobile devices, they do not meet the requirements of the latest 3D and multi-view rendering techniques. Cloud based rendering can potentially enable mobile user to experience the same rich Internet games available to high-end PC users [3], or participate in rich augmented reality and multi-view immersive applications.

In order to enable cloud based rendering, the authors measure the required communication (bandwidth) and computation complexity during rendering operations. In the proposed rendering adaptation method, these requirements will dynamically vary, such that the resulting bit rate and computational load can meet the available network bandwidth and cloud server computing resources, respectively. Therefore, the rendering parameters that can affect communication and computational complexities must be identified. The authors grouped those rendering parameters into four sets: *(i)* realistic effects, which includes color depth, anti-aliasing, texture filtering, etc., *(ii)* texture detail, which refers to how many textures are used to present objects, *(iii)* view distance and *(iv)* environmental details, such as grass, flowers, etc. By varying these four group of settings, different complexity models can be created offline for a specific game or video application. These models are then used to compute rendering complexity levels, such that the root mean square error between the target complexities (communication and computational) and the measured complexities is minimized. Once the levels are available, an online joint rendering and encoding adaptation algorithm (JREA) determines the current settings for each user in near real-time.

The proposed JREA follows three steps. In the first step an encoding bit rate is defined. If during a short time interval the round trip delay keeps increasing, the JREA will reduce the encoding bit rate. On the other hand, if the round trip delay remains bellow a minimum acceptable value the bit rate will be increase. In the second step, the communication complexity level is adapted in order to meet the desired encoding bit rate. In the final and third step, the computational complexity is modified based on the measured server utilization level.

In the simulations not only the response time, but also subjective score were used in order to verify if the proposed algorithm is able to deliver acceptable gaming experience. Tests were performed using a 3G network and different environments (open and close locations). The results showed that the use of the proposed rendering adaptation technique can reduce the network delay and packet loss rate, while maintaining an acceptable video quality.

One minor drawback of the proposed technique is the necessity of creating specific complexity models for each application. Nevertheless, these models can be provided during application development. Future work can focus on the latency and jitter problem associated with the transmission of large amount of data between servers and mobile devices in a network. Another important challenge to be approach is ensuring scalability for a large number of simultaneous users within the limited capacity of mobile networks. Finally, CMM applications involve both mobile service provider and cloud service provider with different services management, methods of payment and prices. In order to develop adequate applications, it is important to consider a business model that includes pricing and revenue sharing between different providers.

References:

[1] Juniper Research, Mobile Cloud: Smart Device Strategies for Enterprise & Consumer Markets 2011-2016, Jul. 2011. [Online]. Available: http://juniperresearch.com/.

[2] MarketsAndMarkets, World Mobile Applications Market—Advanced Technologies, Global Forecast (2010-2015), Aug. 2010. [Online]. Available: http://www.marketsandmarkets.com/.

[3] S.Wang and S. Dey, "Cloud mobile gaming: Modeling and measuring user experience in mobile wireless networks," in Proc. ACM SIGMOBILE MC2R, 2012, vol. 16, no. 1, pp. 10–21

Bruno Macchiavello is an assistant professor at the Department of Computer Science of the University of Brasilia (UnB), Brazil. He received his B. Eng. degree in the Pontifical Catholic University of Peru in 2001, and the M. Sc. and D.Sc. degrees in electrical engineering from the University of Brasilia in 2004 and 2009, respectively. Prior to his current position he helps developing a database system for the Ministry of Transport and Communications in Peru. He also was co-organizer of a special session on Streaming of 3D content in the 19th International Packet Video Workshop (PV2012). His main research interests include video and image coding, image segmentation, distributed video and source coding, multi-view and 3D video processing.

A Framework for the Quality of Service Assessment of Cloud Gaming Systems

A review for "On the Quality of Service of Cloud Gaming Systems"
(Edited by Roger Zimmermann)

Kuan-Ta Chen, Yu-Chun Chang, Hwai-Jung Hsu, De- YuChen, Chun-Ying Huang, and Cheng-Hsin Hsu, "On the Quality of Service of Cloud Gaming Systems", IEEE Transactions on Multimedia, Vol. 16, No. 2, February 2014.

In this study, the authors start from the observation that cloud gaming has attracted significant interest from both researchers and commercial companies. Cloud gaming, as the term is understood nowadays, specifically denotes a gaming system architecture that differs from regular online gaming. In both cases a client device (for example a PC, a set-top box or a mobile device) is connected to a server. However, in regular online gaming, *game state updates* are exchanged between the client and the server and the client contains a graphics rendering engine that can decode and process those game state updates. In contrast, with cloud gaming the real-time rendering is performed on the server- side and what is streamed to the client is a sequence of game screens. In other words, the server renders the game into a local frame buffer, which is then compressed and transmitted at a specific frame rate similar to a video stream transmission.

The authors rightly point out that cloud gaming has a number of advantages in that a client needs relatively modest hardware capabilities while at the same time it can show games that could traditionally only be rendered on high-end gaming hardware. Therefore only a so-called *thin client* is needed.

While a cloud gaming has some compelling advantages, it is important for the end user that the game play experience is good and in some respects comparable to a local setup. The authors in this paper therefore investigate and propose measurement techniques that can be used to assess the quality of service of cloud gaming systems. The authors draw on their experience that they gained while designing their own cloud gaming system called GamingAnywhere [2]. However, importantly, the presented methodology is designed to also work for proprietary games where neither the source code nor physical access to the server are available.

In the study the authors concentrate on obtaining detailed measurements relating to the following aspects and how they influence the quality of service:

- Traffic characteristics.
- Latency, specifically the response delay.
- Graphics quality.

Some of these metrics, such as the traffic characteristics can be relatively directly obtained by executing some monitoring tools on the network connection and by routing traffic through an instrumented router where the bandwidth, de- lay and packet loss can be controlled. Other metrics, such as the latency, require more detailed instrumentation, especially on commercial games where neither the source code nor the internal architecture are publicly available. For the latency measurements the authors propose to decompose it into its individual components such as the *network delay (ND)*, the *processing delay (PD)*, the *game delay (GD)*, and the *playout delay (PD)*. The overall response delay (RD) can then be expressed as:

$$RD = ND + PD + GD + OD$$

While the sub-components of the overall end-to-end latency RD are very interesting to know for the QoS assessment of a cloud gaming setup, they are not all directly observable. More details on how the authors achieve this is presented below. The main contributions of the proposed work are in two areas:

a) The authors propose measurement techniques that are applicable to commercial and proprietary cloud gaming systems. They describe how to instrument closed systems such that some detailed statistics can be obtained.
b) Two existing commercial systems are analyzed with the proposed methods, namely OnLive and StreamMyGame (SMG), and the results are compared in several dimensions.

Approach. The authors quantify the service quality of cloud gaming systems by instrumenting the transmission and client components and then monitoring actual game play. Because different game types have somewhat different requirements to achieve enjoyable game play, as has been shown in other studies [1], the authors perform their tests with three genres of games, namely *action-adventure* (ACT), *first-person shooter* (FPS), and *real-time strategy* (RTS). Within each category the authors consider three individual games.

Not all the measurements are performed with all the games, but a representative mix is used for the experiments. To understand how network characteristics affect the service quality the author route all traffic through a FreeBSD 7.0 router running the dummynet such that they can degrade the traffic bandwidth in a controlled manner. In their measurements the authors are interested in a number of specific metrics as described below.

Response Delay Measurement. The round-trip response delay is one of the most critical aspects of game play. It basically measures how quickly the result of a user input can be observed on the screen (say the user moves an avatar). Users are used to short delays from local game play such as game consoles. Because in cloud gaming the input-response loop includes a client, a network and a cloud server, the response time will naturally be longer. The objective then for any good cloud gaming system is to keep the additional latency introduced by the client-server architecture as short as possible. In order to achieve precise time measurements – as it is not easy visually capture the time instance of a screen change – the authors instrument the Windows client systems by "hooking" into the frame drawing functions. This is quite interesting and allows for accurate end-to-end time measurements.

While the end-to-end response delay (RD) can be measured, it is comprised of several parts, some of which are internal to the game engine and client and cannot be directly observed. Therefore, it is non-trivial to break it down into its constituent components such as the network delay, the processing and game delay and the playout delay. This is one of the most innovative parts of the proposed work. In their framework the authors need estimate the time it takes for a updated frame to be fully transmitted from the game server to the client. Since a frame consists of many (and a variable number) of packets, it is not easy to know when the last packet of a frame has arrived at a client. The authors run a testing procedure where they estimate the playout delay (OD) and at random times block the client from receiving packets. They then observe whether an updated frame is being shown by the client or not. By performing this measurement with random time intervals numerous times the authors are able to obtain scatter plots where the time t_3, as it is called in the paper, can be quite accurately estimated between the point clouds where frames have been received or not received.

The authors consider the responsiveness as one of main criteria to influence the cloud gaming quality of service. To better understand the factors that have an impact on the responsiveness, the authors vary scene complexity, update region size, screen resolution, and the computational power of the server. Results are presented for a variety of parameter settings and some of the interesting observations are that the commercial OnLive system is highly optimized and hence achieves a short response time which is influenced by, for example, the scene complexity. StreamMyGame generally has a longer response time and hence other quality factors play less of a role.

Quantifying Streaming Quality. The second major part of the study is designed to understand the streaming quality (i.e., the frame rate and graphics quality, measured by SSIM [3]) under various network conditions. Here the dummynet router is utilized to vary the network delay, the packet loss rate and the bandwidth available. The results presented for the two tested systems show that OnLive generally achieves a higher visual quality while requiring less network bandwidth. It is still playable with 1 or 2 Mbps bandwidth, while SMG requires higher bandwidth resources and the framerate starves when the bandwidth is too low. This may have much to do with the software frame encoding that SMG is using while OnLive's advantage is its proprietary video compression hardware at the server side.

Limitations. The framework proposed in the manuscript is overall very detailed and it should prove useful for researchers and practitioners working in the area of cloud gaming. The authors show how important quality metrics can be obtained accurately, even with commercial systems where the source code is not available. The study is specifically focusing on metrics that can be measured precisely, and the authors do not employ any user studies to assess how these factors may affect an actual user playing a game. However, generally with a user study only the overall effects can be understood, while the authors' work allows individual factors that influence the quality of service to be analyzed. Therefore the author's framework and user studies are two approaches that are complementary in their use.

References:

[1] M. Claypool, "Motion and Scene Complexity for Streaming Video Games," *In Proceedings of the 4th International Conference on Foundations of Digital Games, FDG '09*, pages 34–41, 2009.

[2] C.-Y. Huang, C.-H. Hsu, Y.-C. Chang, and K.-T. Chen, "GamingAnywhere: An Open Cloud Gaming System," *In Proceedings of the 4th ACM Multimedia Systems Conference, MMSys '13*, pages 36–47, 2013.

[3] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibil-

ity to structural similarity," *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

Roger Zimmermann is an associate professor with the Department of Computer Science at the School of Computing with the National University of Singapore (NUS) where he is also an investigator with the Interactive and Digital Media Institute (IDMI). His research interests are in both spatio-temporal and multimedia in- formation management, for example distributed and peer-to-peer systems, spatio-temporal multimedia, streaming media architectures, georeferenced video management, mobile location-based services and geographic information systems (GIS). He has co-authored a book, six patents and more than hundred-eighty conference publications, journal articles and book chapters in the areas of multi- media and databases. He has received the best paper award at IEEE ISM 2012 and was part of the team who won second place at the ACM SIGSPATIAL GIS Cup 2013. He has been involved in the organization of conferences in various positions, for example program co-chair of ACM Multimedia 2013. He co-directs the Centre of Social Media Innovations for Communities (COSMIC) at NUS and is an investigator with the NUS Research Institute (NUSRI) in Suzhou, China. Roger Zimmermann is an Associate Editor of the ACM Transactions on Multimedia journal (TOMM, formerly TOMCCAP) and the Multimedia Tools and Applications (MTAP) journal. He is a Senior Member of the IEEE and a member of ACM. For more details, see http://eiger.ddns.comp.nus.edu.sg.

Challenges and a Green Solution for Cloud Gaming

A review for "Cloud Gaming: A Green Solution to Massive Multiplayer Online Games"
(Edited by Yan Zhang)

S. P. Chuah, C. Yuen, N. M. Cheung, "Cloud Gaming: A Green Solution to Massive Multiplayer Online Games," Wireless Communication, IEEE, vol. 21, no. 4, pp. 78-87, 2014.

Video gaming is one of the fastest growing applications in mobile media industry. Fueled by the large-scale adoption of smart mobile devices, mobile games such as Angry Birds and Candy Crush have surprised many by hitting a huge number of players in a short period of time. Despite tremendous growth of capability in mobile devices, gaps remain between what mobile devices can support and requirements of great video games. High-end video games often require advanced computer graphics and enormous computational power to render realistic and interactive gaming scenes. Intensive computations for game rendering drain the batteries of mobile devices quickly. Battery capacity has been lagging behind the growth of computation power in mobile devices. As a result, mobile devices are crippled by the limited power supply in running computationally intensive applications. In addition, video games require substantial storage in installation, thanks to their high-fidelity textured graphics.

In the introduction, the authors outlined three major aspects where cloud gaming offers green efficiency. The first aspect is in the software management. Cloud gaming has the advantages of making new releases available to client immediately, and solving compatibility and update issues. As game vendors only need to maintain the game software at cloud servers, software management becomes more cost effective and requires fewer resources. The second aspect lies on the hardware maintenance. In cloud gaming, games are rendered at cloud servers with abundant and scalable resources. Importantly, cloud gaming centrally manages a pool of storage and computing resources rather than each client being constrained by their resources. Economies of scale and statistical multiplexing of client usage ensure overall efficiency in hardware utilization. Besides, power consumption of user devices is reduced as power-consuming computations are carried out in cloud servers, which are often power optimized. Finally, cloud gaming provides green solution in the deployment of video games. Cloud gaming eliminates the cost for sale and distribution of video games by delivering the games to massive numbers of players directly. Besides, pay-per-service saves clients from the large upfront costs of owning hardware and software in which they are likely to lose interest sometime later.

Cloud gaming integrates designs from various technical fields, such as cloud computing, computer graphics, multimedia compressions, and communication networks. Green efficiency has been one of the major research goals in these technical fields recently. The authors presented brief overviews to these green designs which are fundamental building blocks to the overall green solution of cloud gaming.

High-end video games often feature fantastic graphics quality. Rendering these high quality game images constitutes the heave duty computing tasks that require efficient rendering algorithms. In terms of game rendering algorithms, each stage of the rendering pipeline incurs some computation complexity (thus computing power). The energy models of the rendering pipeline have been characterized in [1], which demonstrated that processing of geometry, fragments, and game logic consume the most power in the pipeline. In general, rendering quality and complexity (thus energy consumption) trade-off various aspects of graphics rendering.

The rendered game images are often compressed as video to be streamed to the remote thin clients. Similar to the graphics-rendering module, video coding is another computationally intensive task at the cloud servers. In video compression, there is a trade-off in the optimization of rate and distortion. To compress an image sequence, redundancies are removed via searches of motion vectors and coding mode.

The compressed bitstreams of game videos are then delivered via communication networks. Game video traffics often require high bandwidth, low latency communication networks,

which are power consuming. Thus, green designs in communication and networking are necessary. Video distortion is a key parameter in optimizing the user-perceived video quality. For example, a distortion-aware transmission scheme to optimize the streaming video quality over heterogeneous wireless networks is proposed in [2]. On the other hand, goodput is also of vital significance to enhance both traffic performance and network utilization, hence a novel goodput-aware load distribution model for delivering real-time multimedia traffic in [3].

Next, the authors moved on to review some existing cloud gaming platforms available in commercial and open sources. A notable open source platform is the GamingAnywhere [4], which is an open platform allowed for extensibility, portability, and configurability. More importantly, the fact that Gaming-Anywhere supports major platforms (Windows, Linux, OS X, and Android) is the additional strength to the open platform that the research community would like to take note.
As existing cloud gaming systems are platform-dependent, the authors introduced the VNC-based platform, which is very much in resemblance to the remote desktop system.

The true novelty from the authors' work is however the layered coding scheme [5] which they introduced subsequently. The layered coding scheme leverages on the growing computing capability of mobile devices to render some base layer game images. Instead of delivering the video bitstream of full image quality, cloud server only encodes and sends the enhancement layer images. Upon receipt of the enhancement layer bitstream, the client decodes to obtain the enhancement images, which will be used to enhance the base layer images to obtain the full quality images. Central to the scheme is the optimization of base layer pipeline which are computationally inexpensive, and yet render large amount of information contents in the rendered images. Conversely, enhancement layer images represent the rendering pipeline which is expensive to render but yet contain little information content to the final images. The notable advantage of the layered coding scheme is the successful reduction of bandwidth required to support the cloud gaming.
Finally, the authors outlined some future directions and challenges in the green design of cloud gaming. Among the interesting directions include the cross-layer rendering-coding-delivery optimizations. A network-aware video encoder can adapt the video bitstream based on network statistics to ensure quality of experience in gaming. On the other hand, content-aware network delivery algorithms could prioritize the video packets such that important and urgent video packets are delivered on time and reliably. Joint coding and delivery optimization avoids unnecessary waste of computational and communications energy due to poor coordination of these two technical modules.

In addition, the authors also foresaw trends in device-aware rendering of games. In a cloud gaming service, clients' devices are heterogeneous in display, processing capability, and power supply. They range from wide screen 3D-enabled TVs to power-constrained smart phones. Frame rate and frame size can be adapted accordingly for different game types and device types. Another interesting future direction the authors pointed out is the bandwidth efficient social gaming. The game players and observers are often heterogeneous in their device terminals. Scalable video coding (SVC) offers a competitive solution for this scenario, which would otherwise require transcoding or multiple encodings for various resolutions. Last but not least, the authors also commented on the 3D vision gaming, which attracts much interest within the research and industry of video gaming recently.

References:

[1] J. Pool, A. Lastra, and M. Singh, "An Energy Model for Graphics Processing Units," Proc. 2010 IEEE Int'l Conf. Computer Design, 2010, pp. 409–16.

[2] J. Wu, B. Cheng, C. Yuen, Y. Shang, J. Chen, "Distortion-Aware Concurrent Multipath Transfer for Mobile Video Streaming in Heterogeneous Wireless Networks," to appear in IEEE Transactions on Mobile Computing. DOI: 10.1109/TMC.2014.2334592.

[3] J. Wu, C. Yuen, B. Cheng, Y. Shang, J. Chen, "Goodput-Aware Load Distribution for Real-time Traffic over Multipath Networks," to appear in IEEE Transactions on Parallel and Distributed Systems. DOI: 10.1109/TPDS.2014.2347031.

[4] C.-Y. Huang et al., "Gaminganywhere: An Open Cloud Gaming System," Proc. 4th ACM Multimedia Systems Conf., 2013, pp. 36–47; http://doi.acm.org/10.1145/ 2483977.2483981.

[5] S.-P. Chuah and N.-M. Cheung, "Bandwidth Efficient Mobile Cloud Gaming with Layered Coding and Scalable Phong Lighting," Proc. 21st IEEE Int'l. Conf. Image Processing, 2014, Oct. 2014.

Prof. Yan Zhang is currently Head of Department, Department of Networks at Simula Research Laboratory, Norway. He is also an Ad-

junct Associate Professor at the Department of Informatics, University of Oslo, Norway.

P He received a PhD degree in School of Electrical & Electronics Engineering, Nanyang Technological University, Singapore. He is an associate editor or on the editorial board of a number of journals. He also serves as the guest editor for IEEE Communications Magazine, IEEE Wireless Communications Magazine, IEEE Transactions on Industrial Informatics, IEEE Systems Journal, and IEEE Intelligent Systems. His current research interests include: communications solutions for reliable and secure cyber-physical systems (e.g., transport, smart grid), machine-to-machine communications, Internet-of-Things, economic approaches for networks performance optimization.

Towards Context-Aware Multimedia Content Delivery in Cooperative Satellite and Terrestrial Networks

A review for "Prospects and Challenges of Context-Aware Multimedia Content Delivery in Cooperative Satellite and Terrestrial Networks"
(Edited by Koichi Adachi)

Y. Kawamoto, Z. Md. Fadlullah, H. Nishiyama, N. Kato, and M. Toyoshima, "Prospects and Challenges of Context-Aware Multimedia Content Delivery in Cooperative Satellite and Terrestrial Networks", IEEE Commun. Mag., vol. 52, no. 6, pp.55-61, Jun. 2014.

A seamless access to rich multimedia content on users' devices has becomes more popular due to the rapid increase of mobile multimedia services in Heterogeneous network [1]-[4]. In this paper, cooperative satellite cells and terrestrial wireless communication networks in considered as an enabler for such a seamless multimedia access [5]. To provide seamless multimedia access to users, the authors focus on context information of users. Recently, context awareness has become popular to accompany these services not only to enhance the users' perceived service quality but also to improve the overall utilization of such cooperative networks.

Due to the different nature of two networks, i.e., satellite network and terrestrial network, delivering context-aware multimedia content through these cooperative networks is associated with a significant research challenge. In this paper, the authors address several challenges to provide seamless multimedia content, and propose a dynamic bandwidth allocation method to effectively utilize the satellite and terrestrial networks for providing context-aware contents to many users. Computer-based simulation results are presented to demonstrate the effectiveness of the proposed method. The authors also provide related research works for the satellite-terrestrial cooperative network [7]-[10].

Cooperation among different technologies is likely to form the future heterogeneous network, capable of delivering rich multimedia contents (TV broadcasting, video on demand streaming, and other services) to a large-scale audience. Awareness of context, i.e., users' interests, places and locations, device specification (e.g., display resolution, capacity), subscribed data plan, and so forth, has recently become a popular feature to accompany these services to enhance quality of service as well as overall network utilization. For example, in order to deliver context-aware TV programs to a user through a satellite, the delivery method is typically unicast, which is obviously an expensive operation because satellites are intrinsically useful for broadcasting or multicasting services to many users. On the other hand, usually terrestrial mobile networks are better capable of delivering unicast-based context-aware content delivery. Furthermore, traditional cooperative satellite and terrestrial networks share a fixed bandwidth allocation method. However, due to non-uniform geographical distribution of users, dissimilarity of required contents depending on specific regions and/or users, and so forth, the fixed bandwidth allocation and the conventional way of utilizing these networks (e.g., satellite networks are used during post-disaster recovery operations, while terrestrial networks are used in normal times) are inefficient. In order to address these challenges, in this article, the authors propose a novel method to dynamically assign bandwidth in the satellite/terrestrial frequency sharing system to deliver context-aware contents to many users.

As a first step, the authors provide the clear understanding of the challenges associated with the concept of *context-aware service delivery*. Context awareness implies that the user's information need can be anticipated and responded to in an almost automated fashion. Given a particular user, his/her information and service needs vary depending on his/her immediate and/or individual situation, which dictates his/her context. The context-aware service aims to provide the user with contextually relevant information such as his/her preferences in terms of gender, age, language, used network technology, location, date/time of access, interest, activity, mood, device specifications, and so forth.

Several examples about the association among the context information and the scenarios are provided to highlight their relation. They have an impact mainly on video quality and the choice of video quality is subject to a number of factors (e.g., screen resolution of the device, available bandwidth, and cost). Therefore, on behalf of the user automatically is a key research challenge, particularly in combined satellite and terrestrial networks.

Although the distribution of network users is geographically non-homogenous, in the existing satellite/terrestrial frequency sharing system, the same

bandwidth is allocated to each of the satellite cells. Additionally, the required contents are different from region to region, and also from user to user. Moreover, in the traditional model, the satellite and terrestrial networks are considered to be utilized separately, while these networks share the bandwidth. Thus, the existing bandwidth allocation and the way to utilize these networks are inefficient.

To overcome the aforementioned problems, a dynamic bandwidth allocation method is proposed to provide context-aware contents to many users by using cooperative satellite and terrestrial networks. The proposed method is composed of two steps: the bandwidth allocation and optimization steps in order to maximize the number of users who can enjoy the context-aware contents. In the proposed method, to make efficient use of the satellite's multicast and broadcast capability, the satellite allocates its bandwidth to each cell according to the amount of requirements for multicast or broadcast contents. Based on the allocated bandwidth, the method determines which satellite or terrestrial networks should provide the contents to maximize the number of users enjoying the context-aware contents.

The computer simulation is conducted to show the effectiveness of the proposed method to provide context-aware multimedia contents. It is shown that the number of users who can enjoy context-aware contents can be greatly increased.

This paper addressed the challenges and provided the solution for cooperative satellite and terrestrial network system by utilizing the users' context information. The proposed method could efficiently utilize the limited frequency bandwidth and increase the system capacity. This work can be used as a benchmark scheme for future evolution of cooperative satellite and terrestrial network system.

References:

[1] Y. Kawamoto et al., "Effective Data Collection via Satellite-Routed Sensor System (SRSS) to Realize Global-Scaled Internet of Things," IEEE Sensors J., vol.13, no.10, Oct. 2013, pp. 3645–54.

[2] T. Sakano et al., "Disaster-Resilient Networking: A New Vision Based on Movable and Deployable Resource Units," IEEE Network, vol. 27, no. 4, July–Aug. 2013, pp. 40–46.

[3] Z. Yang et al., "Topology-Aware Modulation and Error-Correction Coding for Cooperative Networks," IEEE JSAC, Special Issue on Cooperative Networking Challenges and Applications, vol. 30, no. 2, Feb. 2012, pp. 379–87.

[4] Y. Kawamoto et al., "A Traffic Distribution Technique to Minimize Packet Delivery Delay in Multi-Layered Satellite Networks," IEEE Trans. Vehic. Tech., vol. 62, no. 7, Sept. 2013, pp. 3315–24.

[5] N. Lee et al., "Providing Seamless Services with Satellite and Terrestrial Network in Mobile Two Way Satellite Environments," in "Managing Next Generation Networks and Services," LNCS, Springer, vol. 4773–2007, pp. 551–54.

[6] T. Aman, T. Yamazato, and M. Katayama, "Traffic Prediction Scheme for Resource Assignment of Satellite/Terrestrial Frequency Sharing Mobile Communication System," Int'l. Wksp. Satellite and Space Commun, Siena-Tuscany, Italy, Sept. 2009.

[7] J. Lachner et al., "Challenges Toward User-Centric Multimedia," 2nd Int'l. Wksp. Semantic Media Adaptation and Personalization, Dec. 2007, pp. 159–64.

[8] C. Baladron et al., "Framework for Intelligent Service Adaptation to User's Context in Next Generation Networks," IEEE Commun. Mag., vol. 50, no. 3, Mar. 2012, pp. 18–25.

[9] C. Park et al., "The Concept and Realization of Context- Based Content Delivery of NGSON," IEEE Commun. Mag., vol. 50, no. 1, Jan. 2012, pp. 74–81.

[10] G. K. Huang et al., "Context-Aware Scalable Multimedia Content Delivery Platform for Heterogeneous Mobile Devices," Proc. Int'l. Conf. Advances in Multimedia, Apr. 2011.

Koichi ADACHI received the B.E., M.E., and Ph.D degrees in engineering from Keio University, Japan, in 2005, 2007, and 2009 respectively. From 2007 to 2010, he was a Japan Society for the Promotion of Science (JSPS) research fellow. He was the visiting researcher at City University of Hong Kong in April 2009 and the visiting research fellow at University of Kent from June to Aug 2009. Currently he is with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include green communication and cooperative communications. Dr. Adachi served as General Co-chair of the Tenth IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS) and Track Co-chair of Transmission Technologies and Communication Theory of the 78th IEEE Vehicular Technology Conference in 2013. He was recognized as the Exemplary Reviewer from IEEE Communications Letters and IEEE Wireless Communications Letters in 2012.

Using Depth Data to obtain a High-level Description of the Hand

A review for "Parsing the Hand in Depth Images"
(Edited by Carl James Debono)

H. Liang, J. Yuan, and D. Thalmann, "Parsing the Hand in Depth Images," IEEE Transactions on Multimedia, vol. 16, no. 5, pp. 1241-1253, August 2014.

Computer vision systems can provide a framework for human-computer interaction. Hand pose tracking and gesture spotting and recognition form new ways for humans to interact with computer systems. These techniques provide natural ways in which humans communicate and therefore can be used in a large number of applications. Examples of applications that can benefit from such systems include: sign language recognition, game interaction, virtual reality, control of machines and devices, and others. In order to develop such solutions, discriminative features that describe the hand are required.

Computer vision-based systems that do not rely on special intrusive equipment such as data gloves [1] and optical marker based techniques [2] are highly desirable to improve the user's computer interaction experience. A lot of research work has been done in hand pose tracking and gesture recognition using natural video, such as [3], [4], and [5], however their results are still not as accurate as the intrusive techniques.

A reason for the lack of accuracy in computer vision-based hand pose tracking and gesture recognition systems is due to the limitations of discrimination features. In the color video or image, the hand is very homogeneous in texture and features such as edge and silhouette are sensitive to environmental lighting and background. These features make pattern recognition difficult to achieve. On the other hand, depth sensors can provide better features and are used in work such as in [6], [7], and [8]. Although providing improvements, these features still provide limited discrimination for compact representations of articulated objects. The parsed hand parts provide high-level features that can be applied for hand gesture recognition and pose tracking, by segmenting the hand into different parts.

The authors of the original paper present a unified framework that makes use of both temporal information and spatial constraints for hand parsing. They use a depth-context feature and improve on its performance by applying a distance-adaptive sampling method. The authors show that the scheme gives better results than the binary depth comparison features used in [9]. Following this, a Superpixel Markov Random Field (SMRF) framework improves the hand parsing performance by modeling the spatial constraints.

The input to the framework consists of a sequence of depth images where only one hand is visible and is the closest object to the capturing device. The output from the system is part label images and key point sets that correspond to various parts of the hand. The process is divided into four main components, which are: (1) Per-pixel classification where an initial part label is assigned to each pixel using a learned Random Decision Forest (RDF) classifier and a temporal position classifier. In this process, a distance-adaptive feature selection method is used together with the depth-context feature to describe each pixel that is then classified. (2) Superpixel partition is used to reduce computational complexity by combining pixels with similar characteristics in the neighborhood. (3) Markov Random Fields (MRF) inference reassigns the hand part labels using superpixel-level inference. (4) Key point extraction then infers the center of the labeled hand parts using their distributions in the 3D space.

The training samples are generated using a 3D hand model. A fully deformable model with 3D closed mesh was developed by the authors and consists of the skeleton and a skin surface mesh. In order to generate the depth images and labels to train the classifier, the authors use the pinhole camera model [10] that allows the projection of the hand model onto an 2D image plane.

The hand classifier assigns a label to each pixel in the depth image of the hand, where the whole hand can be classified in twelve parts. The hand parts are connected with each other and therefore follow a fixed structure, which can be modeled as a tree-structure hierarchy. The RDF classifier is used for the labeling. When temporal information is available, the distribution of the position of an already labeled component can be used from the previous frame to aid the classifier. This improves the overall classification performance. The distance-adaptive sampling is applied since pixels nearer to the currently processed pixel

provide a better description of the context than further ones.

The results from the per-pixel classification are noisy and need smoothening. The MRF is applied to refine the results from the classification but this is done at superpixel level to reduce the data and allow the system to operate in near real time. The superpixel partition is done in the space of posterior distributions and uses the Kullback-Leibler divergence as a measurement for the difference between a pixel and the center of the superpixel cluster. The results reported in the paper indicate that the solution produces good results both for synthesized and real depth camera captured depth images compared to other methods found in literature.

Hand gesture recognition and tracking solutions require high accuracy for such technology to be used in real applications. Misclassification errors can lead to frustrations and erroneous behavior of devices leading to a low quality of experience. Therefore, more work is still needed in this field where better feature extraction solutions and better classifiers are fundamental for the success of such technology. Stereovision and multi-view solutions can provide more information at the expense of a higher computational complexity. Parallel structures and cloud computing can be employed to reduce time complexity and allow more complex solutions to still function in near real time. Furthermore, the issue of having multiple hands in the scene and the fact that in a normal environment the hand might not be the closest object to the capturing device need to be addressed in future work.

References:

[1] CyberGlove [Online]. Available: http://www.cyberglovesystems.com

[2] A. Aristidou, and J. Lasenby, "Motion capture with constrained inverse kinematics for real-time hand tracking," in *Proceedings of the 4th International Symposium on Communications, Control and Signal Processing*, March 2010.

[3] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using Kinect sensor," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1110-1120, August 2013.

[4] Z. Li, and R. Jarvis, "Real time hand gesture recognition using a range camera," in *Proceedings of Ausralasian Conference on Robotics and Automation*, December 2009.

[5] C. Kerdvibulvech, and H. Saito, "Model-based hand tracking by chamfer distance and adaptive color learning using particle filter," *EURASIP Journal on Image and Video Processing*, vol. 2009, January 2009.

[6] A. Baak, M. Muller, G. Bharaj, H.P. Seidel, and C. Theobalt, "A data-driven approach for real-time full body pose reconstruction from a depth camera," in *Proceedings of International Conference on Computer Vision*, November 2011.

[7] R.B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proceedings of IEEE International Conference on Robotics and Automation*, May 2009.

[8] L. Bo, X. Ren, and D. Fox, "Depth kernel descriptors for object recognition," in *Proceedings of IEEE/RSJ International Conference on Robots and Systems*, September 2011.

[9] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, and M. Finocchio, "Real-time human pose recognition in parts from single depth images," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 2011.

[10] J. Heikkila, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 1997.

Carl James Debono (S'97, M'01, SM'07) received his B.Eng. (Hons.) degree in Electrical Engineering from the University of Malta, Malta, in 1997 and the Ph.D. degree in Electronics and Computer Engineering from the University of Pavia, Italy, in 2000.

Between 1997 and 2001 he was employed as a Research Engineer in the area of Integrated Circuit Design with the Department of Microelectronics at the University of Malta. In 2000 he was also engaged as a Research Associate with Texas A&M University, Texas, USA. In 2001 he was appointed Lecturer with the Department of Communications and Computer Engineering at the University of Malta and is now an Associate Professor. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta.

Prof. Debono is a senior member of the IEEE and served as chair of the IEEE Malta Section between 2007 and 2010. He is the IEEE Region 8 Vice-Chair of Technical Activities for 2014. He has served on various technical program committees of international conferences and as a reviewer in journals and conferences. His research interests are in wireless systems design and applications, multi-view video coding, resilient multimedia transmission, and modeling of communication systems.

Exploration of Naturalness in Image Categorization

A review for "From Large Scale Image Categorization to Entry-Level Categories"
(Edited by Jun Zhou)

Vicente Ordonez, Jia Deng, Yejin Choi, Alexander C. Berg, and Tamara L. Berg. "From Large Scale Image Categorization to Entry-Level Categories", Proceedings of the IEEE International Conference on Computer Vision, 2013.

Image categorization has been intensively studied during the past decades. The research in this area has been recently boosted by two engines. First, large amount of image data becomes available in the Internet. This has facilitated the construction of large image datasets [1,2] which can be used to train classifiers that are capable of recognizing thousands or more object categories. Second, the widely available parallel computing facilities has made the fast processing of large amount of image data become possible.

The availability of large amount of image data also brings great challenges. One of them is how to organize image categories. The ImageNet dataset, which contains more than 14 million images, has used the WordNet [3] hierarchy to arrange the structure of object categories. Others, for examples, those collected from Internet and annotated by Web users, may use more natural language type descriptions [4,5]. An example given in this paper is the "Chlorophyllum Molybdites" category in ImageNet, which is more likely to be named as "Mushroom" in natural language. There are then the questions on how to utilize the structure of image categories in the WordNet, and how to combine these different types of labels.

In a previous work [6], Jia et al proposed to compromise the accuracy of classification and specificity in label prediction. On the one hand, the accuracy can be increased by generating a hypernym of the specific object class, for example label dog images as "animal" instead of "dog" or "cat". On the other hand, such vague labeling increases the uncertainty in prediction which can be measured by information gain from coarse to fine predictions. Based on these considerations, an optimization method is developed to maximize the information gain when accuracy is guaranteed. This has enabled effective utilization of the structure of labels.

Based on [6], Ordonez et al proposed a method to combine labels in the ImageNet and those generated in natural language. Entry-level category concept, which was proposed by psychologist, has been introduced to represent the names of categories tend to be used by people. Then the label combination can be implemented by mapping concepts predicted by image categorization systems to the entry-level categories.

A key idea in this paper is to measure the naturalness of labels, which allows translation of concepts only based on language models. The naturalness can be estimated by calculating the frequency that a concept appears on the web. The more it appears, the more likely it is an entry-level concept. Then the goal is to maximize the difference between the naturalness and distance between two labels in the hypernym structure so the optimal concept mapping can be achieved from one label to the other. This is similar to the tradeoff mechanism in [6], but naturalness is not bounded by the structure of labels.

An objective of such concept mapping is to predict the entry-level labels of an image. A direct solution to this problem is making prediction according to the labels in the ImageNet hierarchy, and then using the above naturalness idea to perform the optimal concept mapping. This can also be implemented by training a classifier using datasets with entry-level labels, such as the one introduced in [2]. The authors also showed that these two solutions can be combined to make joint prediction. The proposed methods were evaluated on 1000 image selected from [2]. The results show that the combined solution can achieved the best prediction results, followed by classifier trained on data with entry-level labels.

This paper has won the Marr prize in the 2013 International Conference on Computer Vision. It has covered a very interesting topic of unifying different labeling systems for image categoriza-

tion. The proposed method provides a viable solution to bridge the gap between different labels on the same object category, which can be useful in classification and retrieval of human annotated images.

References:

[1] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database". Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition,* pages 248-255, 2009.

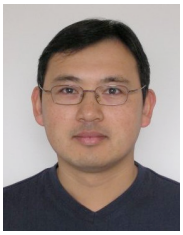
[2] V. Ordonez, G. Kulkarni, T. Berg. "Im2Text: Describing Images Using 1 Million Captioned Photographs". *Proceedings of the Neural Information Processing Systems*, pages 1143-1151, 2011.

[3] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. "WordNet: An online lexical database". *International Journal of Lexicograph*. 3, 4, pp. 235–244, 1990.

[4] S. Guadarrama, E. Rodner, K. Saenko, N. Zhang, R. Farrell, J. Donahue, and T. Darrell. "Open-vocabulary Object Retrieval". *Robotics: Science and Systems*, 2014.

[5] C. Rashtchian, P. Young, M. Hodosh, and J. Hockenmaier. "Collecting Image Annotations Using Amazon's Mechanical Turk". In *Proceedings of the NAACL HLT Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, 2010.

[6] J. Deng, J. Krause, A. Berg, and L. Fei-Fei. "Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3450-3457, 2012.



Jun Zhou received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Canada, in 2002, and the Ph.D. degree in computing science from University of Alberta, Canada, in 2006.

He joined the School of Information and Communication Technology in Griffith University as a lecturer in June 2012. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA. His research interests are in statistical pattern recognition, interactive computer vision, and their applications to hyperspectral imaging and environmental informatics.

Paper Nomination Policy

Following the direction of MMTC, the R-Letter platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia include, but are not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication.

Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings, or other distinguished journals/conferences within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to R-Letter Editorial Board Directors:

Christian Timmerer (christian.timmerer@aau.at), Weiyi Zhang (wzhang@ieee.org), and Yan Zhang (yanzhang@simula.no).

The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper when possible.

Review Process

Each nominated paper will be reviewed by members of the IEEE MMTC Review Board. To avoid potential conflict of interest, nominated papers co-authored by a Review Board member will be reviewed by guest editors external to the Board. The reviewers' names will be kept confidential. If two reviewers agree that the paper is of R-letter quality, a board editor will be assigned to complete the review letter (partially based on the nomination supporting document) for publication. The review result will be final (no multiple nomination of the same paper). Nominators external to the board will be acknowledged in the review letter.

R-Letter Best Paper Award

Accepted papers in the R-Letter are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board).

For more details, please refer to http://committees.comsoc.org/mmc/rletters.asp

Multimedia Communications Technical Committee Officers

Chair: Yonggang Wen, Singapore
Steering Committee Chair: Luigi Atzori, Italy
Vice Chair – North America: Khaled El-Maleh, USA
Vice Chair – Asia: Liang Zhou, China
Vice Chair – Europe: Maria G. Martini, UK
Vice Chair – Letters: Shiwen Mao, USA
Secretary: Fen Hou, China
Standard Liaison: Zhu Li, USA

MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.