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

E-LETTER

Vol. 10, No. 2, March 2015

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from MMTC Chair

Howdy MMTC colleagues,

Welcome to the March 2015 issue of MMTC E-Letter! As the past E-Letter Director, I would like to take this opportunity to thank the past MMTC officers, my co-Directors Drs. Periklis Chatzimisios and Guosen Yue, the MMTC editors, the Interest Groups (IG), and the E-Letter authors for their guidance, cooperation and support in the past two years. Our fruitful collaboration resulted in 12 E-Letter issues with 23 special issues on timely topics.

It is a great honour and pleasure to serve as vice Chair—Letters & Member Communications for the term of 2014 ~ 2016. I am really excited about continuing my involvement with MMTC, one of the most vibrant technical committees of IEEE ComSoc. I will certainly strive to work with the MMTC officers, the E-Letter, R-Letter and Membership boards, the IGs, and MMTC members to better serve the MMTC community and to continue the past success of MMTC.

I would also like to take this opportunity to provide an update of the E-Letter, R-Letter and Membership boards. Last fall, ten E-Letter Editors, including Drs. Florin Ciucu, Markus Fiedler, Michelle X. Gong, Cheng-Hsin Hsu, Zhu Liu, Konstantinos Samdanis, Joerg Widmer, Yik Chung Wu, Weiyi Zhang, and Yan Zhang, and two R-Letter Editors, including Drs. Gene Cheung and Guillaume Lavoué, retired after two years of excellent service. We thank them for their contributions to the two letters. Each retired editors received a Certificate of Appreciation from IEEE ComSoc as a token of our appreciation for their hard work.

It is my great pleasure to introduce to you the new E-Letter and R-Letter Board leaders for the term of 2014 ~ 2016:

- E-Letter Director: Dr. Periklis Chatzimisios (Alexander Technological Educational Institute of Thessaloniki, Greece)
- E-Letter Co-Director: Dr. Guosen Yue (Broadcom, USA)
- E-Letter Co-Director: Dr. Honggang Wang (University of Massachusetts Dartmouth, USA)

- R-Letter Director, Dr. Christian Timmerer (Klagenfurt University Austria)
- R-Letter Co-director, Dr. Weiyi Zhang (AT&T Labs Research, USA)
- R-Letter Co-director, Dr. Yan Zhang (Simula Research Laboratory, Norway)

The lists of new E-Letter and R-Letter editors can be found at the MMTC website. The new E-Letter and R-Letter teams have been working hard on timely publication of the letter issues. Since last September, they have successfully published three E-Letter issues (with six special issues) and four R-Letter issues. It is fortunate to have such a distinguished team and I believe that these two important MMTC publications are in good hands!

In 2013, MMTC established the MMTC Excellent Editor Awards to recognize the outstanding contributions of E-Letter and R-Letter editors. For Year 2014, the MMTC Excellent Editor Awards awardees are:

- Dr. Florin Ciucu, University of Warwick, UK, E-Letter Editor
- Dr. Jun Zhou, Griffith University, Australia, R-Letter Editor

Please join me to congratulate Drs. Ciucu and Zhou for this well-deserved recognition and thank them for their hard work and contributions.

In addition to having two top-notch boards, we also explore other new initiatives to enhance the quality and impact of both letters. While serving on the steering committee of IEEE International Conference on Multimedia & Expo (ICME), Dr. Yonggang Wen and I are working with the ICME Steering Committee Chair, Dr. Jin Li, on fostering cooperation. Starting from ICME 2015, the E-Letter board will work with the ICME panel chairs to publish position papers that summarize the panel theme and discussions; the R-Letter board will participate in the evaluation of a small set of best paper/best paper candidates identified by the ICME TPC team, and review the ICME 2015 best papers in R-Letter. We are considering extend such collaboration to other MMTC sponsored conferences in the future.

With Dr. Yonggang Wen's help, we also started the talk with several major multimedia journals about possible cooperation. As a starting point, we have achieved agreements with the E-i-Cs of IEEE Transactions on Circuits and Systems for Video Technology (CSVT) and IEEE Multimedia, to include the table of contents of their most recent issues in E-Letter. Finally, with help from the Member Service Board, in particular, Dr. Dalei Wu, we have set up Google Analytics accounts for the E-Letter/R-Letter websites, which can provide monthly reports on download statistics of the letters. This would be helpful for assessing the impact of the letters.

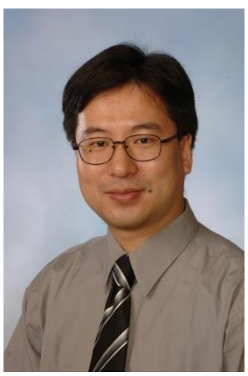
It is also my great pleasure to introduce to you the new Membership Board leaders for the term of 2014 ~ 2016:

- Director: Dr. Zhu Liu (AT&T Labs Research, USA)
- Co-Director: Dr. Lifeng Sun (Tsinghua University, China)
- Co-Director: Dr. Laura Galluccio (University of Catania, Italy)

In my opinion, the Membership board is probably the most important component of MMTC. It is crucial for the success of MMTC to attract active researchers, developers and students in the multimedia area to become a member and get involved in MMTC activities. With help from Drs. Yonggang Wen and Lifeng Sun, we are working on cooperation with China Computer Federation (CCF), a most influential computer related professional organization in China. Recently, we have 59 new members from CCF joined MMTC, thus bringing the number of active MMTC members to over one thousand. Let's congratulate the Membership Board for a good job done!

I hope you enjoy reading this E-Letter issue, and strongly encourage you find the IG of interest to get involved and to contribute to future letter special issues. If you have any suggestions or comments on improving the E-Letter, R-Letter and Membership boards, please do not hesitate to contact me.

Sincerely,



Shiwen Mao
Vice Chair—Letters & Member Communications
Multimedia Communications Technical Committee, IEEE ComSoc

EMERGING TOPICS: SPECIAL ISSUE ON FTV TECHNOLOGY AND APPLICATIONS

Guest Editors: Tasos Dagiuklas[1], Aljoscha Smolic[2] and Wanqing Li[3]

[1]*Hellenic Open University, Greece, dagiuklas@eap.gr*
[2]*Disney Research, Switzerland, smolic@disneyresearch.com*
[3]*University of Wollongong, Australia, wanqing@uow.edu.au*

Free Viewpoint Video (FVV) or Free Viewpoint Television (FVT) is an innovative visual media enabling us to view a three-dimensional (3-D) scene by freely changing our viewpoints. Current FTV standardization targets three very specific application scenarios: Super Multiview Displays where hundreds of very densely rendered views provide horizontal motion parallax for realistic 3D visualization, Integral Photography where 3D video with both horizontal and vertical motion parallax are captured for realistic display and Free Navigation allowing the user to freely navigate or fly through the scene.

This Special Issue of E-Letter focuses on the recent progresses of FTV Technologies. It is the great honor of the editorial team to have four leading research groups, from both academia and industry laboratories, to report their solutions for meeting these challenges and share their latest results.

The first article is contributed by Masayuki Tanimoto and is entitled "FTV Technologies and Standards". The paper introduces FTV technology in terms of Ray-Spacing, Capturing and Multiview Display Technologies. The paper outlines FTV standardization history and presents three applications scenarios for the use of FTV technology.

The second article is contributed by Yebin Liu, Jingtao Fan and Qionghai Dai with the title "3D Reconstruction of Real-world Visual Scenes using Light-View-Time Images". The paper proposes the "LiViTi (Light-View-Time) Space" as a basic data unit used in capture and reconstruction of visual information. The LiViTi space is a three-dimensional space spanning the light, view and time coordinates, with the basic element an image obtained by a standard image/video camera. Recent reconstruction researches based on sparse sampling using new computational cameras and random sampling is all interesting and important extensions of the LiViTi space computing. Such samplings will further free the constraints on the acquisition of visual information and make reconstruction more practical.

The third article is contributed by Hong Ren Wu, Damian M. Tan and David Wu and is entitled "Perceptual Coding of Three-Dimensional (3-D) Video". The paper focuses on a number of issues regarding the paradigm shift from bitrate driven design to visual quality driven design for 3-D video coding (3-DVC) based on rate-perceptual-distortion (RpD) theory, which assures delivery of agreed visual quality in user-centric visual communication, broadcasting and entertainment services.

The forth article is contributed by Hooman Shidanshidi, Farzad Safaei, and Wanqing Li and is entitled "Effective Sampling Density and Its Applications to the Evaluation and Optimization of Free Viewpoint Video Systems". The paper presents the concept of Effective Sampling Density (ESD) for light field based free viewpoint video (FVV) systems have been developed in last six years. It was shown that ESD is a tractable metric that can be determined from FVV system parameters and can be used to directly estimate output video quality without access to the ground truth.

While this special issue is far from delivering a complete coverage on this exciting research area, we hope that the four invited letters give the audiences a taste of the main activities in this area, and provide them an opportunity to explore and collaborate in the related fields. Finally, we would like to thank all the authors for their great contribution and the E-Letter Board for making this special issue possible.

Tasos Dagiuklas received the received the Engineering Degree from the University of Patras-Greece in 1989, the M.Sc. from the University of Manchester-UK in 1991 and the Ph.D. from the University of Essex-UK in 1995, all in Electrical Engineering. He is Assistant Professor at the School of Science and Technology at the

Hellenic Open University, Greece. He is leading the Converged Networks and Services (CONES) Research Group (http://cones.eap.gr). Dr Dagiuklas is Senior Member of IEEE, Chair for IEEE MMTC 3DIG WG and IEEE MMTC E-Board Member. He has served as Vice-Chair for IEEE MMTC QoE WG and Key Member for IEEE MMTC MSIG and 3DRPC WGs. He is serving as Associate Technical Editor for IEEE Communications Magazine. He has served as Guest Editor in many scientific journals. He is a reviewer for journals such as IEEE Transactions on Multimedia, IEEE Communication Letters and IEEE Journal on Selected Areas in Communications. His research interests include FTV, 3DV, Media Optimization across heterogeneous networks, QoE and cloud infrastructures and services. Dr Dagiuklas is a Senior Member of IEEE and Technical Chamber of Greece.

Dr. Aljoša Smolić joined Disney Research Zurich, Switzerland in 2009, as Senior Research Scientist and leader of the "Advanced Video Technology" group. Before he was Scientific Project Manager at Fraunhofer HHI, Berlin, also heading a research group. He has been involved in several national and international research projects, where he conducted research in various fields of visual computing and video coding, and published more than 100 scientific papers in these fields. In current projects he is responsible for applied research in visual computing in close cooperation with various business units of The Walt Disney Company. Dr. Smolić received the Dipl.-Ing. Degree in Electrical Engineering from the Technical University of Berlin, Germany in 1996, and the Dr.-Ing. Degree in Electrical Engineering and Information Technology from Aachen University of Technology (RWTH), Germany, in 2001. He received the "Rudolf-Urtlel-Award" of the German Society for Technology in TV and Cinema (FKTG) for his dissertation in 2002. He is Area Editor for Signal Processing: Image Communication and served as Guest Editor for the Proceedings of the IEEE, IEEE Transactions on CSVT, IEEE Signal Processing Magazine, and other scientific journals. He is Committee Member of several conferences, including ICIP, ICME, and EUSIPCO and served in several Chair positions of conferences. He chaired the MPEG ad hoc group on 3DAV pioneering standards for 3D video. In this context he also served as one of the Editors of the Multi-view Video Coding (MVC) standard. Since many years he is teaching full lecture courses on Multimedia Communications and other topics, now at ETH Zurich.

Wanqing Li received his PhD in electronic engineering from The University of Western Australia (1998). He was a Lecturer (1987-90) and Associate Professor (1991-92) at the Department of Computer Science, Hangzhou (now Zhejiang) University. He joint Motorola Lab in Sydney (1998-2003) as a Senior Researcher and later a Principal Researcher. He received Motorola CTO's Award in 2003 for his significant technical contribution to video-based motion analysis. He was a visiting researcher at Microsoft Research US in 2008, 2010 and 2013. He is currently an Associate Professor and Director of Advanced Multimedia Research Lab (AMRL) of University of Wollongong, Australia. His research areas are 3D computer vision and 3D multimedia signal processing, including 3D reconstruction, human motion analysis, detection of objects and events, and free-viewpoint video. He has published one book, over 100 referred articles in the leading journals and international conferences in multimedia and computer vision.

Dr. Li is a Senior Member of IEEE and currently a co-chair of the 3D Rendering, Processing and Communications interest group, Multimedia Technical Committee of IEEE Communication Society. He is the guest editor of the special issue on Human activity understanding from 2D and 3D data (2015), International Journal of Computer Vision, and the special issue on Visual Understanding and Applications with RGB-D Cameras (2013), Journal of Visual Communication and Image Representation. He served as a Co-organizer of the IEEE International workshop on Human Activity Understanding from 3D Data (HAU3D) (2011-2013) and Hot Topics in 3D multimedia (Hot3D) (2014), an area chair of International Conference on Multimedia & Expo (ICME) 2014, a publication chair of IEEE Workshop on Multimedia Signal Processing (MMSP) 2008, General Co-Chair of ASIACCS'09 and DRMTICS'05, and technical committee members of numerous international conferences and workshops.

FTV Technologies and Standards

Masayuki Tanimoto

Nagoya Industrial Science Research Institute, Japan
tanimoto@nagoya-u.jp

1. Introduction

4k/8k UHDTV (Ultra High-Definition TV) has realized viewing at the highest resolution in visual media. However, it transmits only a single view and users can't change the viewpoints. It is still far away from our viewing in the real world.

The next challenge comes to FTV (Free-viewpoint TV) [1]-[5]. FTV enables users to view a scene by freely changing the viewpoints as we do naturally in the real world. FTV is the ultimate 3DTV with infinite number of views and ranked as the top of visual media. FTV will give a great impact on the various fields of our life and society.

FTV has been realized by novel capture, processing and display technologies. MPEG has been developing FTV standards since 2001. In this paper, FTV technologies and standards are summarized.

2. FTV Technologies

Ray-Space

FTV was developed based on the ray-space method [6]. Two types of ray-space are defined. One is orthogonal ray-space and another is spherical ray-space. The orthogonal ray-space and the spherical ray-space are used for FTV with linear camera arrangement and circular camera arrangement, respectively.

An example of the orthogonal ray-space is shown in Figure 1. The horizontal cross-section of orthogonal ray-space has a line structure. A free-viewpoint image is generated by cutting the ray-space vertically with a knife at a position determined by the viewpoint. Parallel knives are used to generate multiple views for a multiview display.

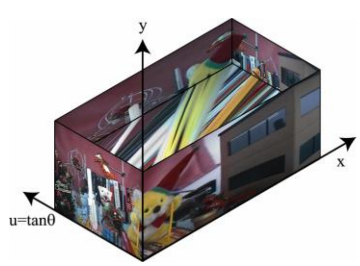


Figure 1. Orthogonal ray-space and a horizontal cross-section.

An example of the spherical ray-space is shown in Figure 2. The horizontal cross-section has a sinusoidal structure. Sinusoidal knives are used for view generation from the spherical ray-space.

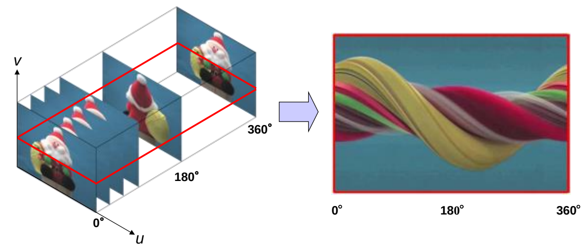


Figure 2. Spherical ray-space and a horizontal cross-section.

Capture

A "100-camera system" with a hundred of HDTV cameras was developed for view capture. The camera arrangement is flexible as shown in Figure 3.

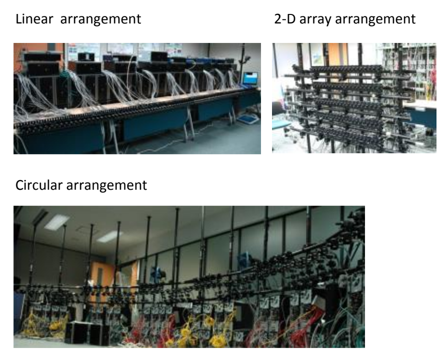


Figure 3. Flexible arrangement of 100-camera system.

User Interface

Various types of user interface were developed for FTV as shown in Figure 4.

Figure 4. Various types of FTV user interface.

Super Multiview Technologies

Super multiview denotes a very high number and high density of views. It is needed for realistic 3D viewing with smooth motion parallax and without eye fatigue. Super multiview ray-reproducing FTV shown in Figure 5 captures 360 all-around views by mirror-scan and displays them on a cylindrical display [7].

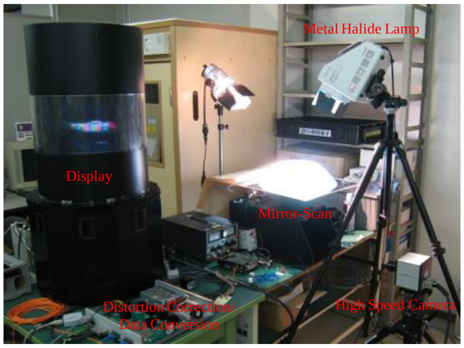


Figure 5. Super multiview ray-reproducing FTV.

3. History of FTV Standardization

MPEG has been developing FTV standards in 3 phases since 2001 as shown in Figure 6. The first phase of FTV is MVC (Multi-view Video Coding) [8] and the second phase of FTV is 3DV (3D Video) [9]. Although MVC has the same number of input and output views, 3DV transmits a small number of views and increases number of views for multiview displays by view synthesis. The third phase of FTV started in August 2013 targets super multiview and free navigation applications [10].

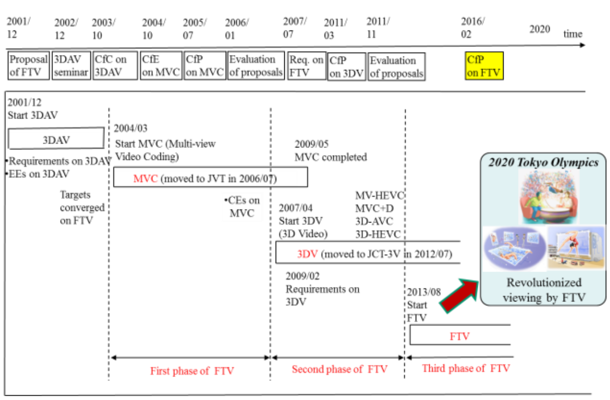


Figure 6. History of FTV standardization in MPEG.

4. MVC Standard

The framework of MVC is shown in Figure 7. MVC is an efficient coding standard of multiview video. In MVC, the number of input views is the same as that of output views. The view synthesis function of FTV is not included in MVC. MVC was standardized as the extension of H.264/MPEG4-AVC [8]. It has been adopted by Blu-ray 3D.

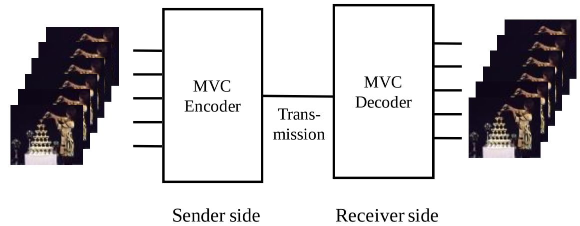


Figure 7. Framework of MVC.

5. 3DV Standard

The framework of 3DV is shown in Figure 8. 3DV gives efficient joint coding standards of multiview plus depth. The coding is based on AVC or HEVC. 3DV sends a set of view and depth, and increases the number of views at the receiver side for multiview displays by depth-assisted view synthesis. 3DV enables multiview display adaptation and viewing adaptation [9].

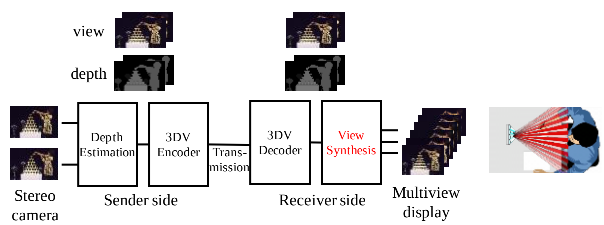


Figure 8. Framework of 3DV.

6. FTV Standard

The framework of FTV is shown in Figure 9. The output of FTV is super multiview for super multiview displays or a single view for free navigation.

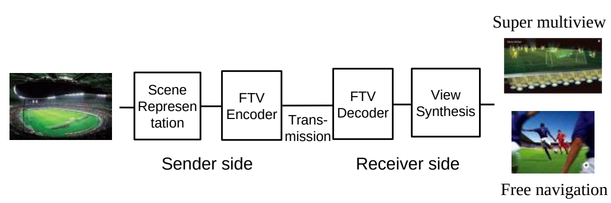


Figure 9. Framework of FTV.

Motivation and Background

In 2010, FIFA World Cup Japan Bid Committee planned to deliver the excitement on soccer stadium of 2022 FIFA World Cup to the world by FTV. It aimed to revolutionize the viewing of the soccer game by free navigation and super multiview 3D viewing as shown in Figure 10. This has become strong motivation for the third phase of FTV.

Figure 10. Revolutionized viewing by FTV.

As a technical background, super multiview capture and display technologies are already available as shown in Figure 11 and Figure 12, respectively.

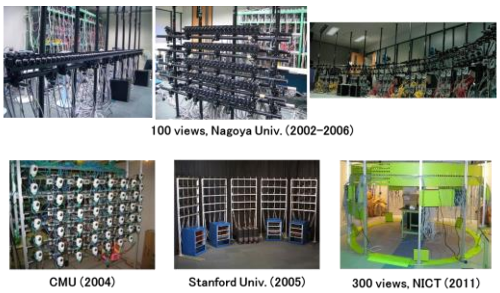


Figure 11. Super multiview capture technology.

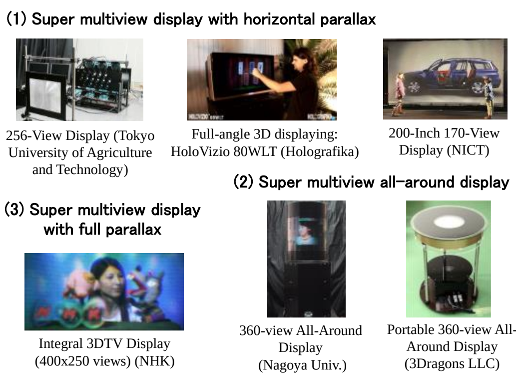


Figure 12. Super multiview display technology.

Based on these motivation and background, the third phase of FTV sets the following application scenarios.

Application Scenario 1: Super Multiview

In this application, users can enjoy very realistic glasses-free 3D viewing. An example is shown in Figure 13. It is a 360-degree viewable 3D display, "Holo-Table" [11]. It displays 360 views with 1-degree interval. The features of this display are not only a high number of views but also high density of views. Because of these features, users can see a 3D scene from any directions with smooth motion parallax and without eye fatigue.

Figure 13. Holo-Table (3Dragons LLC).

Another example is Integral 3DTV [12] shown in Figure 14. It has 400x250 views and provides both horizontal and vertical parallax.

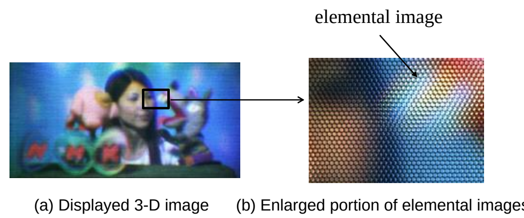


Figure 14. Integral 3DTV (NHK).

Application Scenario 2: Free Navigation

Another application is free navigation, where users can experience walk-through or fly-through of 3D scene. Figure 15 shows free navigation viewing of soccer games. Users can enjoy free navigation on mobile devices as shown in Figure 16.

Figure 15. Free navigation viewing of soccer games (KDDI).

Figure 16. FTV on mobile devices (left: Nagoya University, right: Orange Labs Poland).

Challenges

FTV standard for super multiview addresses the following challenges [10]: (a) very wide viewing range, (b) both horizontal and vertical parallax, (c) smooth transition between adjacent views and motion parallax, i.e., “walk-around” viewing, and (d) reduced eye fatigue. Furthermore, FTV standard has to support free navigation of a natural 3D scene with wide baseline and arbitrary camera positioning.

7. Conclusion

FTV is the top of visual media and provides many interesting research topics. The examples are scene representation of a large-scale space, integration of image-based and model-based approaches, integration of coding and synthesis, allocation of computational power between center and terminal, use of emerging capture and display devices etc.

Super multiview and free navigation services of FTV will revolutionize viewing of 3D scene. Attractive FTV services will be introduced by integrating FTV technologies and standards.

The FTV reflector of MPEG can be joined at the following site.
http://lists.uni-klu.ac.at/mailman/listinfo/ftv

Acknowledgement

This research was partially supported by JSPS KAKENHI Grant Number 25289117, Hoso Bunka Foundation Grant and Grant of Support Center for Advanced Telecommunications Technology Research.

References

[1] Masayuki Tanimoto, “Free Viewpoint Television,” The Journal of Three Dimensional Images, Vol.15, No.3, pp.17-22, September 2001 (in Japanese).

[2] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, Tomohiro Yendo, “Free-Viewpoint TV”, IEEE Signal Processing Magazine, Vol.28, No.1, pp.67-76, January 2011.

[3] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, Tomohiro Yendo, “FTV for 3-D Spatial Communication”, Proceedings of the IEEE, Vol. 100, No. 4, pp. 905-917, April 2012. (invited paper)

[4] Masayuki Tanimoto, FTV: Free-viewpoint Television” , Signal Processing : Image Communication, Vol. 27, Issue 6, pp. 555-570, June 2012. doi:10.1016/j.image.2012.02.016 (online 24 February 2012). (invited paper)

[5] Masayuki Tanimoto, “FTV (Free-viewpoint Television)”, APSIPA Transactions on Signal and Information Processing, Vol. 1, Issue 1, e4 (14 pages) (August 2012). doi: 10.1017/ATSIP.2012.5 (invited paper)

[6] T. Fujii, T. Kimoto and M. Tanimoto, “Ray Space Coding for 3D Visual Communication,” Picture Coding Symposium 1996, pp. 447-451, March 1996.

[7] T. Yendo, T. Fujii, M. P. Tehrani and M. Tanimoto, “All-Around Ray-Reproducing 3DTV”, IEEE International Workshop on Hot Topics in 3D (Hot 3D), July 2011.

[8] A. Vetro, T. Wiegand, and G. J. Sullivan, “Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard”, Proc. IEEE, Vol. 99, No. 4, pp. 626–642, Apr. 2011.

[9] “Applications and Requirements on 3D Video Coding”, ISO/IEC JTC1/SC29/WG11 N10570, April 2009.

[10] “ Use Cases and Requirements on Free-viewpoint Television (FTV),” ISO/IEC JTC1/SC29/WG11 MPEG2014/N14178, San Jose, US, January 2014.

[11] Hideyoshi Horimai and Masayuki Tanimoto, “Super Multiview 3D Display Holo-Table/Holo-Deck Using Direct Light Scanning and Its Specification”, ISO/IEC JTC1/SC29/WG11 (MPEG), M35037, October 2014.

[12] J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Haino, M. Yoshimura, M. Furuya, and M. Sato, “Integral Three-Dimensional Television Using a 33-Megapixel Imaging System,” Journal of Display Technology, Vol. 6, No. 10, pp.422-430, October 2010.

Masayuki Tanimoto received his B.E., M.E., and Dr.E. degrees in electronic engineering from the University of Tokyo in 1970, 1972, and 1976, respectively. He joined Nagoya University in 1976. He was a Professor at Graduate School of Engineering, Nagoya University from 1991 to 2012. In 2012, he retired from Nagoya University and became an Emeritus Professor. Currently, he is a Senior Research Fellow at Nagoya Industrial Science Research Institute. He has been engaged in the research of image coding, image processing, 3D imaging, FTV and ITS.

Prof. Tanimoto was the president of the Institute of Image Information and Television Engineers (ITE). He is IEEE Life Fellow, Honorary Member of the ITE and Fellow of the IEICE (Institute of Electronics, Information, and Communication Engineers). He received the ITE Distinguished Achievement and Contributions Award, the IEICE Achievement Award, and the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science, and Technology.

3D Reconstruction of Real-world Visual Scenes using Light-View-Time Images

Yebin Liu, Jingtao Fan and Qionghai Dai

Automation Department, Tsinghua University
{*liuyebin, fanjingtao, qhdai*}*@tsinghua.edu.cn*

1. Introduction

Vision is the most important perception channel for humans to perceive the world. The acquisition and reconstruction of visual information using computer vision technologies helps humans to better understand the visual world, and now serves as an indispensable component in computer science. This paper proposes the "LiViTi (Light-View-Time) Space" as a basic data unit used in capture and reconstruction of visual information. The LiViTi space is a three-dimensional space spanning the ***light***, **view** and ***time*** coordinates, with the basic element an image obtained by a standard image/video camera, as shown in Fig.1.

The three 1D subspace of the LiViTi space consists of computer vision problems focusing on the view, light and time dimensions, respectively. First, with fixed view and illumination, optical flow is the main problem located in the time dimension. Second, stereo matching and image based rendering are the two methods mainly considered in the view dimension, and both of them assume constant illumination and static object targets (constant time). Third, image based relighting and photometric stereo problems are the key problems defined in the light dimension for images of a static scene sharing the same viewpoint.

In the higher subspaces, the combinations of any of the two dimensions define three planes in the LiViTi space. First, the green plane is the view-time space, targeting problems such as spatio-temporal 3D reconstructions and free-viewpoint video. The second is the view-light space (the pink plane), which contains, for example, multi-view photometric stereo techniques. Third is the light-time space (the blue plane), which has not been well investigated but still implies problems such as optical flow under varying illumination. Finally, the full LiViTi computing is the killer-application, refers to simultaneously sampling and reconstruction of all the three dimensions, and then rendering real-world dynamic scenes or objects at arbitrary time instant, from arbitrary view, under arbitrary illumination.

Since the problems in the low dimension LiViTi space are well studied, this paper will focus on the survey of some representative researches on acquisition and reconstruction of real-would visual information located on the higher and full dimension space.

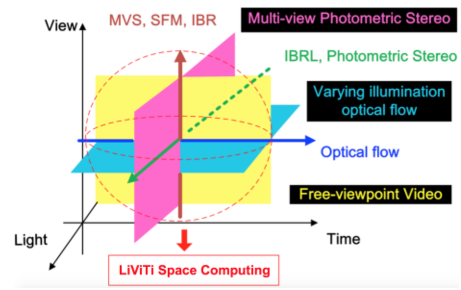


Fig. 1. ViLiTi space computing: sampling and reconstruction of real-world visual information, with typical vision and graphics problems fall into the subspaces.

2. View-Time Space Computing

View-time space computing involves the spatial-temporal sampling and the reconstruction of a dynamic scene, to synthesize images under views and at times that were not physically captured. The sampling of the view-time space usually requires the using of multi-camera systems.

The first multi-camera system for 3D reconstruction of dynamic object is the "VirtualRealty"[1] project. Multi-view stereo technique is adopted for free viewpoint video of human actors based on the multi-view images filmed by 51 cameras arranged on a dome. Later, with technique progresses on shape-from-silhouette and stereo, the number of cameras is greatly reduced to around 10. Typical examples includes multi-view depth estimation for open scene [2], real-time visual hull based methods [3], depth-map fusion methods [4], and graph-cut based methods [5]. Fig.2 shows the reconstruction models of the girls and the corresponding view-independent texture mapping results using our proposed depth-map fusion method [6]. Analysis and discussion of the relationship between the synthesized performance and the number of cameras are recently discussed [7, 8].

Figure 2. 3D Reconstruction and texture mapping of dynamic scene using a depth-map fusion method [6].

For view-time space computing, the taking advantage of the temporal information is vital for applications such as motion capture, analysis and editing of the target 3D model. If a 3D model of a specific time instant is available, more efficient reconstruction scheme would be to deform the model based on the temporal features such as silhouette, SIFT, optical flow and scene flow [7]. In this manner, temporal coherence on model topology may be maintained and spatial characteristic can be employed to improve the reconstruction accuracy. Examples of this temporal dominated reconstruction include skeleton based performance capture [8] and mesh based performance capture [9].

The reconstruction of much more complex scene in the view-time space with multiple non-rigid moving objects have already been considered. A markerless motion capture approach [10] is proposed to reconstruct the skeletal motion and the detailed time-varying surface geometry of multiple closely interacting people. Due to ambiguities in feature-to-person assignments and frequent occlusions, it is not feasible to directly apply markerless motion capture approaches [11] to the multi-person case. To solve this puzzle, a combined image segmentation and tracking approach was designed to overcome these difficulties. With the labeled pixels (see the second row of fig.3 left), single-person markerless motion and surface capture approach can be applied to each individual body to archive high quality reconstruction results as shown in the last row of the left figure in fig.3.

Figure 3. 3D Reconstruction and texture mapping of dynamic scene using depth-map fusion method [10].

3. Light-View Space Computing

The light-view space computing is another hotspot for scene reconstruction. Multi-view photometric stereo and 3D relighting are problems defined in this subspace. Given the sampled images, the former tries to recover view geometry, scene geometry, illumination and scene reflectance entangled in the light-view space. The later renders unknown images lying in the view-light space span by the given images.

Wu et al. [12] capture compact but sparse sampling on the light-view space. A multi-view stereo algorithm under multiple illuminations is first introduced to get an initial 3D model. Then, based on the fact that the light transport matrix on each model vertex is sparse, the normal of each vertices and the illuminations can be computed, which serve to improve the accuracy of each model. The algorithm assumes Lambertian surface but allows for non-uniform surface and have neither pre-knowledge nor constraint on the illuminations. The right of Fig.3 shows some of the reconstruction results from this work.

4. Time-Light Space Computing

Data capture on the time-light space requires only a single camera with multiple illuminations. As stated above, it is impossible to capture 2 images of the dynamic scene under the same viewpoint with different lights. Therefore, the sampling of time-light subspace is usually incomplete and successive frames are under different illuminations. If multiple illumination images can be registered to an anchor frame, then photometric stereo can be imposed to solve both the surface normal of the object and the outer space illumination. However, it is often difficult to register neighboring images using state-of-the-art optical flow technique.

The "Light stage" [13] is one of the representative works belonging to the time-light space computing. The system uses multiple controllable LEDs and a high-speed camera to capture human face performance under varying illumination. The captured images can be classified into two kinds: target frame (or tracking frames, captured under standard illumination) and basic frames (captured under cycling basic illuminations). By using conventional optical flow techniques on the tracking frames, and assuming linear motion between neighbor tracking frames, images under varying illuminations can be registered to the same time instant to obtain compact image set in the time-light space. Finally, the image based relighting technique [14] can be adopted to relight any time instant frame.

If the frame rate of the cameras used is not fast enough, usually it is impossible for available optical methods to register the serious motion of successive tracking frames. To resolve this difficulty, Hernandez [15] introduced the using of color lights in the light-view space sampling with technique named "color multiplex sampling". By assuming that the color and the reflectance of the object are uniform, the single view video captured by the color camera can be decomposed into (R, G, B) channels. As a result, each time instant may have three images under different illuminations, which allows for surface normal estimation using photometric stereo techniques.

5. Full LiViTi Space Computing

The technique for the full LiViTi space computing is to simultaneously reconstruct geometry, reflectance and motion of dynamic scenes using the full view-light-time space images. Intuitively, this would require a complicated and systematic sampling using multiple cameras and multiple illumination equipment. The reconstruction problem is a challenging and mostly relies on the combinative optimization using techniques from the low dimensional spaces.

Fu et al. [16] present an acquisition and reconstruction system for high-quality 3D reconstruction of dynamic objects working in the full LiViTi space. Firstly, multi-view images of the moving object are captured under periodically varying illumination using a multi-camera multi-light system. The dome system is a 6m-diameter hemisphere with 20 FLEA2 cameras uniformly located on a ring in the dome. The camera resolution is 1024 × 768 and the frame rate is 30fps. The acquisition system is hardware controlled to provide 6 periodical varying illuminations (Fig.4) to illuminate the scene where a character performs arbitrary motions.

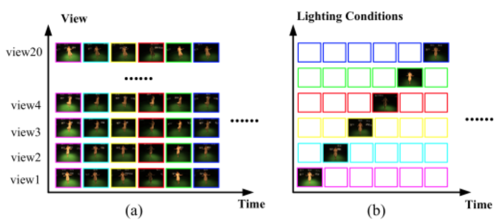


Figure 4. Image interpolation under variant illuminations. (a) the different color box indicates the images captured under different illuminations; (b) the images are captured by a camera on the dome, the white blocks denotes the images required to be interpolated based on the optical flow estimation.

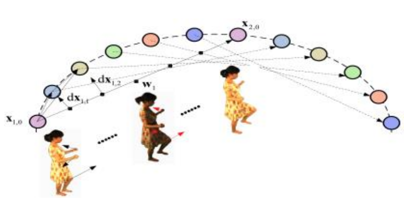


Figure 5. Periodical-illumination optical flow under variant illumination.

Second, a periodical-illumination optical flow estimation method is proposed to register images from the same camera (Fig.5), and then synthesize images under different illuminations for each viewpoint (Fig.4). By taking advantage of the periodical illuminations, this method calculates pixel correspondences between each pair of consecutive frames. Using the pixel correspondences, in each camera view, frames under different illuminations are warped to a certain time step to get the interpolated multi-lighting frames at this time step, as shown in Fig.5.

Finally, for each time instant, multi-view images under the same illumination are used to perform multi-view stereo [10] and obtain 3D models for each instant. The problem is then convert to the light-view space computing. Based on the initial 3D models, the surface normals are recovered using the proposed light-view space [12] described in Sect.2. Finally, the 3D models are improved by fusing the recovered normal and the vertex position of the initial model. The final results on many captured sequences are shown in Fig. 6.

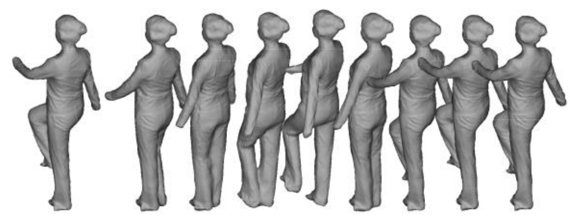


Figure 6. Reconstruction of surface details in the full LiViTi space.

6. Conclusion

This paper proposes the LiViTi space computing as a framework for summarize available researches on acquisition and reconstruction of visual information. Future works may cover more interesting sampling strategies in the LiViTi space. Recent reconstruction researches based on sparse sampling using new computational cameras [17] and random sampling [18] are all interesting and important extensions of the LiViTi space computing. Such samplings will further free the constraints on the acquisition of visual information and make reconstruction more practical.

References

[1] P. Rander, P. J. Narayanan, and T. Kanade, "Virtualized real- ity: constructing time-varying virtual worlds from real world events," in IEEE Visualization, pp. 277–284, 1997.

[2] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. A. J. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM TOG (Proc. SIGGRAPH), vol. 23, no. 3, pp. 600–608, 2004.

[3] F. Jean-Sbastien and B. Edmond, "Exact polyhedral visual hulls," in British Machine Vision Conference (BMVC), pp. 329–338, 2003.

[4] D. Bradley, T. Popa, A. Sheffer, W. Heidrich, and T. Boubekeur, "Markerless garment capture," ACM Trans. Graph., vol. 27, no. 3, pp. 1–9, 2008.

[5] J. Starck and A. Hilton, Surface capture for performance based animation, IEEE Computer Graphics and Applications, vol. 27(3), pp. 21–31, 2007.

[6] Y. Liu, Q. Dai, and W. Xu, A point-cloud-based multi-view stereo algorithm for free-viewpoint video, IEEE trans. Visualization and Computer Graphics, vol. 16, no. 3, pp. 407–418, May/June 2010.

[7] H. Shidanshidi, F. Safaei, A.Z. Farahani, W. Li, Non-uniform sampling of plenoptic signal based on the scene complexity variations for a free viewpoint video system. ICIP 2013, 3147-3151, 2013.

[8] L. Fang, N. -M. Cheung, D. Tian, A. Vetro, H. Sun, O. C. Au, An Analytical Model for Synthesis Distortion Estimation in 3D Video, IEEE Trans. Image Processing, 23(1), 185-199, 2014.

[9] S. Vedula, S. Baker, P. Rander, R. T. Collins, and T. Kanade, Three-dimensional scene flow, IEEE Trans. PAMI, vol. 27, no. 3, pp. 475–480, 2005.

[10] D. Vlasic, I. Baran, W. Matusik, and J. Popović, "Articu- lated mesh animation from multi-view silhouettes," ACM Trans. Graph, vol. 27, no. 3, 2008.

[11] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun, "Performance capture from sparse multi-view video," ACM TOG (SIGGRAPH), vol. 27, pp. 1–10, 2008.

[12] Y. Liu, J. Gall, C. Stoll, Q. Dai, H.-P. Seidel, C. Theobalt, Markerless Motion Capture of Multiple Characters Using Multi-view Image Segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, 35(11), pp. 2720-2735, 2013

[13] J. Gall, C. Stoll, E. Aguiar, C. Theobalt, B. Rosenhahn, and H.-P. Seidel, "Motion capture using joint skeleton tracking and surface estimation," in CVPR, pp. 1746–1753, 2009.

[14] C. Wu, Y. Liu, Q. Dai, and B. Wilburn, "Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination," IEEE Transactions on Visualization and Computer Graphics, vol. 17, pp. 1082–1095, 2011.

[15] A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, and P. Debevec, "Performance relighting and reflectance transformation with time-multiplexed illumination," ACM Trans. Graphics, vol. 24, July 2005.

[16] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in ACM SIGGRAPH, 2000.

[17] C. Hernandez, G. Vogiatzis, G. Brostow, B. Stenger, and R. Cipolla, Non-rigid photometric stereo with colored lights, ICCV, 2007.

[18] Ying Fu, Yebin Liu, Qionghai Dai, Dynamic Shape Capture via Periodical-illumination Optical Flow Estimation and Multi-view Photometric Stereo, 3DIMPVT, Hangzhou, China, May 17-21, 2011.

[19] Genzhi Ye, Yebin Liu, Yue Deng, Nils Hasler, Xiangyang Ji, Qionghai Dai, Christian Theobalt, Free-viewpoint Video of Human Actors using Multiple Handheld Kinects, IEEE Trans. System, Man & Cybernetics Part B, 43(5), pp. 1370-1382, 2013.

[20] A. Elhayek, C. Stoll, K.I. Kim, C. Theobalt, Outdoor human motion capture by simultaneous optimization of pose and camera parameters. Computer Graphics Forum, 2014.

Yebin Liu received the BE degree from the Beijing University of Posts and Telecommunica- tions, China, in 2002, and the PhD degree from the Automation Department, Tsinghua Univer- sity, Beijing, China, in 2009. He was a research fellow in the Computer Graphics Group of the Max Planck Institute for Informatik, Germany, in 2010. He is currently an associate professor at Tsinghua University. His research areas include computer vision and computer graphics.

Jingtao Fan received the B.E degree and M.E degree in computer science and technology, Ph.D degree in optical engineering from the Changchun University of Science and Technology, Changchun, China, in 2003, 2007 and 2013, respectively. He is currently a Post Doctor of Tsinghua University in China. His research interests mainly include the 3D video processing and computer vision.

Qionghai Dai received the BS degree in mathematics from Shanxi Normal University, China, in 1987, and the ME and PhD degrees in computer science and automation from North- eastern University, China, in 1994 and 1996, respectively. Since 1997, he has been with the faculty of Tsinghua University, Beijing, China, where he is currently a professor and the director of the Broadband Networks and Digital Media Laboratory. His research areas include computer vision, and graphics. He is a senior member of IEEE.

Perceptual Coding of Three-Dimensional (3-D) Video

Hong Ren Wu[1], Damian M. Tan[2], David Wu[2]

[1] Royal Melbourne Institute of Technology, Australia, henry.wu@rmit.edu.au
[2] HD[2] Technologies Pty. Ltd., Australia {damain.tan,david.wu}@hd2tech.com

1. Quality of visual experience

Technological advances in visual communications, broadcasting and entertainment continue to captivate the general public, offering new heights of viewing quality and experience which aim at clarity, precision and visual sensation by increasing spatiotemporal resolutions and bit depths of ultra-high definition (UHD) television (TV)/video [1], and realism and visual impact by three-dimensional digital video (3-DV) [2]. Multi-view video which goes beyond the stereoscopic 3-D video and free view-point TV/video offer a much higher degree of freedom and choice of view-points or viewing perspectives [3, 4].

A price has to be paid though, in order to deliver the quality increments and heightened visual experience promised by the aforementioned new video formats. For example, there is an eight-fold increase in the uncompressed data rate for 3-D full HD video compared with the standard definition (SD) video [3]. For an 8K (i.e., 7680×4320 pixels per frame) UHD video in its maximum configuration, there incurs a transmission data rate over 143 Gbps (Gigabits per second) in uncompressed form [1] in contrast with approximately 166 Mbps (Megabits per second) of uncompressed SD 4:2:2 component video. Multi-view video beyond the stereoscopic 3-D rendition and free view-point TV yield even higher data rates and require more advanced modelling and/or synthesis techniques [4, 5]. As a result, transportation of videos in these new formats puts an upward pressure on bandwidth demands of communication networks [6] and terrestrial broadcast bandwidth allocations [2, 7]. This requires new and effective compression theory, methods and techniques, as well as provides tremendous challenges and opportunities for ICT (information and communications techno-logy) and associated industries to bring about economical and social benefits to the global society.

Efficient and affordable digital visual communications have been made possible by data compression techniques based on Shannon's entropy theory for information lossless compression and rate-distortion (R-D) theory for information lossy compression [8]. The latter has supported most of practical applications to date. Entropy based coding design is obviously signal quality driven which mandates perfect reconstruction of a given original signal. Based on the R-D theory, two distinct design approaches are possible, i.e., bitrate controlled coding and quality (or distortion) controlled coding. For a constant bitrate coder to lay a claim to its effectiveness in compression performance, it has to hold a fixed bitrate and then to maximize picture quality, whereas for a constant quality coder to be effective, it must, first and foremost, be able to hold a given picture quality and, then, to do so at the lowest possible bitrate [8, 9].

While constant bitrate coding practice has been prevalent to date, insufficient coding bitrates, due to transmission bandwidth and storage constraints and/ or unregulated visual quality across spatiotemporal dimensions and views, confront existing bitrate driven design philosophy and approaches. This exacerbates visual traits of various spatial and temporal coding artifacts [10] and their spatial and/or temporal variations/fluctuations [11], which markedly erode quality visual experience as intended by HD/UHD videos, and render ineffective and/or inefficient 3-D video presentations [9].

This letter focuses on a number of issues regarding the paradigm shift from bitrate driven design to visual quality driven design for 3-D video coding (3-DVC) based on rate-perceptual-distortion (RpD) theory [9, 12], which assures delivery of agreed visual quality in user-centric visual communication, broadcasting and entertainment services. In Section 2, advances in 3-D video coding and assessment are briefly reviewed along with human visual perception-based coding of visual signals relevant to 3-DVC as the background. This identifies differences between the two design methodologies which require alternative quality assessment approaches. Section 3 refines a visual quality regulated 3-DVC framework extended from a previous work [13, 14] to illustrate key components in an RpD optimization (RpDO) based design.

2. Coding and assessment of 3-D video

The current research and development in 3-DVC have been predominantly led and propelled by international standards activities associated with multi-view video coding (MVC) extensions to ITU-T H.264/ISO/IEC advanced video coding (H.264/AVC) [15, 3] and ITU-T H.265/ISO/IEC high efficiency video coding (H.265/ HEVC) [5], respectively, with test conditions for 3DV core experiments specified in [16] and subjective quality assessment methods in [17]. Inter-view prediction is the central theme of MVC for reduction of spatiotemporal redundancies between the base view and its enhancement view(s), while MVC plus depth map (MVC+D) caters for depth image based rendering and synthesis for a multitude of views beyond the conventional stereoscopic 3-D video [2, 15, 3, 5]. The R-D performance is a key criterion where the bitrate and the PSNR (peak signal to noise ratio) are used as the measures [16], and subjective assessments are based on the MOS (mean opinion score) using a five-level or an eleven-level absolute category rating (ACR) [18, 17].

The supporting evidence was abundantly shown in the literature that these 3-DVC frameworks as the extensions to international video coding standards were effective in lifting the above said objective or subjective picture quality significantly with a given (anchor) bitrate compared with their prior work [15, 5]. What had been missing was the evidence to show whether these coding frameworks would be able to consistently deliver a prescribed or designated visual quality as discernible by human observers in terms of, e.g., JNND (just-not-noticeable difference) as well as JND (just-noticeable difference) levels [9] or VDUs (visual distortion units) [19].

Annoying visual experience due to various spatio-temporal video coding artifacts and their variations is, with all due respect, commonplace in applications of the existing single-view video coding standard implementations based on the rate driven design approach [10, 11]. 3-DVC which follows the same design principle with unregulated or ineffectively controlled visual quality will inhere all the artifacts associated with its single-view counterpart, and induce additional, often more severe, visual discomfort due to perceivable difference between coded matching areas of the two views above the JND and disparity-compensated coding residual fluctuations in the enhancement view which differs from those in the base view, and variation or even loss of depth perception due to visual quality differences above a certain threshold between the two views when visual quality is not (or cannot be) adequately regulated [9, 20]. It begs the question whether perceptual 3-DVC based on visual quality driven design and RpDO ought to be considered as a viable alternative to address these issues.

Perceptual coding techniques for single-view visual signal compression as classified in [9] may be extended to 3-DVC either as a constant bitrate coder to maximize visual quality for a given bitrate or as a constant quality coder to minimize coding bitrate for a designated visual quality. While there had been numerous and on-going reports on consistent performance gains in terms of bitrate savings and visual quality by perceptual coding of (single-view) visual signals, most of these perceptual coders did not maintain a constant visual quality. To date, successful visual quality regulated still image coding results reported in the literature are limited to the JNND or perceptually lossless level [21-23]. There were very limited reports on human perception-based 3-DVC which extended HVS (human visual system) models used in single-view video coding to 3-DVC [20, 24] and/or experimented with inter-view masking effect in asymmetric 3-D stereoscopic video coding [20, 25], where a bitrate driven coding design platform was often used as was the case with most of perceptual (single-view) video coding with few exceptions [13, 26]. Considering that performance in the three key primary perceptual dimensions of 3-DVC, including picture quality, depth quality/acuity and visual comfort [17], is significantly impacted by the visual quality of compressed base and enhancement views of 3-D video [20, 9], a visual quality driven design approach may offer a viable, if not imperative, solution to quality assured and sustainable 3-D visual communication services. The causes of a seemly impasse to the much needed paradigm shift from bitrate driven design to visual quality driven design for 3-DVC, or visual signal coding in general, deserve a closer examination and a number of issues have been highlighted in [27].

3. Towards visual quality regulated 3-DVC

The difference between the classical R-D optimization (R-DO) and the RpDO was examined in [27, 28] where "0" measured by the MSE (mean squared error) is replaced by the JNND (or JND_0 if you will) as the reference point while other nonzero MSE values by JNDs (or VDUs) on the distortion scale. In the RpDO based constant quality coder design, maintaining a bitrate deems redundant or ineffective if, e.g., it is neither sufficient to guarantee a distortion level at the JNND for perceptually lossless coding nor necessary to achieve JND_1 (i.e., JND) for a perceptually lossy coding.

For a perceptual distortion measure (PDM) to regulate spatiotemporal and stereoscopic visual quality in RpDO based 3-DVC, it is required to predict reasonably accurately discernible quality levels, i.e., JNDs, by human observers, to ensure a designated visual quality and to prevent severe visual discomfort due to, e.g., unstable or loss of depth perception [9]. A preliminary experimental investigation has been reported in [27, 28] to ascertain if various perceptual image distortion/ quality metrics reported in recent years are able to consistently grade different images at various JND levels. Images were generated using an open source JPEG 2000 coder at increasingly higher compression ratios for a total of eighty-one variations for each of forty-one well-known test images, providing a range of test pictures which capture the transition points between JND levels. An image at JND_n is determined relative to the image at $JND_{(n-1)}$ for $n > 1$, except for JND_1 which was relative to the reference such that JND_2 is relative to JND_1 and JND_3 to JND_2, etc. Perceptual distortion or quality measures were computed for sets of images at JND_1, JND_2, JND_3, JND_4, and JND_5, respectively. Assuming small data samples with normal distribution, the lower and the upper bounds of the metrics' predictions were computed with the mean and the standard deviation of each metric using a 95% confidence interval (CI), and prediction was considered inaccurate if the variation was such that most of the responses from a metric (i.e., >50%) sat outside the 95% CI range. It was found, as one might have expected, that acceptance rates of these metrics were lower than 50%, indicating that they were ineffective as a measure to predict discernible visual quality levels in terms of JNDs.

There may be a need for alternative subjective test methods, where quality 3-D video test sequences[1] are used to anchor the reference point with respect to visual quality scale tailored to applications to produce subjective data in terms of JNDs (or VDUs) which can be used in place of ACR based subjective data to model and parameterize PDMs for quality regulated 3-DVC design and automated assessment methods.

Another difficulty in visual quality regulated 3-DVC design is related to the fact that the available perceptual distortion measures embedded in a video coder to date are often formulated using an image based HVS model in either pixel or transform domain, instead of residual images which results from temporal or inter-view prediction [9]. JND thresholding [30] and JND adaptive inter-frame residual coding [31] have been reported where the prediction errors below the JND are eliminated from further coding and only residuals above the JND are adjusted with respect to the JND for further coding, which has been extended to perceptual 3-DVC in [24]. In either case, coding of inter-frame prediction residual to a designated visual quality level has not been established and is subject to further examination.

4. A visual quality regulated 3-DVC framework

HVS model based perceptual coding approach has demonstrated superior performance in terms of RpDO for still image coding and intra-frame coding of digital video against standard benchmarks, noticeably at perceptually lossless quality [21-23, 26, 9]. To address the issue with the lack of HVS models based on prediction residual image for video coding [9] and perceptual coding of 3-D video prediction error image [24], perceptual 3-DVC with hybrid inter-frame and inter-view prediction is formulated using RpD criterion [13, 14].

Perceptually lossless 3-D video coding

Fig.1 illustrates a perceptually lossless 3-DVC framework with hybrid motion- and disparity-compensated prediction, based on the discrete wavelet transform (DWT) decomposition as an example, where the left-view is the base view as reference and the right-view the enhancement view. In Fig.1, function block "EBCOT Encode" is the embedded block coding with optimal truncation [32] and "Optimise" finds the minimum bitrate of the input coded streams to output along with selected coding mode. The aim of function block "VLF" is to remove, via a visual filtering, as much data and psychovisual redundancies as possible from transform domain representations, minimizing the bitrate while keeping the PDM value equal to or below the JNND level [13, 14, 21, 23], i.e.,

$$\begin{aligned}\mathbf{v}_F &\triangleq T_{VLF}\left(\mathbf{v},\mathbf{v}_{ref}\right)\\ &= arg\min_{\forall \mathbf{v}_f} R_{BIT}\left\{\mathbf{v}_f \mid PDM\left(\mathbf{v}_O,\ \mathbf{v}_{O_f}\right) \le JNND\right\}\end{aligned} \quad (1)$$

where the operator T_{VLF} on a vector (matrix) variable $\mathbf{v}$ with a given reference vector variable $\mathbf{v}_{ref}$ returns the visually filtered coefficient vector $\mathbf{v}_F$ which delivers the perceptually lossless compression using the perceptual distortion measure *PDM* at the minimum bitrate, function $R_{BIT}\{\cdot\}$ computes the bitrate, $\mathbf{v}_O$ is the original image with respect to $\mathbf{v}$ in the DWT domain, $\mathbf{v}_{O_f}$ is a visually filtered coefficient vector as an approximation of $\mathbf{v}_O$ with respect to $\mathbf{v}$, and defined as

$$\mathbf{v}_{O_f} \triangleq T_f(\mathbf{v},\ \mathbf{v}_{ref}) = \mathbf{v}_f + \mathbf{v}_{ref}\ , \quad (2)$$

$\mathbf{v}_f$ is a visually filtered version of $\mathbf{v}$ by operator T_f, and $\mathbf{v}_{ref}$ is zero in intra-frame coding with no reference. Given that $\mathbf{x}$ represents a given image and $\mathbf{X}$ is its DWT, here are examples of coding operations in Fig.1.

The (base) left-view image is denoted by $\mathbf{x}_L$, its visually filtered coefficient matrix $\mathbf{X}_{L_F}$ in intra-frame mode without predictive coding (or $\mathbf{v}_{ref} = 0$), which delivers the perceptually lossless compression using the perceptual distortion measure *PDM* at the minimum bitrate, is given by (1) as follows,

[1] Component color video sequence for each view is preferably in 4:4:4, or at least in 4:2:2, format since chorma subsampling to 4:2:0 or 4:1:1 is known to induce noticeable chroma distortions [29].

$$\mathbf{X}_{L_F} = T_{VLF}\left(\mathbf{v}, \mathbf{v}_{ref}\right)\Big|_{\substack{\mathbf{v}=\mathbf{X}_L \\ \mathbf{v}_{ref}=0}}$$
$$= arg \min_{\forall \mathbf{X}_{L_f}} R_{BIT}\left\{\mathbf{X}_{L_f} \mid PDM\left(\mathbf{X}_L,\ \mathbf{v}_{O_f}\right)\Big|_{\mathbf{v}_{O_f}=\mathbf{X}_{L_f}} \le JNND\right\}, \tag{3}$$

noting that $\mathbf{v} = \mathbf{v}_O = \mathbf{X}_L$ in (1), and in (2),

$$\mathbf{v}_{O_f} = T_f(\mathbf{v},\ \mathbf{v}_{ref})\Big|_{\mathbf{v}=\mathbf{X}_L} = \mathbf{X}_{L_f} + 0\,. \tag{4}$$

EBCOT encoded $\mathbf{X}_{L_F}$ is output as the perceptually lossless coded left-view bitstream.

For the intra-coding of the (enhancement) right-view where the left-view $\mathbf{X}_{L_F}$ is used as a reference for inter-view disparity prediction, the minimum bitrate R_{Rmin} to encode the right-view frame, $\mathbf{X}_R$, to perceptually lossless level is given by

$$R_{Rmin} = \min\left(R_{BIT}(\mathbf{X}_{R_F}), R_{BIT}(\Delta\mathbf{X}_{R_F})\right) \tag{5}$$

where $\mathbf{X}_{R_F}$ is coded as shown in (3) by substituting $\mathbf{X}_R$ for $\mathbf{X}_L$, and

$$\Delta\mathbf{X}_{R_F} = T_{VLF}\left(\mathbf{v}, \mathbf{v}_{ref}\right)\Big|_{\substack{\mathbf{v}=\Delta\mathbf{X}_R \triangleq \mathbf{X}_R - \mathbf{X}_{L_F} \\ \mathbf{v}_{ref}=\mathbf{X}_{L_F}}}$$
$$= \arg\min_{\forall \Delta\mathbf{X}_{R_f}} R_{BIT}\left\{\Delta\mathbf{X}_{R_f} \,\Big|\, PDM\left(\mathbf{X}_R, \mathbf{v}_{O_f}\right)\Big|_{\mathbf{v}_{O_f}=\mathbf{X}_{L_F}+\Delta\mathbf{X}_{R_f}} \le JNND\right\}, \tag{6}$$

noting that $\mathbf{v}_O = \mathbf{X}_R$ in (1).

For inter-coded right-view frame with hybrid motion- and disparity-compensated prediction, the minimum bitrate R_{PRmin} to encode the right-view frame, $\mathbf{X}_R$, to perceptually lossless level is given by

$$R_{PRmin} = \min\left(R_{BIT}(\mathbf{X}_{R_F}), R_{BIT}(\Delta\mathbf{X}_{RR_F}), R_{BIT}(\Delta\mathbf{X}_{RL_F})\right) \tag{7}$$

where $\mathbf{X}_{R_F}$ is coded by substituting $\mathbf{X}_R$ for $\mathbf{X}_L$ in (3),

$$\Delta\mathbf{X}_{RL_F} = T_{VLF}\left(\mathbf{v}, \mathbf{v}_{ref}\right)\Big|_{\substack{\mathbf{v}=\Delta\mathbf{X}_{RL} \triangleq \mathbf{X}_R - \mathbf{X}_{L_F} \\ \mathbf{v}_{ref}=\mathbf{X}_{L_F}}}$$
$$= \arg\min_{\forall \Delta\mathbf{X}_{RL_f}} R_{BIT}\left\{\Delta\mathbf{X}_{RL_f} \,\Big|\, PDM\left(\mathbf{X}_R, \mathbf{v}_{O_f}\right)\Big|_{\mathbf{v}_{O_f}=\Delta\mathbf{X}_{RL_f}+\mathbf{X}_{L_F}} \le JNND\right\}, \tag{8}$$

and

$$\Delta\mathbf{X}_{RR_F} = T_{VLF}\left(\mathbf{v}, \mathbf{v}_{ref}\right)\Big|_{\substack{\mathbf{v}=\Delta\mathbf{X}_{RR} \triangleq \mathbf{X}_R[i] - \mathbf{X}_{R_F}[i-1] \\ \mathbf{v}_{ref}=\mathbf{X}_{R_F}[i-1]}}$$
$$= \arg\min_{\forall \Delta\mathbf{X}_{RL_f}} R_{BIT}\left\{\Delta\mathbf{X}_{RL_f} \,\Big|\, PDM\left(\mathbf{X}_R, \mathbf{v}_{O_f}\right)\Big|_{\mathbf{v}_{O_f}=\Delta\mathbf{X}_{RR_f}+\mathbf{X}_{R_F}[i-1]} \le JNND\right\}, \tag{9}$$

noting that $\mathbf{v}_O = \mathbf{X}_R$ and $i \in I$, (I denotes a set of sequenced video frames), is the time index of the left and right video sequences which is omitted when all frames in the equation are at the same time instance.

Not all coding paths shown in Fig.1 will be used depending on application constraints, e.g., backward compatibility requires $\Delta\mathbf{X}_{LR}$ and $\Delta\mathbf{X}_{LR_F}$ not to be used if the left-view video is used in the base view decoding for single-view video applications. Using the (base) left-view sequence coding path, this framework allows backward compatibility with perceptually lossless compression of single-view HD video for high quality digital cinematic distribution applications [23, 26]. A more generic inter-frame and inter-view prediction based coding such as biprediction can be similarly formulated and included in this framework.

Visual quality regulated 3-D video coding

Substituting *JNND* in (1) by a different JND level, i.e., JND_1, JND_2, etc. [27], perceptual 3-DVC with perceptual distortion control can be achieved [9], where JND levels are mapped by predicted values using a PDM measure or profile [19, 23, 27]. User or service specific alternative quality/distortion scales may be defined in terms of JNDs or VDUs.

Visual decomposition transform is not limited to the DWT which is used herein with the EBCOT as a vehicle to demonstrate the theoretical framework built on previously published work [9, 19, 23].

The suppression theory of human visual perception of a 3-D scene from stereoscopic video states that if right and left views are transmitted and displayed with unequal spatial, temporal and/or quality resolutions, the overall 3-D video quality is determined by the view with the better resolution [20, 25]. Therefore, rate adaptation of 3-D video may be achieved at constant perceived 3-D video quality by adaptation of the spatial, temporal resolution of one of the views with controlled perceptual distortion while encoding the base view at a designated visual quality.

4. Conclusion

In this contribution, a theoretical framework was described for visual quality regulated 3-D video coding after highlighting a number of relevant issues.

References

[1] ITU-R, "Parameter values for ultra-high definition television systems for production and international programme exchange," Rec. BT.2020, Aug. 2012.

[2] A. Vetro, et al., "3D-TV Content Storage and Transmission," IEEE Trans. Broadcasting, vol.57, no.2, pp.384-394, Jun. 2011.

[3] Y. Chen, et al., "Overview of the MVC+D 3D video coding standard," J. Vis. Commun. Image R., vol.25, pp.67-688, May 2014.

[4] M. Tanimoto, "FTV: Free-viewpoint Television," Signal Processing: Image Communication, vol.27, no.6, pp.555–570, Jul. 2012.

[5] K. Müller, et al., "3D High-Efficiency Video Coding for Multi-View Video and Depth Data," IEEE Trans. Image Process., vol.22, no.9, pp.3366-3378, Sep. 2013.

[6] G.B. Akar, et al., "Transport Methods in 3DTV-A survey," IEEE Trans. Circ. and Syst. for Video Tech., vol.17, no.11, pp.1622-1630, Nov. 2007.

[7] A. Vetro, "Frame Compatible Formats for 3D Video Distribution," Proc. ICIP 2010, pp.2405-2408, Sep. 2010.

[8] N.S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video. Prentice-Hall, 1984.

[9] H.R. Wu, et al., "Perception-based Visual Signal Compression and Transmission", (invited paper) Proc. IEEE, vol.101, no.9, pp.2025-2043, Sep. 2013.

[10] M. Yuen and H.R. Wu, "A Survey of Hybrid MC/ DPCM/DCT Video Coding Distortions", Signal Process., vol. 70, pp.247-278, Oct. 1998.

[11] Y. Gong, et al., "An efficient algorithm to eliminate temporal pumping artifact in video coding with hierarchical prediction structure," J. Vis. Commun. Image R., vol.25, no.7, pp.1528-1542, Oct. 2014.

[12] N.S. Jayant, et al., "Signal Compression Based on Models of Human Perception," Proc. IEEE, vol.81, no.10, pp.1385-1422, Oct. 1993.

[13] H.R. Wu, et al., "A Theoretical Framework for Perceptually Lossless Coding of Stereo 3-D Video", Proc. 13th IASTED Int. Conf. on Signal and Image Proc. (SIP 2011), pp.50-55, Dec. 2011.

[14] K.R. Rao and H.R. Wu, Perceptual Coding of Digital Pictures, Tutorial, ICME 2013, San Jose, California, USA, Jul. 2013.

[15] A. Vetro, et al., "Overview of the stereo and multiview video coding extensions of the H.264/AVC standard," in Proc. IEEE, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[16] K. Müller and A. Vetro, "Common Test Conditions of 3DV Core Experiments," ITU-T SG16 WP3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JCT3VE1100, San Jose, US, Jan. 2014.

[17] ITU-R, "Subjective methods for the assessment of stereoscopic 3DTV systems", Rec. BT.2021, Aug. 2012.

[18] ITU-R, Methodology for the subjective assessment of the quality of television pictures. Rec. BT.500-13, Jan. 2012.

[19] M.G. Ramos and S.S. Hemami, "Suprathreshold Wavelet Coefficient Quantization in Complex Stimuli: Psychophysical evaluation and analysis," J. Opt. Soc. Am. A, vol.18, no.10, pp.2385-2397, Oct. 2001.

[20] M.G. Perkins, "Data compression of stereo pairs," IEEE Trans. Commun., vol. 40, no. 4, pp. 684–696, Apr. 1992.

[21] D. Wu, et al., "Perceptually Lossless Medical Image Coding", IEEE Trans. Med. Imag., vol. 25, no. 3, pp 335 - 344, Mar. 2006.

[22] H. Oh, et al., "Visually Lossless Encoding for JPEG-2000," IEEE Trans. Image Process., vol.22, no.1, pp.189-201, Jan. 2013.

[23] D.M. Tan and D. Wu, Perceptually lossless and perceptually enhanced image compression system & method, Patent Int. Pub. No.: WO2013/063638 A2. WIPO, Geneva, Switzerland, 10 May 2013.

[24] L. Zhang, et al., "Stereoscopic Perceptual Video Coding Based on Just-Noticeable-Distortion Profile," IEEE Trans. Broadcast., vol.57, no.2, pp. 572-581, Jun. 2011.

[25] G. Saygih, et al., "Quality Assessment of Asymmetric Stereo Video Coding," Proc. ICIP 2010, pp.4009-4012, Sep. 2010.

[26] D. Wu, et al., "Perceptual Coding At The Threshold Level For The Digital Cinema System Specification," Proc. ICME 2010, Pascal Frossard et al (eds.), pp. 796-801, Jul. 2010.

[27] H.R. Wu, "Introduction – State of play and challenges of visual quality assessment," In: C.W. Deng, et al., (eds) Visual Signal Quality Assessment – Issues of Quality of Experience, Springer International Publishing, pp.1-30, Nov 2014.

[28] H.R. Wu, et al., "Rate-perceptual-distortion optimization (RpDO) based picture coding – Issues and Challenges," in Proc. the 19th Int. Conf. Digital Signal Process., Hong Kong, P.R. China, Aug. 2014, pp.777-782.

[29] H.R. Wu, "QoE Subjective and Objective Evaluation Methodologies," In: T. Dagiuklas and C.W. Chen (eds.), Multimedia Quality of Experience (QoE) Current Status and Future Requirements, Wiley, New Jersey, (in press) 2015.

[30] Z.Wei and K.N. Ngan, ''Spatio-temporal just notice-able distortion profile for grey scale image/video in DCT domain,'' IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 3, pp. 337–346, Mar. 2009.

[31] X. Yang, et al., "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," IEEE Trans. Circuits Syst. Video Technol. vol. 15, no. 6, pp. 742–752, Jun. 2005.

[32] D.S. Taubman, "High performance scalable image compression with EBCOT," IEEE Trans. Image Proc., vol. 9, pp. 1158-1170, Jul. 2000.

Hong Ren Wu received B.Eng. and M.Eng. degrees from University of Science and Technology, Beijing (USTB), China, in 1982 and 1985, respectively, and the Ph.D. degree from University of Wollongong, Wollongong, N.S.W., Australia, in 1990.

From 1982 to 1985, he was an Assistant Lecturer in the Department of Industrial Automation, USTB. He was on academic staff of Chisholm Institute of Technology and then Monash University, Australia from 1990 to 2005, last as an associate professor. Since February 2005, he has been with Royal Melbourne Institute of Technology (RMIT), Australia, as Professor of Visual Communications Engineering, serving concurrently to 2010 as head of computer and network engineering. He is a coeditor of *Digital Video Image Quality and Perceptual Coding* (CRC Press, 2006). He was a guest editor for the special issues on Multimedia Communication Services of CIRCUITS, SYSTEMS AND SIGNAL PROCESSING (2001), Quality Issues on Mobile Multimedia Broadcasting of IEEE TRANSACTIONS ON BROADCASTING (2008), and QoE Management in Emerging Multimedia Services of the IEEE COMMUNICATIONS MAGAZINE (2012).

Damian M. Tan received the B.Comp. (Hons.) and the Ph.D. degrees from Monash University, Australia, in 1998 and 2003, respectively. He is a research fellow with the School of Electrical and Computer Engineering, Royal Melbourne Institute Technology (RMIT), from 2006 to 2008, and a research associate of RMIT since 2009. He is director of HD^2 Technologies Pty Ltd. His primary research interests are in the areas of image and video processing.

David Wu received the B.CompSci (Hons I) degree from Monash University, Australia, in 2003 and the Ph.D. degree in Electrical and Computer Engineering from Royal Melbourne Institute Technology (RMIT) in 2007. He is director of HD^2 Technologies Pty Ltd. His primary research interests are in the areas of image and video processing.

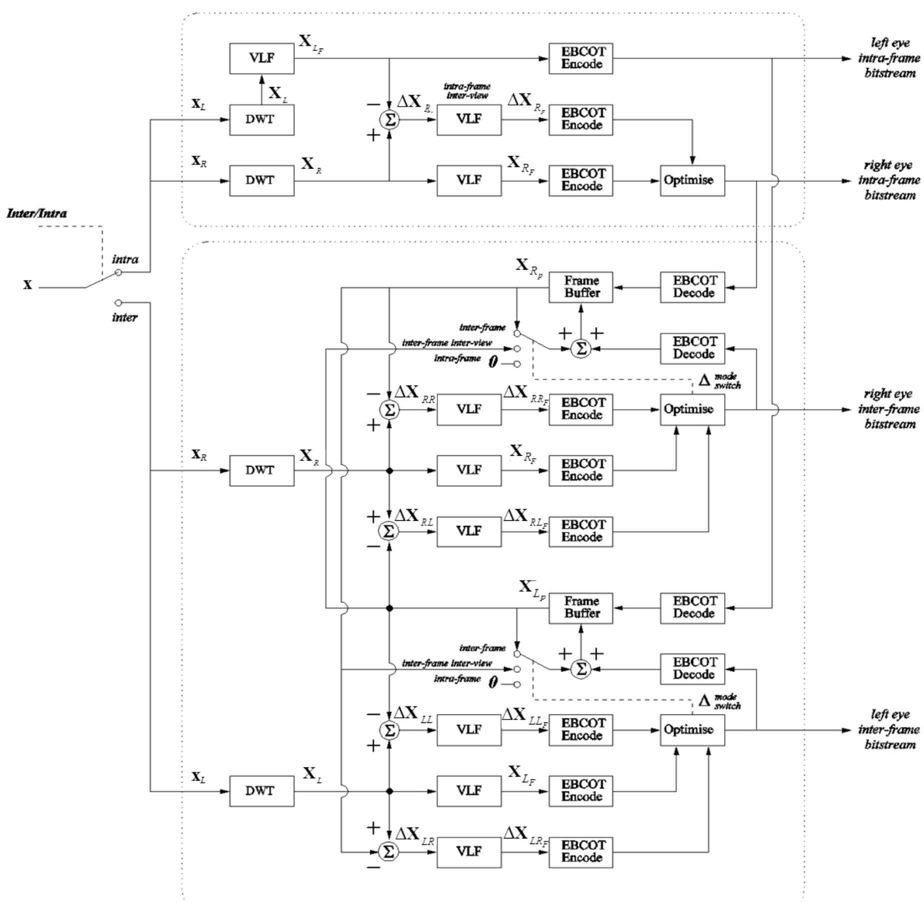


Figure 1. An example of perceptual 3-D video coding framework with hybrid inter-frame and inter-view prediction [11,31]. VLF: Visual Lossless Filtering as defined in (1); EBCOT: embedded block coding with optimal truncation [39]; Optimise: finds the minimum bitrate of the input coded streams as the output along with selected coding mode.

Effective Sampling Density and Its Applications to the Evaluation and Optimization of Free Viewpoint Video Systems

Hooman Shidanshidi, Farzad Safaei, Wanqing Li

ICT Research Institute, University of Wollongong
{*hooman, farzad, wamqing*}*@uow.edu.au*

Abstract
The concept of Effective Sampling Density (ESD) for light field based free viewpoint video (FVV) systems have been developed in last six years. It was shown that ESD is a tractable metric that can be determined from FVV system parameters and can be used to directly estimate output video quality without access to the ground truth. This paper provides an overview of the ESD and its applications in evaluation and optimization of acquisition and rendering subsystems.

1. Introduction

Free Viewpoint Video (FVV) [1, 2] consists of three main components: *acquisition* that captures the scene using a number of cameras, *rendering* that reconstructs the desired view from the acquired information, and *transmission* of captured or processed information. The performance, and in particular the quality of the output video of a FVV system depends on the efficacy of these components and their collaboration.

The *acquisition subsystem* has a variety of configurations and topologies, such as the well-known camera grids. Acquisition typically includes two processes, *ray capturing* and *depth estimation*. *Ray capturing* process results in a certain sampling density (SD) at any point of the scene space. SD at a given location is defined as the number of rays acquired per unit area of the convex hull of the surface of the scene in that location. *Depth estimation* process provides an estimation of depth (e.g. depth map) to assist rendering.

The *rendering subsystem* utilizes the captured rays and depth information from *acquisition* to synthesize the desired view, i.e., light field signal reconstruction. The view synthesize process requires to estimate a subset of unknown rays creating the desired view. To estimate an unknown ray r, the rendering goes through two processes: (i) the *ray selection* that chooses a subset of acquired rays, purported to be in the vicinity of r; and (ii) the *interpolation* that provides an estimate of r from these selected rays.

The *ray selection* process, in particular, is often prone to error. For example, imperfect knowledge of depth may cause this process to miss some neighbouring rays and choose others that are indeed sub-optimal (with respect to proximity to r). It is shown in [3] that the output of *ray selection* is an *effective sampling density* (ESD) which is lower than the original SD obtained from *acquisition*. ESD is defined as the number of rays per unit area of the scene that have been captured by *acquisition* component and chosen by *ray selection* process to be employed in the rendering. Clearly, ESD $\leq$ SD with equality holding only when the rendering process has perfect knowledge of depth and sufficient computational resources. Not surprisingly, ESD can be a true indicator of output quality, *not* SD, and its key advantage is that it provides a tractable way for evaluating the influence of the imperfections of *both* acquisition and rendering components. The theory of ESD was first introduced in [4] followed by its application in evaluation and optimization of FVV acquisition and rendering subsystems in [5, 6]. A comprehensive description of ESD and a framework for analytical derivation of ESD for different rendering methods can be found in [3,7]. It is also shown that how theoretical calculated ESD can be used to empirically predict the output video quality in term of objective signal distortion in PSNR as well as high correlation between ESD and perceived quality. Other applications of ESD include calculation of the minimum number of cameras for a regular camera grid [7], non-uniform light field acquisition based on the scene complexity variations [7], and optimization of acquisition and rendering Subsystems [7].

In this paper the concept of ESD is reviewed in section II. In section III, several FVV research problems are discussed and it is shown that how ESD has been used or can be extended to address these problems. Section IV concludes the paper.

2. Effective Sampling Density

Fig. 1 demonstrates a generic LF rendering method. Ray r is the unknown ray that needs to be estimated for an arbitrary viewpoint reconstruction. r intercepts the scene on point p at depth d. However the estimated depth of p is in error Δd which makes the rendering algorithm to assume p' as the intersection instead of p. *Ray selection* process has chosen a number of rays from all rays intersecting imaginary point p' the approximation of p. These rays are all passing through vicinity of p, but due to depth estimation error Δd, intersect with the scene in different points bounded by a convex hall A, i.e., the interpolation area. These rays are captured by a subset of cameras in camera plane uv. These cameras are bounded by a convex hull A'. It is easy to show that interpolation convex hull A is proportional to A'. Note that the sampling density has been effectively reduced because of depth estimation error Δd producing larger A.

Finally a 2D interpolation over convex hull A' is applied to estimate r. Note that if the selected rays are not passing through known pixels in image plane st, a neighbourhood estimation or a bilinear pixel interpolation is also applied to estimate them. The rendering method with depth information by using 2D interpolation over camera plane and neighbourhood estimation in image plane is called UV-DM. If pixel interpolation over st is used instead of neighbourhood estimation, it is called UVST-DM. The details of these methods and ESD derivations for them are discussed in [3]. In this review, the concept is only demonstrated for UV-DM which can be generalized to UVST-DM.

By using geometrical analysis to calculate the area of A, the ESD for the UV-DM demonstrated in Fig. 1 can be derived as:

$$\mathrm{ESD}_{\mathrm{UVDM}} = \frac{|\omega|}{A} = \frac{|\omega|}{\frac{\Delta d}{d}A' + \mu(l(d+\Delta d), A')} \quad (1)$$

where $|\omega|$ is the number of rays employed for interpolation and μ is a function to calculate the effect of pixel interpolation over st plane on the area A.

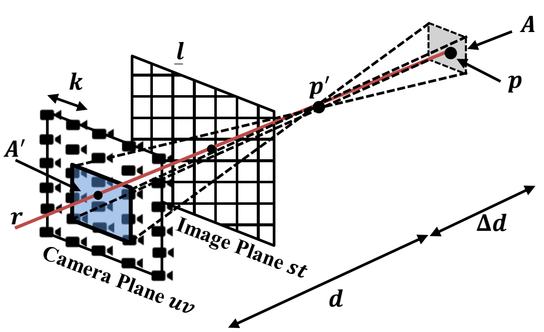


Fig. 1. Light field rendering method using depth information

In a simple form of UV-DM, the rays are selected by *ray selection* process in a way that A' becomes rectangular, i.e., 2D grid selection and therefore 2D interpolation over A' can be converted into a familiar bilinear interpolation. As shown in [3], this generic rendering method can be represented as $\mathrm{UVDM}(d, \Delta d, k, l, |\omega|)$ and its ESD can be calculated from (1) as follows:

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}(\sqrt{|\omega|}-1)\right)^2} \quad (2)$$

where k is the distance between two neighbouring cameras in the grid, l is the length of the pixel, d is the estimated depth of point p, the intersection of unknown ray r with the scene, Δd is the bound on depth estimation error, and $|\omega|$ refers to the number of rays employed in interpolation.

While (1) is derived by assuming a regular camera grid for acquisition and square area of interpolation, but ESD equations can be derived for any other scenarios.

One of the main problems in any FVV system analysis and design is acquisition and rendering evaluation and comparison. For any given acquisition configuration and rendering method, the ESD can be calculated from (1) or extensions of (1). To evaluate an acquisition component or a rendering method, it was shown in [3] that the configuration or method with higher ESD has a better output video quality. Hence, ESD can be used as an unbiased tractable indicator to directly compare acquisition configurations and rendering methods.

Another important problem is acquisition and rendering optimization. To optimize the parameters of an acquisition system, e.g., camera density for a regular camera grid or the parameters of a rendering method, e.g., number of rays for interpolation, the optimization problem can be derived using the concept of ESD and solved numerically or analytically.

Another related problem is output video prediction and estimation from system parameters without the need for implementation and experiments. In [3] it was shown that there is a high correlation between ESD and output video quality both in term of objective signal distortion in PSNR and subjective quality perceived by users. In addition, an empirical method was proposed to map calculated ESD directly to rendering quality in PSNR. This allows predicting output video quality directly from FVV system parameters.

The mathematical framework to calculate ESD for a given FVV system and to solve problems of evaluation and optimization are fully addressed in [3, 7]. In following section a summary of some of these problems are given to show the applications of ESD.

3. Applications of ESD

Calculating the minimum number of cameras

Regular camera grids are widely used for FVV acquisition. Several studies are reported to calculate the minimum number of cameras for regular grids which can be categorized in three main approaches: a) plenoptic signal spectral analysis [8,9] and, the light field spectral and frequency analysis [10, 11], b) view interpolation geometric analysis such as [12], and c) optical analysis of light field [13]. However, these methods are essentially based on several simplifying assumptions (e.g., Lambertian scene, no occlusion, linear interpolation over 4 or 16 rays, and calculating the Nyquist sampling rate without considering under-sampling), and also suggest an impractically high number of cameras [5, 7].

In contrast, using ESD to address this problem has several advantages such as studying under-sampled light field under realistic conditions (non-Lambertian reflections and occlusions) and rendering with complex interpolations. The optimization method based on ESD proposed in [5, 7] is summarized here.

In $\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ expression given as (2), d is given by scene geometry and Δd is determined by the depth estimation method and cannot be altered by us. Changing the other three parameters could potentially improve the rendering quality. By assuming a given camera resolution, i.e., a fixed value of l, two other parameters can be tuned to compensate for the depth estimation error while maintaining the rendering quality. These parameters include k as a measure of density of cameras during acquisition and $|\omega|$ as an indicator of complexity of rendering method. ESD is proportional to $|\omega|$ and inversely proportional to k, i.e., higher camera density (smaller k) and employing more rays for interpolation results in higher ESD. The optimization of k is summarized here and optimization of $|\omega|$ will be discussed in next subsection.

The problem of calculating the minimum number of cameras can be expressed in term of minimum camera density, i.e., maximum k to provide required ESD in each point of the scene to compensate for the adverse effect of depth map estimation errors. This minimum required ESD can be calculated for the ideal case when there is no error in depth estimation and there are n rays employed for interpolation. Hence the optimization method can be written as:

Find the maximum k to satisfy

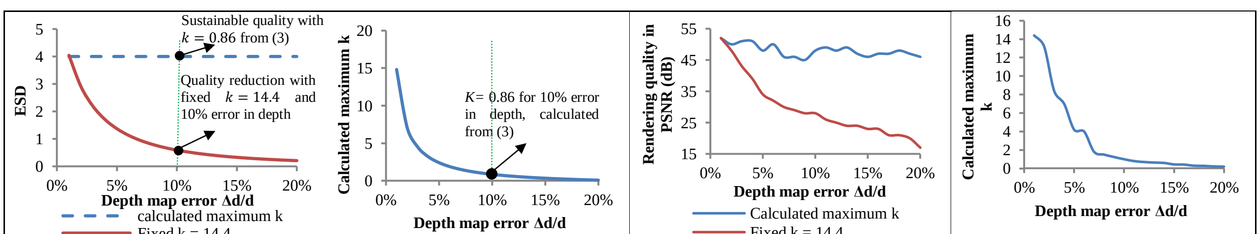
$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \mathrm{ESD}_{\mathrm{Ideal}} \rightarrow$$

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \mathrm{ESD}_{\mathrm{UVDM}(d,0,k,l,n)} \rightarrow$$

$$k = \frac{ld(d\sqrt{\frac{|\omega|}{n}} - d - \Delta d)}{\Delta d(\sqrt{|\omega|}-1)} = \frac{l\left(\left(\sqrt{\frac{|\omega|}{n}}-1\right)d^2 - d\Delta d\right)}{\Delta d(\sqrt{|\omega|}-1)} \quad (3)$$

where $\Delta d > 0$ and $|\omega| > n(\frac{d+\Delta d}{d})^2$

Fig. 2 shows the summary of theoretical expectations and experimental results for the optimization process. Fig. 2(a) and Fig. 2(b) illustrate the theoretical expectations. It is assumed that $l = 0.01$, average depth of scene $\bar{d} = 100$, relative depth map error $\frac{\Delta d}{d}$ between 1% to 20% and $|\omega|$ is calculated as follows to satisfy the condition of (3): $|\omega| > 4(\frac{100+20}{100})^2 > 5.76 \rightarrow |\omega| = 6$. For any given depth estimation error $\Delta d \leq 20\%$, k is calculated directly from (3) to maintain $\overline{\mathrm{ESD}}$ at 4.00, the ideal ESD calculated for $n = 4$ and $\Delta d = 0$. Fig. 2(a) demonstrates the ESD for fixed $k = 14.4$ and optimum k calculated from (3). Fig. 2(b) shows the calculated k in such a scenario. The corresponding point for 10% error in depth estimation is highlighted in Fig. 2(a) and Fig. 2(b), respectively, to show the relation of these two Figures. Fig. 2(c) shows that the rendering PSNR is maintained at a prescribed value (for instance 50 dB) with calculated k in contrast with the average PSNR for fixed $k = 14.4$, the required k to maintain the quality is demonstrated in Fig. 2(d). Fig. 2 shows that for high error rates, changing k using (3) results in significant improvements over the fixed camera density and can maintain the quality around the prescribed 50 dB.



a) Theoretical impact of depth estimation error on rendering quality ($\overline{\mathrm{ESD}}$) for k = 14.4 and k from (3)

b) Theoretical calculation of k from (3) to maintain the quality ($\overline{\mathrm{ESD}}$) at a constant value

c) Experimental rendering quality for fixed k = 14.4 vs. calculated maximum k demonstrated in Fig. 2(d)

d) Maximum k from to maintain the mean PSNR at a prescribed value of 50 dB

Fig. 2. Summary of theoretical and experimental optimization of k (camera density) based on ESD

Calculating the minimum interpolation complexity

The number of rays selected by *ray selection* process of a given rendering method is an important parameter of the rendering complexity. On one hand, increasing the number of rays results in increasing ESD in each point of the scene resulting in higher output quality. On the other hand, this also increases the interpolation complexity resulting in slower rendering which might not be acceptable in real time applications. To calculate the optimum number of rays for interpolation to satisfy both required rendering quality as well as rendering efficiency, an optimization method is proposed in [7]. With the same approach as previous subsection the minimum $|\omega|$ to avoid quality deterioration due to errors in depth maps can be calculated as:

Find the minimum $|\omega|$ to satisfy

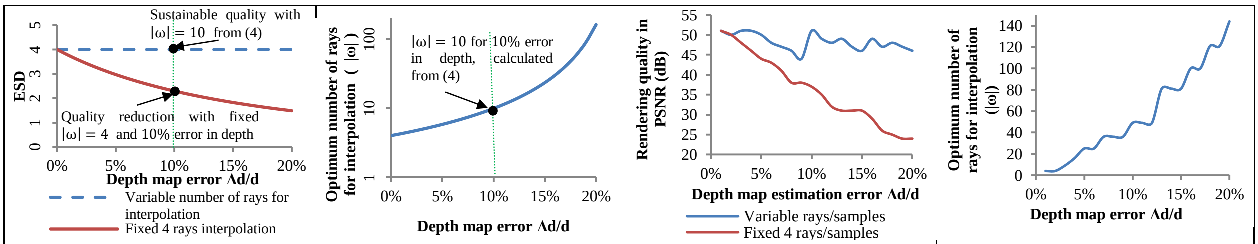
$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \mathrm{ESD}_{\mathrm{Ideal}} \rightarrow$$

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \mathrm{ESD}_{\mathrm{UVDM}(d,0,k,l,n)} \rightarrow$$

$$|\omega| = \left(\frac{l(d+\Delta d) - \frac{\Delta d.k}{d}}{\frac{ld}{\sqrt{n}} - \frac{\Delta d.k}{d}}\right)^2 \quad (4)$$

where $k < \frac{ld^2}{\Delta d\sqrt{n}}$

Fig. 3 shows the summary of theoretical expectations and experimental results for the optimization process.
Fig. 3(a) and Fig. 3(b) show the theoretical expectations for this optimization model. l, $\bar{d}$ and $\overline{\Delta d}$ are the same as Fig. 2. k is calculated as follows to satisfy the condition of (5): $k < \frac{0.01x100^2}{20\sqrt{4}} < 2.5 \rightarrow k = 2.2$. For any $\Delta d < 20\%$, $|\omega|$ is calculated directly from (4) to maintain $\overline{\text{ESD}}$ at 4.00, the ideal ESD calculated for $n = 4$. Fig. 3(a) demonstrates the ESD for fixed 4 ray interpolation and for optimum number of rays calculated from (4). Fig. 3(b) shows the actual number of rays $|\omega|$, employed in interpolation in such a scenario. The corresponding point for 10% error in depth estimation is highlighted in Fig. 3(a) and Fig. 3(b), respectively, to show the relation of these two Figures. Fig. 3(c) shows that the rendering PSNR is maintained at a prescribed value (for instance 50 dB) with calculated optimum number of rays $|\omega|$ in contrast with the average PSNR for conventional fixed 4 ray interpolation, calculated number of rays $|\omega|$ is demonstrated in Fig. 3(d). Fig. 3 shows that for high level of error in depth, the use of optimum $|\omega|$ using (4) results in significant improvements over the conventional fixed 4 ray interpolation and can maintain the rendering quality around the prescribed 50 dB.



a) Theoretical impact of depth estimation error on quality ($\overline{\text{ESD}}$) for $|\omega| = 4$ and $|\omega|$ from (4)

b) Theoretical calculation of $|\omega|$ form (4) to maintain the quality ($\overline{\text{ESD}}$) at a constant value

c) Experimental rendering quality for fixed 4 ray interpolation ($|\omega| = 4$) vs. calculated optimum $|\omega|$

d) Optimum $|\omega|$ from (4) to maintain the mean PSNR at a prescribed value of 50 dB

Fig. 3. Summary of theoretical and experimental optimization of $|\omega|$ (number of rays employed in interpolation)

Irregular acquisition based on the scene complexity

As noted before, FVV acquisition is typically performed by using a regular camera grid. While a regular acquisition itself results in non-uniform sampling density, this non-uniformity does not match the scene complexity and frequency variations. The simplest non-uniform acquisition can be done by using an irregular camera grid. The problem is then to find the positions and orientations of the camera in the gird to provide higher ESD in the parts of the scene with higher complexity and vice versa. The theory of irregular/non-uniform signal sampling has been widely investigated and it is shown that irregular sampling can reduce the number of required samples for perfect reconstruction of the signal. However to the best of our knowledge, this property has not been explored for FVV acquisition and rendering. An optimization method based on ESD for this problem is proposed in [6]. It is shown that ESD can be regarded as a set of utility functions $U_h(ESD)$ based on the given scene complexity factor h. The higher the scene complexity, more ESD would be required for a given reconstruction fidelity. Each acquisition configuration and rendering method results in an ESD pattern, which varies in the scene space. Assume that the scene could be partitioned into a number of smaller 3D regions or blocks, each having a fixed average complexity h, determined from the highest frequency components of the block computed by applying DCT transform. Then, the aim of the optimization problem could be to find the optimum acquisition configuration which provides the minimum required ESD for all blocks. This optimization problem is discussed in [6,7] and is shown that an analytical dynamic programming solution is available to compute the optimum irregular camera grid. Theoretical analysis and experimental validation showed that the output video quality can be significantly improved (around 20% in mean PSNR) by employing the proposed irregular acquisition compared with the regular camera grid. Fig. 4 shows the initial regular camera grid and the optimum irregular camera grid for 169 cameras. The average of rendering PSNR from 1,000 virtual cameras was improved from 39.10 for regular grid to 46.60 for optimum irregular grid.

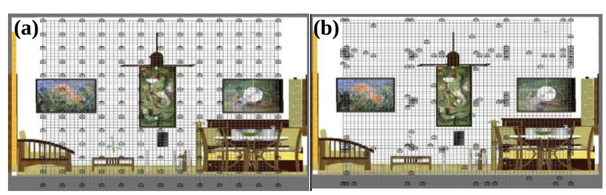


Fig. 4. a) Regular camera grid with 169 ($13x13$) cameras; b) optimum irregular camera grid for 169 cameras

4. Conclusion

ESD is an effective way to estimate signal distortion for light-field based FVV. It offers a new approach to addressing and studying various problems in FVV systems including component and system optimization.

References

[1] M. Tanimoto, *et al.*, "Free-Viewpoint TV," *IEEE Signal Processing Magazine*, vol. 28, pp. 67-76, 2011.

[2] M. Tanimoto, "FTV: Free-viewpoint Television," *Signal Processing: Image Communication*, vol. 27, pp. 555-570, 2012.

[3] H. Shidanshidi, *et al.*, "Estimation of Signal Distortion using Effective Sampling Density for Light Field based Free Viewpoint Video," *IEEE Transactions on Multimedia*, 2015 (to appear)

[4] H. Shidanshidi, *et al.*, "Objective evaluation of light field rendering methods using effective sampling density," in *MMSP*, 2011, pp. 1-6.

[5] H. Shidanshidi, *et al.*, "A Method for Calculating the Minimum Number of Cameras in a Light Field Based Free Viewpoint Video System," in *ICME*, 2013, pp. 1-6.

[6] H. Shidanshidi, *et al.*, "Non-uniform Sampling of Plenoptic Signal based on the Scene Complexity Variations for a Free Viewpoint Video System," in *ICIP*, 2013, pp. 3147 - 3151

[7] H. Shidanshidi,"Effective Sampling Density for Quality Assessment and Optimization of Light Field Rendering and Acquisition", PhD Thesis, Univesity of Wollongong

[8] J.X. Chai, *et al.*,"Plenoptic sampling," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 307-318, Jul. 2000.

[9] M. N. Do, *et al.*, "On the bandwidth of the plenoptic function," *IEEE Transactions on Image Processing*, vol. 21, pp. 708-717, 2012.

[10] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 1038-1050, 2003.

[11] C. Zhang and T. Chen, "Light field sampling," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, pp. 1-102, 2006.

[12] L. Zhouchen and S. Heung-Yeung, "A Geometric Analysis of Light Field Rendering," *Int. J. Comput. Vision*, vol. 58, pp. 121-138, 2004.

[13] T. Feng and H. Y. Shum, "An optical analysis of light field rendering," in *Proceedings of Fifth Asian Conference on Computer Vision*, 2000, pp. 394-399.

INDUSTRIAL COLUMN: BIG MOBLE DATA AND MOBILE CROWD SENSING

Guest Editors: Jun Guo and Kan Zheng

Beijing University of Posts & Telecommunications, China,
{guojun, zkan }@bupt.edu.cn

The big mobile data provides unprecedented opportunities to understand behaviors and requirements of mobile users, which in turn allow the delivery of intelligence for real-time decision making in various real-world applications. Analyzing this data can support different applications smartly, which vary from personal and community healthcare, urban sensing, marketing industry and social networking. This special issue of E-Letter focuses on the promising current progresses in big mobile data and mobile crowd sensing.

In the first article titled, "Analyzing Social Events in Real-Time using Big Mobile Data", *Gavin McArdle1, Giusy Di Lorenzo1, Fabio Pinelli1, Francesco Calabrese1, Erik Van Lierde* from IBM Research-Ireland and Maynooth University, Ireland discuss a new application called Social Event Analytics (SEA) which provides both real-time and historic information about crowd and vehicle densities. The application can be used by authorities and event organizers to manage events and gauge their success. Also, they analyze the related data to determine the amount of time individuals spent at events and to understand mobility occurring in the region.

In the second article, "Connecting human mobility with browsing behavior in mobile Interne", *Yuanyuan Qiao, Jie Yang, and Xiaoxing Zhao* from Beijing University of Posts & Telecommunications, China propose a scoring framework to quantify the relationship between human mobility and user's browsing behavior. They discusse the mobility features that impact on browsing behavior and demonstrated their points through the analysis results.

The third article titled "Multimedia Big Mobile Data Analytics for Emergency Management" by *Yimin Yang and Shu-Ching Chen* from Florida International University in USA propose propose a multimedia big mobile data computing framework for emergency management, which consists of three stages, namely data collection, data processing, and data representation. They discuss the proposed framework in details and present the system evaluation results as well.

Finally, the fourth article, titled "Smart and Interactive Mobile Healthcare Assisted by Big Data", by *Yin Zhang and Min Chen* from Huazhong University of Science and Technology, China present their recent work of social networking and cyber-physical systems especially for mobile Healthcare. The enhanced intelligence supported by mobile big data will be applied in their future work.

These articles provide different viewpoints for big mobile data and mobile crowd sensing, from data analysis to the potential applications. It is believed that big mobile data will totally change our life. We are very grateful to all the authors for making great contribution and the E-Letter Board for giving this opportunity to this special issue.

Jun Guo (guojun@bupt.edu.cn) is a full professor and a vice president of Beijing University of Posts and Telecommunications (BUPT). He received B.E. and M.E. degrees from BUPT, China in 1982 and 1985, respectively, Ph.D. degree from the Tohuku-Gakuin University, Japan in 1993. His research interests include pattern recognition theory and application, information retrieval, content based information security, and bioinformatics.
He has published over 200 papers on the journals and conferences including SCIENCE, Nature Scientific Reports, IEEE Trans. on PAMI, Pattern Recognition, AAAI, CVPR, ICCV, SIGIR, etc. His book "Network management" was awarded by the government of Beijing city as a finest textbook for higher education in 2004.
He has served on numerous TPCs for networking and theoretical computer science conferences, e.g., the first and second IEEE International Conference on Network Infrastructure and Digital Content, respectively.

KAN ZHENG [SM'09] (zkan@bupt.edu.cn) is currently a professor in Beijing University of Posts &Telecommunications (BUPT), China. He received the B.S., M.S. and Ph.D degree from BUPT, China, in 1996, 2000 and 2005, respectively. He has rich industry experiences on the standardization of the new emerging technologies. He is the author of more than 200 journal articles and conference papers in the field of resource optimization in wireless networks, M2M networks, VANET and so on. He holds editorial board positions for several journals. He has organized several special issues in famous journals including IEEE Communications Surveys & Tutorials, Transactions on Emerging Telecommunications Technologies (ETT). Dr. Zheng is a Senior IEEE member.

Analyzing Social Events in Real-Time using Big Mobile Data

Gavin McArdle[1, 2], Giusy Di Lorenzo[1], Fabio Pinelli[1], Francesco Calabrese[1], Erik Van Lierde[3]

[1]IBM Research-Ireland, Dublin, Ireland
{fcalabre, giusydil, fabiopin}@ie.ibm.com

[2]National Centre for Geocomputation, Maynooth University, Maynooth, Ireland
gavin.mcardle@nuim.ie

[3]Mobistar, Brussels, Belgium
erik.vanlierde@mail.mobistar.be

1. Introduction

Managing public safety at large events is important. Crowd control and traffic management are particularly relevant for non-ticketed events in public spaces. In such cases, it can be difficult for organizers to anticipate the number of people who will attend and to validate an event's success [1]. Given the ubiquitous nature of mobile phones, Call Detail Records (CDRs) and IP Detail Records (IPDRs), which are the logs of user transactions with a mobile phone service provider, have been widely used to study urban processes [2, 3, 4]. By building on the work of [5, 6] our research explores the potential of analyzing big mobile data in real-time to estimate the density of crowds in different areas of a city - while events are taking place. The research has also been extended to estimate the density of vehicles on the main access routes to a city.

This paper describes a new application called Social Event Analytics (SEA) which provides both real-time and historic information about crowd and vehicle densities. The application can be used by authorities and event organizers to manage events and gauge their success. The application was used in January 2015 for monitoring city wide events in Mons, Belgium which marked the launch of *Mons 2015* (Mons as a European Capital of Culture). Using SEA local police simultaneously monitored the density of vehicles on the road network and the crowd density in different areas of the city. Additionally, offline analysis of the CDR data provided the organizers with valuable information about visitors to the opening ceremony, such as their country or province of origin, duration of their trip to Mons and the city in which they stayed.

In this paper, we briefly describe the data analysis which we carried out for the specific Mons case study in which over 20 million CDRs and IPDRs were analyzed each day. The results are useful for authorities but will also help to further our knowledge and understanding of human processes in urban environments and demonstrate the benefits of using big mobile data.

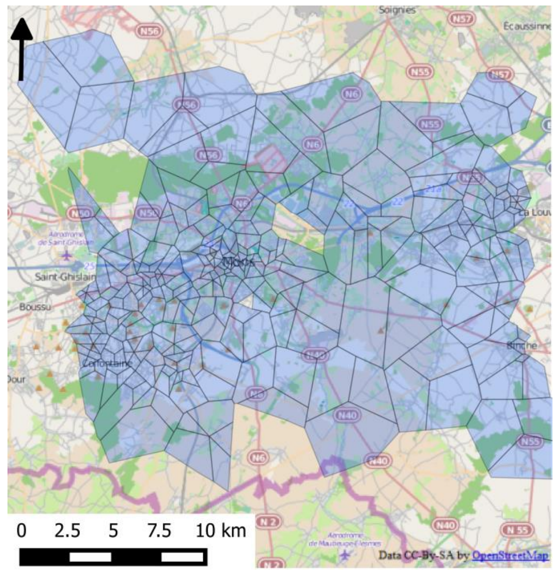


Figure 1. 319 Voronoi cells were computed from the cell tower distribution and azimuth data provided by Mobistar.

2. Data Preparation and Analysis

Data from Mobistar, a Belgium telecommunication operator, was used in this research. Mobistar provided details about the distribution and azimuth of cell towers in the city of Mons and the surrounding area. This allowed us to apply a Voronoi tessellation [7] to generate 319 distinct Voronoi polygons (figure 1). These convex polygons formed the basic unit on which crowd and vehicle density were analyzed and estimated. Prior to the opening ceremony, the police and organizers provided an inventory of specific public squares in the city where events were taking place along with a list of important access routes to the city. Our real-time analysis focused on providing population estimations and densities for these spatial features.

Call Detail Records.

Anonymous CDR and IPDR data for individuals connected to cell towers in Mons and the surrounding area were supplied directly by Mobistar. Each row of CDR data consists of an anonymous device ID, a time-stamp, the ID of the cell tower the device is connected to, the home country of the device and the type of record (call, SMS or data). Typically, data connections are an always on service which generate comprehensive CDRs and IPDRs to provide a good estimate of individual mobility. In our study the data were recorded at a rate of approximately 250 records per second and analyzed by our application in near real-time.

Population Density Estimation.
By analyzing the cell IDs, it was possible to determine the location of individuals, at the level of a Voronoi polygon, within the city at any given moment. Using aggregation, the number of users connected to each cell tower was calculated. By combining this data with the known area of each Voronoi cell and the known customer penetration rate of the mobile operator, the density of crowds in each Voronoi polygon was estimated.

A further estimation of the number of people in the whole city was calculated by summing the number of people connected to the cell towers which cover the spatial extent of the city.

Visitors to the city.
While the police are interested in public safety and crowd control, organizers are also interested in assessing the success of the event and the marketing campaigns used to attract people. In addition to the numbers attending the event, the organizers also wanted to know where people were travelling from in order to attend the events in the city. Therefore, the number of users by nationality was calculated using the home country of the device. This information is available directly from each CDR and IPDR.

In order to obtain the home city of Belgian users, further analysis of the complete CDR data for Belgium was carried out offline. Several weeks of historic CDR data were analyzed in order to determine the city where each device generally connected to the network between Midnight and 6 AM. This is similar to the approaches used to calculate the significant places people visit [8]. The results of this were used to provide detailed estimates of where visitors to Mons originated.

Vehicle Density.
In addition to crowd estimation, the density of vehicles travelling on the main access roads to the city was estimated using a minimal computational approach. Each road of interest was segmented so each segment was contained within a single Voronoi cell. This produced 86 road segments (figure 2). While the density of crowds in these cells can be estimated using the techniques described above, we are interested in those travelling through the cell. Historic CDR data was analyzed to determine users who regularly spend time in that cell. These typically represent people who live or work in the area. These users were removed from the analysis of the traffic density which allowed us to estimate the vehicle density of the major roads in the cell. When multiple road segments are contained within a single Voronoi polygon, we estimated the density of such roads by averaging the densities of the road segments in the polygons on each side of the one containing multiple major roads.

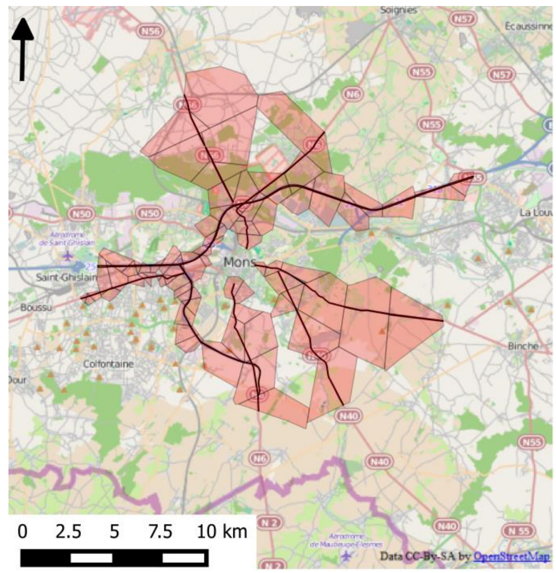


Figure 2. 86 road segments were computed based on their intersection and overlap with the Voronoi cells.

3. SEA Real-Time Application

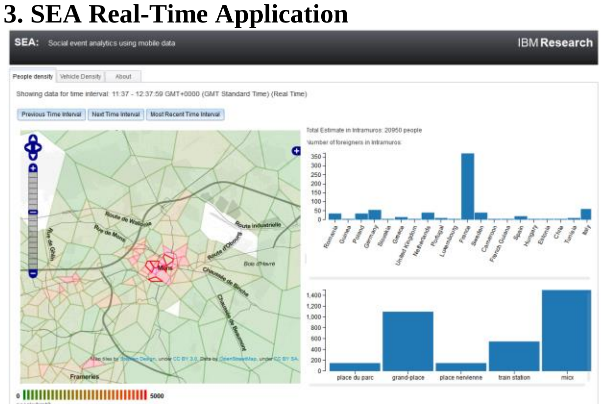


Figure 3. A screen capture from the SEA application showing the estimated crowd density for the Voronoi cells in Mons. Additional bar charts show the estimated volume of people at several locations in the city and the home country of visitors.

The CDR analysis approaches described above were incorporated into an application, called SEA, which monitors vehicle and crowd density in near real-time and provides the output in a digestible format via a graphical user interface. The data was supplied by the telecommunication operator every 15 minutes and contained the anonymous CDR and IPDR data for the previous 15 minutes. Based on the analysis described above, dynamic visualizations were produced to present the results to the police and organizers.

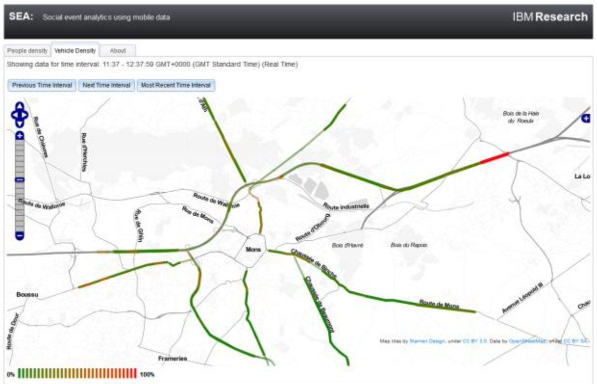


Figure 4. A screen capture from the SEA application showing the estimated traffic density on several access routes for Mons.

The main visualization, seen in figure 3, consists of a map with the Voronoi polygon structure superimposed. The polygons are colored according to the crowd density. The polygons containing the squares where events are taking place are further highlighted. A bar chart shows the estimated volume of people for each of these squares. An additional bar chart shows the home country of international visitors along with an estimate of the total number of people in the city. A separate interactive map shown in figure 4 presents the roads of interest. Each road segment is colored according to the estimated traffic density. Roads which share cells are colored with 50% transparency to signify an average of the surrounding road segments was used to calculate its density.

4. SEA Offline Analysis

In addition to the interactive real-time application, a summary report was produced after the opening event. The report provided an hourly estimate of attendance at the opening ceremony based on the number of individuals detected within the Voronoi polygons contained within the city boundary. A summary of the number of Belgian visitors to the event per home province was also calculated using the home province detection algorithm described above. Similarly, an outline of the number of international visitors by home country was provided.

By analyzing the Voronoi polygons, where mobile devices connected to the network during nighttime hours, the cities and towns where visitors to Mons stayed were determined and reported. The distribution of the duration (in days) of visits to the region was also calculated by analyzing the number of consecutive days which a mobile device was seen on the network in Mons during, prior to or subsequent to the opening ceremony. In all cases devices seen frequently in the Mons area (indicating a strong possibility of living or working in this area) were removed prior to analysis.

The report provided new insights and important feedback to the event organizers regarding the value of the event to Mons and the surrounding region. Traditional survey methods such as questionnaires would not provide the same level of detail as that offered by big mobile data analysis.

5. Conclusion

Big mobile data is a powerful tool for understanding how individuals and crowds interact in urban spaces. When the data are delivered in real-time, new opportunities to further our knowledge of the processes occurring in a city at any given moment are created.

In this paper we have described an application to monitor big crowds and events in cities in real-time by analyzing big mobile data collected from a telecommunications operator. The application was used by police and organizers to monitor social events occurring during the opening ceremony of *Mons 2015*. In conjunction with the real-time data analysis, we also analyzed the CDR and IPDR data to determine the amount of time individuals spent at events and to understand mobility occurring in the region. These can be used as indicators regarding the success of events while also furthering our understanding of urban processes. We will advance this research by examining more detailed movement patterns within the city and developing new techniques to produce crowd density estimates with a higher spatial resolution than a Voronoi polygon representing a single cell tower.

Acknowledgments

This publication has emanated from research conducted in part with the financial support of Science Foundation Ireland under Grant "SFI 13/IF/I2783".

References

[1] F. Calabrese, F. C. Pereira, G. Di Lorenzo, L. Liu, and C. Ratti, “The geography of taste: analyzing cell-phone mobility and social events” in Pervasive computing. Springer, 2010, pp. 22–37.

[2] F. Calabrese, L. Ferrari, and V. D. Blondel, “Urban sensing using mobile phone network data: A survey of research” ACM Computing Surveys (CSUR), vol. 47, no. 2, p. 25, 2014.

[3] J. Reades, F. Calabrese, and C. Ratti, "Eigenplaces: analysing cities using the space-time structure of the mobile phone network" Environment and Planning B: Planning and Design, vol. 36, no. 5, pp. 824–836, 2009.

[4] R. C'aceres, J. Rowland, C. Small, and S. Urbanek, "Exploring the use of urban greenspace through cellular network activity" in Proc. of 2nd Workshop on Pervasive Urban Applications (PURBA), 2012.

[5] T. Sohn, A. Varshavsky, A. LaMarca, M. Y. Chen, T. Choudhury, I. Smith, S. Consolvo, J. Hightower, W. G. Griswold, and E. De Lara, "Mobility detection using everyday gsm traces" in UbiComp 2006: Ubiquitous Computing. Springer, 2006.

[6] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz, "Redrawing the map of great britain from a network of human interactions" PloS one, vol. 5, no. 12, 2010.

[7] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," Nature, vol. 453, no. 7196, pp. 779–782, 2008.

[8] S. Isaacman, R. Becker, R. C´aceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky, "Identifying important places in peoples' lives from cellular network data" in Pervasive computing. Springer, 2011, pp. 133–151.

[9] F. Calabrese, Z. Smoreda, V. D. Blondel, and C. Ratti, "Interplay between telecommunications and face-to-face interactions: A study using mobile phone data" PloS one, vol. 6, no. 7, p. e20814, 2011.

Gavin McArdle is a Research Fellow at Maynooth University in Ireland, he is also a Science Foundation Ireland funded Industry Fellow at IBM Smarter Cities Technology Centre in Dublin. His research interests include human mobility, urban dynamics, spatial data analysis, city dashboards and interfaces.

Francesco Calabrese is an Advisory Research Staff Member at the IBM Research - Ireland center in Dublin, Ireland. Francesco manages the Smarter Urban Dynamics group, focusing on developing analytics and tools to better understand and optimize urban dynamics.

Fabio Pinelli is a Research Scientist at the IBM Research – Ireland centre in Dublin. His fields of research include spatio-temporal data mining, urban dynamics and intelligent transportation systems.

Giusy Di Lorenzo is a Research Scientist at the IBM Research - Ireland center in Dublin, Ireland. Her research explores the analysis of urban dynamics and human mobility using data gathered from sensor networks and social media. In particular, she is interested in developing applications to improve urban living experiences of citizens.

Erik Van Lierde works as Operational Service Manager (M2M) and Mobile Network Data Scientist at Mobistar in Brussels, Belgium. Since 2009, Mobistar has hosted the France Telecom Machine-to-Machine (M2M) International Center of Competence and is as such an important player in the worldwide M2M market. The exploration of mobile network data is a logical extra service that we deliver to M2M and other customers.

A Case for Making Mobile Device Storage Accessible by an Operator

Aaron Striegel, Xueheng Hu, Lixing Song

Department of Computer Science and Engineering, University of Notre Dame, USA
{*striegel, xhu2, lsong2*}*@nd.edu*

1. Introduction

The past few years have seen a veritable explosion of wireless data consumption across a wide variety of wireless devices. The aptly dubbed *wireless data tsunami* finds its roots in growth trends predicting 1000x growth over a period of ten years [1]. With such massive demands on capacity, wireless carriers are forced to embrace an `all of the above' strategy embracing capacity gains or demand reductions wherever possible as no individual solution is likely to fully satisfy long-term needs.

In the near term, solutions such as WiFi offloading offer tremendous appeal with technologies such as ANDSF (Access Network Discovery and Selection Function) and Hotspot 2.0 poised to dramatically streamline seamless WiFi roaming [2]. Challenges emerge though with respect to how truly seamless such roaming will be and the ability to deliver consistent Quality of Experience (QoE) to end users by virtue of the fundamental nature of unlicensed spectrum.

From the cellular side, solutions such as small cells as supported by LTE-A bring the ability to augment capacity while preserving the seamless roaming and coverage afforded by cellular [3]. Whereas the cellular network can bring a more consistent QoE, challenges emerge though with respect to the expense, logistical mechanics, and newfound system complexity management. More dramatic gains can be found in various 5G efforts with significant potential viewed in the millimeter wave bands for dramatic increases in capacity. Unfortunately, significant challenges abound in such higher frequency bands with many research challenges yet to be solved [4].

While access-based solutions seek to increase the capacity to mobile users, an alternative technique is to change how and when mobile devices retrieve their information. Efforts have ranged from characterizing data and energy consumption by smartphone apps [5] to actively examining the efficiency of the data transfers themselves [6]. Broader efforts such as Information Centric Networks (ICN) and Content-Centric Networks (CCN) could be broadly viewed similarly though not necessarily particular to wireless. More recent research efforts have explored the extent to which D2D communications [7-9] might be leveraged to share cached information owing inspiration in part to prior concepts from Delay Tolerant Networks (DTN).

2. Time Shifting as a Foundational Service

We posit that the ability to time shift demand (ex. taking advantage of elasticity in the transfer time) will emerge as one of the most important mechanisms for satisfying user QoE in wireless networks. Time shifting can allow for flattening the demand curve and hence avoid catastrophic overages during peak demands which we believe is essential for user QoE.

Effective time shifting though requires two key components: (1) the availability of storage for time shifting and (2) steerable control of said storage towards network level objectives. With regards to the first component, storage remains a relatively inexpensive component for a mobile device. Our own studies have found an average of 25% to 50% of storage free on most user devices (4-8GB) [10] with user-modifiable storage able to easily add significant capacity. The second component though is considerably more difficult. Although ICN / CCN arrangements and D2D caching shift demand, such shifts may not be necessarily nor as controllable as needed to reasonably impact user QoE.

The contribution of this paper is to make the fairly radical argument operator-accessible mobile device storage should be a key feature of future mobile devices. We outline several mechanisms by which said storage could be achieved and then discuss two scenarios (Operator Push, User

Triggered Push) to demonstrate the mechanisms for such a system and to elicit further discussion from the community at large.

System Makeup

Consider one of the following two system variations for the purposes of illustration: a user-shared system variation and operator-dedicated system variation.

In the first system variation (user-shared), an underlying file system mechanism serves as an abstraction layer. To the user, the system behaves exactly as before with the user having both allocated storage (music, apps, movies, photos) and free storage. Writes and reads to the file system proceed as expected at the user level. From the perspective of the abstracted file system (less than best effort file system), the free space may be allocated as seen fit to the wireless operator. The 'free' space may be overwritten at any time by the user with a reasonable expectation that storage changes will occur gradually rather than rapidly (ex. a user may grab several hundred MBs but is unlikely to eliminate 4GB+ of storage quickly [10]). Prioritization of content for eviction would be provided via operator-defined policies.

In the second system variation (operator-dedicated), the storage is moved to be part of the network adapter(s) / chipset. No prioritization is needed for eviction as all space is exclusively accessible by the operator.

Storage Accessibility

With said storage in place, an API would be exposed for the operator to be able to write (or read if necessary) from the storage space. The API would need to be appropriately protected via appropriate security mechanisms (ex. secure connection and / or signed content). Connections would be initiated on the device side leaving open effectively a secure FTP channel equivalent for the carrier to drive content. Note that a secure device-initiated connection would also afford content steering even when the device is connected via WiFi. Support for pushing either via LTE Broadcast or multicast would also be possible.

Content would be stored locally in a block-wise manner with appropriate identifiers for access. Larger objects would likely be split into multiple blocks to accommodate cases where the device itself may not be able to cache the entirety of the larger object. An API would allow for localized trapping of content requests to the operator-managed storage as a function of the underlying operating system. Content could be sandboxed with respect to individual applications or shared across all applications.

3. Example Scenarios

Operator Push

In the first scenario, we assume that all content in the operator space must be pushed by the operator to the client. The operator could consider this content from one of two perspectives. In the first perspective, machine learning would be applied to pro-actively fetch content for a user. For instance, a user might access CNN or ESPN every morning. The connection is pre-fetched and saved in the local storage whereby the local storage is viewed as exclusively benefitting the QoE perception of the carrier (ex. carrier X's speed is acceptable). Alternative scenarios might involve LTE Broadcast and staging of larger system or application updates (ex. Angry Birds). Other alternative scenarios might involve the active involvement of on-line social networking sites effectively tagging shareable high volume content (ex. videos) for pro-active pushing as the networks allows.

From the second perspective, the storage is viewed as a potential mechanism for bringing in additional revenue. A content provider might pay a carrier for localized staging of content or even perhaps the operator might offer different levels of content distribution (localized in the geographic area, on the device, etc.). Pricing could be used to competitively prioritize cases where intense storage constraints were observed. Advertisements generated based on location might be pre-staged to avoid issues with connectivity ensuring smooth delivery (ex. payment for ad, payment for initial site content / information).

User Triggered Push

In the second scenario, we assume the presence of D2D caching whereby when mobile nodes come into range of one another, the nodes will share / exchange caches. Normally in such cases, information will propagate as mobile nodes `infect' one another with the data. In pure D2D schemes though, the only way to realize a beneficial time shift is if a mobile node comes in contact with another mobile node that already possesses the data prior to the mobile node needing the data. Effectively, the perfect arrangement of time, location, and access must occur for the full utility of D2D caching to be realized.

Although the Operator Push allows the operator to observe and potentially push popular content, the Operator Push model may miss content, particularly when content is received across multiple access mechanisms (ex. WiFi, cellular). The operator could observe checksums of similar content (via reads or simple relaying of the number of matches during a cache exchange) and use that early warning as a tripwire to push out content to specific users or even groups of users. With proper tuning, the operator could push out popular content much sooner and across broader time windows as the network allows. The key property is that by making the operator aware of such staging with the ability to control the staging, the operator can offer a better QoE and overall system optimization.

4. Conclusions

In this paper, we made the fairly radical argument that mobile device storage should be operator-accessible rather than exclusively owned and controlled by the user. Whether such storage is effectively hidden as a lower-level / less reliable storage or expressly hidden as part of the wireless chipset, the notion of a foundational network service for operator storage at the mobile device is the end goal. We believe that such storage opens up intriguing new opportunities for services whether those services are primarily in the benefit of the carrier or going even further viewed as a revenue stream for the operator. Finally, we note that there exist interesting constructs as well as applied to D2D caching and operator management of said services.

References

[1] Cisco, “Cisco VNI mobile data traffic forecast 2012-2017,” Feburary 2013.

[2] InterDigital, Cellular-WiFi-Integration, White Paper, 2011.

[3] 3GPP Release 10.

[4] Andrews, J.G.; Buzzi, S.; Wan Choi; Hanly, S.V.; Lozano, A.; Soong, A.C.K.; Zhang, J.C., "What Will 5G Be?," Selected Areas in Communications, IEEE Journal on , vol.32, no.6, pp.1065,1082, June 2014

[5] N. Ding, D. Wagner, X. Chen, A. Pathak, Y. C. Hu, and A. Rice. Characterizing and modeling the impact of wireless signal strength on smartphone battery drain. In Proc. of SIGMETRICS, pages 29-40, New York, NY, USA, 2013. ACM.

[6] F. Qian, J. Huang, J. Erman, Z. M. Mao, S. Sen, and O. Spatscheck. How to reduce smartphone traffic volume by 30%?. In Proceedings of the 14th international conference on Passive and Active Measurement (PAM'13), Matthew Roughan and Rocky Chang (Eds.). 42-52, 2013.

[7] Al. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger. Is there a case for mobile phone content pre-staging?. In Proceedings of the ninth ACM conference on Emerging networking experiments and technologies (CoNEXT '13). ACM, New York, NY, USA, 321-326, 2013.

[8] A. Asadi, Q.Wang, and V. Mancuso, “A Survey on Device-to-Device Communication in Cellular Networks,” IEEE Communications Surveys & Tutorials, vol. PP, no. 99, pp. 1–1, 2014.

[9] M. Ji, G. Caire, and A. F. Molisch, “Wireless Device-to-Device Caching Networks: Basic Principles and System Performance,” Computing Research Repository, vol. abs/1305.5216, 2013.

[10] X. Hu, L. Meng, A. Striegel, Evaluating the Raw Potential for Device-to-Device Caching via Co-Location, in Proc. Of MobiSPC, Aug. 2014.

Aaron Striegel received his BS and Ph.D degrees from Iowa State University in Computer Engineering in 1998 and 2002, respectively. He is currently an associate professor and associate chair of Computer Science Engineering at the University of Notre Dame which he joined in 2003. He also serves on the Executive Committee of the Wireless Institute at Notre Dame. Prof. Striegel's research interests include large scale network instrumentation, wireless system dynamics, and content distribution across wireless and wired networks.

Xueheng Hu is a graduate student of Prof. Striegel in the Department of Computer Science and Engineering at the University of Notre Dame. He received his B. Eng. In Software Engineering from Beijing University in 2006 and his MS in Computer Science from Miami University of Ohio in 2011.

Lixing Song is a graduate student of Prof. Striegel in the Department of Computer Science and Engineering at the University of Notre Dame. He received his BS in Electrical Engineering from Wuhan University in 2011 and his MS in Computer Science from Ball State in 2014.

The role of keypoint detection and description in human action recognition from videos

Sio-Long Lo and Ah Chung Tsoi

Macau University of Science and Technology, Taipa, Macau SAR, China
{sllo, actsoi}@must.edu.mo

1. Introduction

Human action recognition from completely un-marked video clips is an important problem in mobile big data, with the ever increasing number of videos being made available on repository web sites such as Youtube, or from television broadcasts. The problem of human action recognition, if solved, would facilitate quick and fast indexing and retrieval of these ever increasing numbers of video clips.

There are in general two approaches to solving this problem, one using explicit feature extraction techniques to extract features for each frame in the video clip, and then use some classification techniques like support vector machines, or its nonlinear counterpart, kernel machines, to classify the concatenated features. This is commonly referred to as bag of visual word (BoVW), bag of words (BoW) or bag of features (BoF) approach (see e.g., [1]). The other approach, the deep neural network approach is to use an integrated implicit feature extraction method consisting of a multiple hidden layered feedforward neural network with specific architectures, e.g., convolutional neural networks, and then another fully connected multiple hidden layered feedforward neural network is used to perform the classification task (see e.g., [2]). It was commented in [3] that these two approaches are closely related.

As it is impractical to extract features from all pixels in each frame of the video in the BoF approach, it is customary to extract features only at some selected points, the so-called keypoints [4, 5, 6, 7, 8], which are points in which the features become significant. For the deep neural network approach [2], e.g., the two streamed deep neural network (TSDNN) approach [3], the appearance cues and the motion cues pass through individual and largely identical deep neural network pipelines, and then the results are combined at the end to produce a classification [3]. No explicit keypoint detection and extraction are performed.

In this paper we will consider the BoF approach, using three feature extraction techniques: (1) scale invariant feature transform (SIFT) [5], (2) speedup robust feature (SURF) approach [9], and (3) Oriented Rotated BRIEF (ORB) approach [10], though we will make some comments on the (TSDNN) approach [3] when it is relevant to illustrate the relationships between the two. Our aim is to compare the performance of these three keypoint and feature extraction techniques and apply them to the same benchmark dataset: UCF Sports, the baseline result will be the one obtained using the same feature extraction technique using normal appearance and motion cues and then to draw some conclusions on the suitability of the technique to large scale mobile big data processing.

There are a number of possible ways to detect and extract keypoints and features. The latest one is KAZE features [11] and AGAST [8], based on extension of FAST [6, 7]. The reason why we have chosen to use those three: SIFT [5], SURF [9] and ORB [10] is that their keypoint detection techniques are quite different, and that the features extracted are different and these features are used quite often in image processing. KAZE features [11] and its accelerated version [12] has not been applied to video processing task yet, to the best of our knowledge, while AGAST [8] appears to be an extension of the FAST (feature accelerated segment test) technique [6, 7] of determining the keypoints, which is the underlying keypoint detection technique used by ORB [10]. A future research task is to combine AGAST [8], and accelerated KAZE [12] and use the ideas contained in this paper to apply it to both appearance cues and motion cues for video processing.

2. Keypoint detection

A common method in keypoint detection and description would be to consider points in each frame, in which some features, e.g., pixel intensities, describing the frame undergo some significant changes. For example the set of keypoints can be determined using a 2D Harris corner condition [4]. Harris corner condition [4] is to consider the eigenvalues of the 2D matrix of the directional gradients of pixel intensities in the vicinity of a candidate point A with coordinates x,y. If these eigenvalues fall within a certain bound then the point A is considered significant in which some changes of features, in this case, the gradient of pixel intensities, occurred. It is well known that Harris corner condition gives many false key points [5], and hence the signal to noise ratio is quite low.

Another method would be to consider the following observation (see e.g., [5]): a keypoint would be invariant to scale changes in the image [5]. In a certain sense, this [5] can be considered a generalization of the Harris corner condition, in that the gradient condition is carried out in different scales and the point which is invariant to scale changes will most likely be the point in which significant changes in the features occurred. This is the method undergirding the determination of the so called scale invariant feature transform (SIFT) [5].

In the SURF approach [9], the keypoints are detected using the Hessian matrix of the pixel intensities, a second order description, instead of the gradient matrix, a first order description, of the pixel intensities. The Hessian matrix is the gradient of the gradient matrix, and hence would quantify the curvature of the variation of pixel intensities.

The ORB (Oriented and Rotated BRIEF) uses the FAST (features accelerated segment test) method [6, 7] in locating the keypoints. The FAST method [6, 7] essentially draws a Bresenham circle [13] with radius 3 centered on the candidate pixel and considered the gradient of the points located on this circle with the pixel intensity of the candidate point A. If the sum of the gradients exceeds a certain threshold, then it concludes that the candidate point A is a keypoint. There exists a preselection algorithm in which only four points on the circle, the north, south, east and west, are evaluated [6, 7]. There also exists a way in which two candidate keypoints close to one another can be reduced to one keypoint [6, 7]. Using this algorithm it is possible to find all keypoints in an image. This way of detecting the keypoints though superficially looks quite different from those used in Harris corner condition [4], or the one used in SURF [9], can be considered as utilizing the third order moment of the gradient of the pixel intensities, using a circular neighborhood instead of the common square neighborhood used in [4, 9].

3. Features related to appearance cues

The set of keypoints once obtained can be used as the basis of obtaining features describing the appearance cues, conveyed by the pixel intensities, in grey level or in colour. For the keypoints detected using Harris corner condition [4] we can compute the histogram of the gradients of points in a square neighborhood around the keypoint [14]. In this case, consider the neighboring points to A in the n×n grid. The gradients at these points are evaluated, and a histogram can be obtained. This is called a histogram of gradients (HOG) [14], a feature vector characterizing the local behavior of point A. If the histogram is normalized then this can be considered as a measure of the probability of the gradient distribution around the point A. If the histogram has Nh bins then this is an HOG with length Nh HOG.

In the SIFT case [5], we can compute the HOG of the invariant point A on multiscales though in this case, it uses the Difference of Gaussian and Laplacian of Gaussian for the actual computations [5]. This can be concatenated together to form a feature vector characterizing the point A. For historical reasons, this is called a SIFT vector, or simply SIFT [5].

In SURF [9], a boxed version of the Gaussian functions is used. The evaluation of the SURF features while different from that of the SIFT; in many ways it is an adaptation of the SIFT approach [5] making use of some of the alternative fast computational techniques, e.g., integral image, to compute the features [9].

The ORB (Oriented FAST and Rotated BRIEF – binary robust independent elementary features [15]) essentially recognizes that in FAST [6, 7], unlike in SIFT [5] or SURF [9], there is no rotational invariance included, and hence the major contribution of the ORB is to include the rotational invariance into the formulation [10]. The BRIEF is used to obtain the binarized version of the features, so that comparison between two vectors (as performed in feature matching) can be performed very rapidly using logical operations on digital devices, e.g., mobile phones. Note that there is no multiscale version of ORB and this underlies the reason why we conducted the experiments in the way which we did as shown in Section 8.

The outcome of this is that there will be feature vector which will characterize the appearance cues, i.e., based on an image instead of based on videos.

In the TSDNN [3] a deep neural network pipeline is dedicated to compute the feature vectors using appearance cues. It does not however, has the multiscaled version as in SIFT [5], or SURF [9], or the explicit dedicated architecture, exemplified by using a Gaussian mask as in SURF [9] or FAST [6, 7].

4. Motion cues

Motion cues will form a significant part in understanding how an object moves in the video. A simple way in which motion cues can be used would be to consider what is commonly referred to as optical flow approach. A vector connecting a point A (x,y) in a frame, with the same point A $(x+\Delta x, y+\Delta y)$ in the succeeding frame is called an optical flow vector, or simply an optical flow [16]. The optical flow encapsulates some information on motion. The velocity of point A moving between the current frame and the succeeding frame is given by $(\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t})$ where Δt is the frame rate. The histogram formed by considering the points in the neighborhood of point A is called a histogram of oriented optical flow (HOF).

Since SIFT[5]can be considered as a generalization of the gradient of the pixel intensities it makes sense to consider a SIFT flow [17], the vector which connects the SIFT at a keypoint A in one frame and the SIFT of the same point A in the succeeding frame. As SURF and ORB are in some sense the off-springs of SIFT, it makes sense to consider SURF flow and ORB flow respectively as ways to describe the motion behavior of the objects in the video. To the best of our knowledge SURF flow, or ORB flow has never been considered previously, and this is the first time such a concept is considered.

Another characteristic of motion is the acceleration of the object moving through the video. A simple characterization of such a notion would be to consider the change of velocity of point A (x,y) among 3 frames, i-th frame (x,y), i+1-th frame, ($x+\Delta x$,$y+\Delta y$) and i+2-th frame, ($x+\Delta x+\Delta\xi$,$y+\Delta y+\Delta\zeta$), i.e., $\left(\frac{\Delta\xi-\Delta x}{\Delta t}, \frac{\Delta\zeta-\Delta y}{\Delta t}\right)$, where Δt is the frame rate. For historical reasons, this is called a motion boundary [**Error! Reference source not found.**, **Error! Reference source not found.**] and histogram of the points in the neighborhood of A is called a motion boundary histogram (MBH) [14,18]. It is simple to see that one can form the acceleration descriptors of the features SIFT[5], SURF[9]and ORB [10]respectively. It is worth noting that apart from the so called motion boundary [[19], no one has ever considered the accelerations of the SIFT flow, SURF flow and ORB flow respectively in the literature.

It is possible to concatenate the motion cue feature vectors on each two frame pair, throughout the entire video, and obtain a characterization of the motion conveyed by the objects in the video clip.

In the TSDNN [3], a separate deep neural network pipeline is set up to consider the motion cues. It is claimed that such pipeline encapsulates the spirit of the HOF and MBH, though it is difficult to see from the feature formation part of the deep neural network.

5. Global saliency description using trajectories

By connecting the movements of point A (x,y) in frame i through the optical flow vector ($x+\Delta x$,$y+\Delta y$) in frame i+1, for i=1,2,…,N where N is the total number of frames in the video clip, it is possible to form the trajectory spanned by point A in the video [20][21]. The shape of this trajectory informs us on the global behaviour of the point A. This shape descriptor can be computed simply by connecting the trajectory pieces between every two frame pair, normalized by the total length of the trajectory . To make it less dependent on the behaviour of a single keypoint, it is common to consider a group of points in the vicinity of A moving through the video together. This provides a trajectory tube of the set of points moving in the video. It is possible to work out the shape of the trajectory tube.

6. Appearance cue and motion cue features

For each point in the keypoint set, we can evaluate the set of descriptors, gradients, SIFT, SURF, and ORB, related to appearance cues, and motion cues, and the trajectory shape descriptor. Collect this set of descriptors of the keypoints over the entire length of the video, we will have a non-parametric model of the video as described by these feature vectors.

7. Classification

One may perform classification of these feature vectors if there is a training dataset, in which the labels of each video are provided. The classification can be performed using a deep neural network [2][3] in recognition of its highly nonlinear discriminating surface in the high dimensional feature space, or a kernel machine [20][21] Conceptually this would be a relatively simple task. For recognition of a video clip without any labels, one simply extract the features, gradients, SIFT, SURF, ORB, and then pass such set of features through the classifier, which will provide the classification of the video with unknown label.

There is a question which might be raised: given that there are a number of methods for extracting the appearance and motion cue features, e.g., gradient, SIFT, SURF, and ORB, which one might give the best performance on a given video benchmark. If we can settle this question, then we can consider how to make such a method run efficiently on mobile devices, which might have limited memory and processing capabilities. The results of such processing, with its reduced volume, can then be collected centrally so as to form some quasi-real time assessment of the situation. This is exactly what we wish to do in this paper, to provide some preliminary results of running such sets of feature extraction methods over a benchmark video set.

8. Experimental results

We compare our proposed techniques on the UCF-Sports benchmark dataset [22]. The UCF-Sports dataset contains 150 videos from ten action classes, diving, golf swinging, kicking, lifting, horse riding, walking, running, skating, swinging (on the pommel horse and on the floor), and swinging (at the high bar). These videos are taken from real sports broadcasts and the bounding boxes around the subjects are provided for each frame. We follow the protocol proposed in [22] using the same training/testing split of the dataset. In our experiment, we extract the motion feature only on one scale. This is because the ORB features are only extracted using one scale. To make the comparison fairer, we decided to evaluate the results using only one scale in all the feature extraction approaches. The experimental results are shown in Table I.

TABLE I. ACTION RECOGNITION USING DIFFERENT FEATURE EXTRACTION TECHNIQUES

	SIFT	SURF	ORB	Optical
Traj. Shape	57.4%	62.1%	67.8%	66.8%
HOG	67.7%	66.8%	65.5%	78.3%
HOF	71.9%	68.9%	74.0%	78.3%
MBH	72.8%	73.2%	71.5%	77.4%
Combined	78.3%	78.7%	76.0%	78.3%

The experimental results show that SURF gives the best performance, followed by SIFT, the baseline and ORB. It is not surprising to find that SURF performs best as intuitively we expect its performance will be similar to those provided by SIFT, except that SURF uses a second order measure to determine the keypoints. It is not surprising to see that SIFT and the baseline results are the same. The slightly degraded performance of the ORB relative to others might be due to binarized version of the features in BRIEF [15] as the keypoint detection used [6][7] should be superior to that used in the baseline or SURF [9].

9. Conclusion

In this paper, we compared the performance of a number of feature extraction methods, viz., SIFT, SURF and ORB to be used in BoF approach for human action recognition; the baseline results are those obtained from using the gradient information, and the simple processing of the motion cues. We evaluated these approaches on the UCF-Sport dataset, and found that SURF gives better recognition results than the others using a relatively short histogram vector of 64. In future, we wish to explore the aspect of multiscale feature extraction using SIFT, SURF and ORB, and a longer histogram length, as well as deploying the techniques to other more challenging datasets, e.g., Hollywood2 [23], UCF101 [24].

Acknowledgment The authors wish to thank the Fundo para o Desenvolvimento das Ciências e da Tecnologia, Macau SAR, for financial support in the form of a project grant number: 034/2011/A2, which made this research possible.

References

[1] B. Schiele and J. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *Int J Computer Vision*, vol. 36, no. 1, 2000.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Neural Info Proc Syst*, 2012, p. 1106-1114.

[3] K. Simonyan and A. Zissermann, "Two-stream convolutional networks for action recognition in videos," *Arxiv 1406.2199*, 2014.

[4] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc Alvey Vision Conf*, 1988, p. 147–151.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2, p. 91-110, 2004.

[6] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *IEEE Int Conf Computer Vision*, vol. 2, 2005, pp. 1508–1511.

[7] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal Machine Intel*, 2009.

[8] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *Euro Conf Computer Vision*, 2010, pp. 183–196.

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Euro Conf Computer Vision*, 2006.

[10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Int. Conf. Computer Vision*, 2011, p. 2564-2571.

[11] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Euro Conf. Computer Vision (ECCV)*, 2012, pp. 214–227.

[12] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Brit Mach Vision Conf*, 2013.

[13] Bresenham, "Algorithm for computer control of a digital plotter," *IBM Sys J*, vol. 4, no. ! , pp. 25–30, 1965.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Int Conf Computer Vision Pattern Recog*, 2005, pp. 886–893.

[15] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 34, no. 7, pp. 1281–1298, 2011.

[16] B. K. P. Horn and B. G. Schunk, "Determining optical flow," *Art Intel*, vol. 17, pp. 185–203, 1981.

[17] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Trans on Pattern An. Machine Intel*, vol. 33, no. 5, 2011.

[18] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Euro Conf Computer Vision*, vol. 3952, 2006, p. 428–441.

[19] S. L. Lo and A. C. Tsoi, "Motion boundary trajectry for human action recognition," in *Asian Conf Computer Vision, Workshops*, 2014.

[20] H. Wang, A. Kläser, C. Schmid, and C. L. Liu, "Dense trajectories and motion boundary descriptors for action recognition." *Int J Computer Vision*, 2013.

[21] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Int Conf Computer Vision*, 2013.

[22] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action mach: A spatio-temporal maximum average correlation height filter for action recognition," in *Int Conf Computer Vision Pattern Recog*, 2008, pp. 1–8.

[23] I. Laptev, M. MarszaNek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in *Int Conf Computer Vision Pattern Recog*, 2008, pp. 1–8.

[24] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human action classes from videos in the wild," Univ Cent Florida, Tech. Rep. CRCV-TR-12-01, November 2012.

Dr. S L Lo received his BSc, MSc, and PhD degrees respectively in 2006, 2010, and 2012 from Macau University of Science and Tech. After one year of postdoc, he is currently an assistant professor in the same institution.

Prof A C Tsoi received his MSc and PhD respectively in 1970 and 1972 from University of Salford. After having worked in Britain, New Zealand, Australia, and Hong Kong, he took up his current position as Dean, Faculty of Information Technology, Macau University of Science and Technology in 2010.

Multimedia Big Mobile Data Analytics for Emergency Management

Yimin Yang and Shu-Ching Chen

School of Computing and Information Sciences, Florida International University, USA
{yyang010, chens}@cs.fiu.edu

1. Introduction

The world has stepped into a big data era with the development of advanced technologies and the growth of the Internet of Things (IoT). The volume of big data is expected to reach yottabyte (10^{24}) in the near future, among which over 60% will come from wireless mobile devices as opposed to desktops by the year of 2016 [1]. Except for publicly available data released by organizations and government, more and more private individuals begin to share multimedia data through mobile devices across the world. For example, people may take pictures and videos instantly at a disaster scene and share them via social media tools on mobile devices. Accompanying the exponentially growing big data is the challenge of how to analyze and make sense of those data to provide better services to the world. A concrete example is the problem of associating a situation report with plain textual information with the multimedia data collected at a disaster scene to support the timely and efficient decision-making process [2].

With the ever increasing enormous big data, a single-pass framework is infeasible to process the data in real-time. Therefore, many researchers in both industries and academia look for solutions on large-scale data processing. MapReduce (MR) [3] is the framework of choice for large-scale distributed applications. Recent research work in the literature has shown the effectiveness of MR-based frameworks on the tasks of semantic classification [4], information retrieval [5], and so on. More recently, Hadoop Spark [6], a successor system that is more powerful and flexible than Hadoop MapReduce, is merging due to its advantages of lower latency, iterative computation and real-time processing.

In this paper, we propose a multimedia big mobile data computing framework for emergency management. The framework can be described in three stages: the first stage is data collection through vertical search engine and web services; the second stage is data processing, including textual document analysis and multimedia classification leveraging our previous work on the MapReduce Multiple Correspondence Analysis (MR-MCA)-based semantic classification framework; and the third stage is data representation through an iPad application with an intuitive and friendly interface. Section 2 discusses the proposed framework in details. Section 3 presents the system evaluation results and section 4 concludes the paper.

2. A Multimedia Big Mobile Data Framework

We present the proposed multimedia big mobile data framework for emergency management. There are several major components, including data collection, document analysis, key frame and feature extraction, MR-MCA-based classification, report-multimedia association and the final presentation through well-defined user interface as shown in Figure 1.

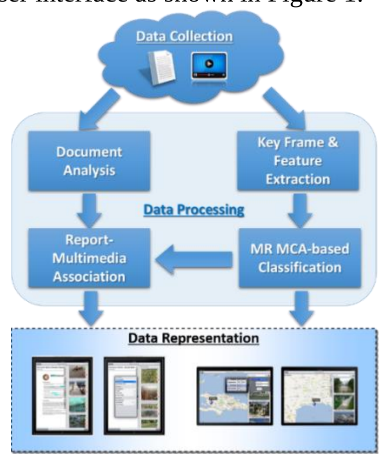


Figure 1. Multimedia big mobile data computing framework for emergency management.

Data Collection.
High-quality, real-time and relevant information is critical for effectively dealing with emergency situations. There are two ways for data collection, i.e., vertical search engine and web services. Specifically, we design and implement a vertical search engine that provides an initial solution for continuously crawling, organizing, indexing and retrieving disaster information. The built in crawler of the vertical search engine runs on a MR cluster. In addition to the data automatically crawled through the vertical search engine, we also provide web services to enable users to upload disaster related data, such as reports, images and videos.

Key Frame and Feature Extraction.
Key frame selection is a critical step for video processing before feature extraction. In this paper, we propose an effective key frame extraction method

based on camera take detection, which can dramatically reduce redundant data in videos while keeping the semantic information.

1)Key frame extraction based on camera take detection
A camera take is a series of consecutive frames taken by a camera. It can be cut into a sequence of segments and interleaved with other camera takes to form a scene which completes an event or a story in a video program. This is a common process in film editing. Figure 2 shows an example of camera take editing results for a video downloaded from FEMA website. Each sub-image in the figure is a key frame selected from a shot (as illustrated in Figure 2(a)), and thus frames (a) to (h) represent consecutive shots, composing a scene. To take a closer look at the key frames, it is obvious that frames (b), (e), and (g) are from the same camera take, so are frames (a) and (c), as well as (f) and (h). Apparently, the shots from the same camera take could be grouped together and represented by one or more frames. It will highly reduce the throughput for further processing.

Figure 2. Examples of camera takes.

Figure 3 depicts the process of camera take detection.
Specifically, it takes the following four steps for camera take detection:

- **Frame difference calculation:** based on the assumption that two consecutive frames in a video shot should have a high similarity in terms of visual content. The frame difference is calculated using color histogram (or raw pixel values for saving the computational cost) as a measurement of similarity between two frames.
- **Shot detection:** if the frame difference is above some preset threshold, then a new shot is claimed. The selection of threshold is critical since it may cause over segmentation or down segmentation depending on the types of video programs (action, drama, etc.). To determine a proper threshold and further refine the detection results, certain constraints may apply, such as the shot duration.
- **Key frame selection:** a key frame should properly represent the visual content of a shot. Without loss of generality, the last frame of a shot is selected as the key frame for later processing. It is worth mentioning that more advanced techniques may be utilized to select (or generate) the most representative key frame(s).
- **Camera take detection:** each detected shot (represented by a key frame) will be matched with the last shot in each detected camera take. If a certain matching criterion is satisfied, then the current shot will be added to the end of the matched camera take. It is based on the assumption that a shot is the most related to the one with the closest temporal relationship. Initially, within a certain time period, we may assume the first shot as a camera take. The matching strategies vary from sift point matching to frame difference matching, depending on various performance requirements.

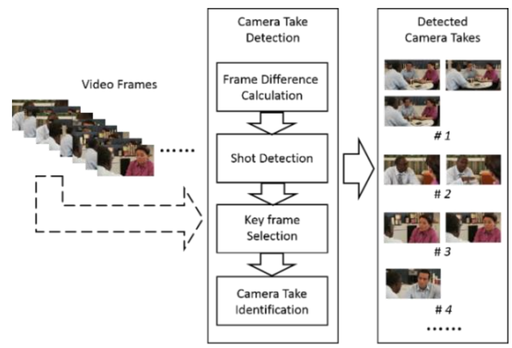


Figure 3. Camera take detection.

2)Feature extraction
Both visual and textual features are extracted for multimedia classification. Specifically, the visual features include visual descriptors, such as Histogram of Oriented Gradient (HOG) [7], color and edge directivity descriptor (CEDD) [8], as well as other low-level visual features, e.g., color histogram, color moment, and texture wavelet [9]. The textual features are extracted from the meta data such as titles and descriptions based on *tf-idf* schema. Except for the aforementioned keyframe-based features, we also extract shot-based features for videos [10]. Since there are totally over 700 dimensional features, some of the features might be redundant or even irrelevant. Therefore, a feature selection step is carried out based on the Hidden Coherent Feature Groups (HCFG) analysis method [9] to identify exemplar features for final classification.

Document Analysis.
Location and subject (e.g., hurricane, flood, and earthquake) are two critical characteristics to describe an emergency event, which are the same for the associated multimedia (such as the images and videos taken at the event scene). The purpose of document analysis is to identify the potential location-subject pairs in a document (e.g., situation report) and relate it to the classified multimedia data. There are mainly three steps for document analysis as follows.

1)Entity extraction
GATE system [11] is a popular tool for natural language processing (NLP). It is applied in our framework to extract tokens and identify certain types of entities, such as data and location, for further analysis.

2)Synonym extraction
Considering the same subject may be expressed in different words in various documents, it is necessary to search all possible synonyms for a specific subject in a document. The synonym extraction is performed using an open source package based on WordNet [12].

3)Location-subject pair identification
After retrieving a list of candidate locations through the GATE system and all synonyms for the corresponding subjects via WordNet, a matching procedure is carried out to identify the final location-subject pairs by examining all the tokens in a document. For more details of the analysis, please refer to [2].

MR-MCA-based Classification.
When dealing with large-scale big mobile data, it is not suitable to use a single-pass framework since the huge volume of data will easily use up memory, not to mention the speed of processing. To accommodate big data requirement and tackle those aforementioned issues, a distributed MCA framework [4] based on MR [3] technique is proposed to perform large-scale correlation-based semantic classification tasks. The MCA algorithm has been proved to be effective for disaster image classification [13], and it was further improved by incorporating temporal information and principle components analysis [10]. In this study, we update MR-MCA and adapt it to video classification as described in Algorithm 1.

The algorithm receives as its inputs the training and testing data sets, Tr and Te. First the MR-MCA algorithm [4] is applied to Tr to obtain the training model, denoted as $\{F_j, W_j\}_{j=1}^{J}$, where F_j and W_j represent the feature and the corresponding weight set, with J as the total number of features (line 2). Then for each video instance V_i in Te, and each key frame X_q in V_i, the algorithm iterates through the instance's feature items (line 5) and the class labels (line 6, where $|C|$ is the total number of classes), to accumulate the score S_l for each X_q (line 7). The label of X_q is determined by the highest score of S_l, represented as C_l. Finally, the video instance V_i is classified to the one with the largest number of key frames (lines 13-14).

Report-Multimedia Association.
After document analysis and MR-MCA-based classification, we are able to associate the processed situation report with the classified multimedia data based on the identified location-subject pairs from both sides. In the next section, we will introduce the developed iPad application based on the proposed framework, where the functionalities such as filtering based on locations and subjects as well as keywords are provided. The users are also allowed to give feedback to the processed results and to further improve the association.

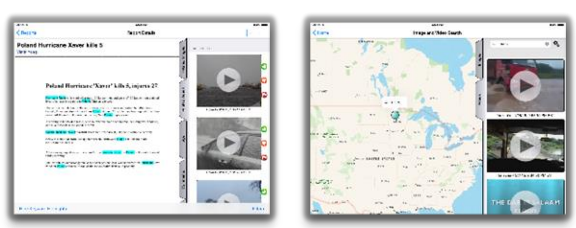
Algorithm 1 MR-MCA for video classification

Input: Training data set Tr, testing data set Te
Output: Classification results for Te

1: **procedure** CLASSIFICATION(Tr, Te)
2: $\quad \{F_j, W_j\}_{j=1}^{J} \leftarrow$ MR-MCA(Tr);
3: $\quad$ **for** all $V_i \in Te \quad (i = 1, \cdots, N)$ **do**
4: $\quad\quad$ **for** all $X_q \quad (q = 1, \cdots, Q)$ **do**
5: $\quad\quad\quad$ **for** all $F_{j,k} \in \{\{F_k\}^j\}_{j=1}^{J}$ **do**
6: $\quad\quad\quad\quad$ **for** $l \leftarrow 1, \ldots, |C|$ **do**
7: $\quad\quad\quad\quad\quad S_l \leftarrow S_l + abs(w_{j,k}^l)$;
8: $\quad\quad\quad\quad$ **end for**
9: $\quad\quad\quad$ **end for**
10: $\quad\quad\quad C_l \leftarrow \arg\max_l(\{S_l\})$;
11: $\quad\quad\quad \{X_q\}^{C_l} \leftarrow X_q$;
12: $\quad\quad$ **end for**
13: $\quad\quad C_l^* \leftarrow \arg\max_{C_l}(|\{X_q\}^{C_l}|)$;
14: $\quad\quad \{V_i\}^{C_l^*} \leftarrow V_i$;
15: $\quad$ **end for**
16: $\quad$ **return** $\{\{V_i\}^{C_l^*}\}$
17: **end procedure**

3. System Evaluation

Since the evaluation of report-image association has been conducted in our previous work [2], we will mainly evaluate the report-video association in this paper. We have crawled over 1500 videos from the Internet and processed the data using the proposed framework. Figure 4 shows two of the major interfaces in our developed iPad application, where Figure 4(a) shows the situation report with related videos that are classified based on the pre-defined ontology. The users are able to filter the videos based on locations, subjects, and keywords. In addition, users can offer feedback to the retrieved results and the system will automatically refine the association accordingly. Furthermore, we also provide the functionality of retrieving disaster videos using keywords and displaying them on a map based on locations as shown in Figure 4(b).



(a) (b)
Figure 4. iPad application interfaces.

4. Conclusion

In this paper, we have presented a multimedia big mobile data computing framework consisting of three stages, namely data collection, data processing, and data representation. In the data collection stage, vertical search engine and web services are used to collect high quality disaster data. In the data processing stage, multimedia data analytics implemented on top of the MR framework are performed to associate situation reports (in plain texts) with multimedia data (such as images and videos). Finally, an iPad application with interactive interfaces is developed.

Acknowledgement

This research is partially supported by DHS under grant Award Number 2010-ST-062-000039, DHS's VACCINE Center under Award Number 2009-ST-061-CI0001, NSF HRD-0833093 and CNS-1126619. The authors would like to thank Xiaoyu Dong for his assistance in collecting data and implementing the iPad application.

References

[1] http://venturebeat.com/2015/01/22/big-data-and-mobile-analytics-ready-to-rule-2015/.

[2] Y. Yang, W. Lu, J. Domack, T. Li, S.-C. Chen, S. Luis, and J. K. Navlakha, "MADIS: A Multimedia-Aided Disaster information Integration System for emergency management," *The 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, pp. 233-241. 2012.

[3] J. Dean and G. Sanjay, "MapReduce: Simplified Data Processing on Large Clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107-113, 2008.

[4] F. C. Fleites, H.-Y. Ha, Y. Yang, and S.-C. Chen, "Large-Scale Correlation-Based Semantic Classification Using MapReduce," *Cloud Computing and Digital Media: Fundamentals, Techniques, and Applications*, pp. 169-190, CRC Press, 2014.

[5] S. P. Dravyakar, S. B. Mane, and P. K. Sinha, "Private Content Based Multimedia Information Retrieval Using Map-Reduce," *International Journal of Computer Science Engineering and Technology (IJCSET)*, vol. 4, no. 4, pp. 125-128, 2014.

[6] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster Computing with Working Sets," *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, pp. 10-10, 2010.

[7] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886-893, 2005.

[8] S. A. Chatzichristofis and Y. S. Boutalis, "CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval," *Computer Vision Systems*, pp. 312-322, 2008.

[9] Y. Yang, H.-Y. Ha, F. C. Fleites, and S.-C. Chen, "A Multimedia Semantic Retrieval Mobile System Based on HCFGs," in *IEEE MultiMedia*, vol. 21, no. 1, pp. 36-46, 2014.

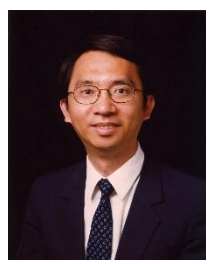
[10] Y. Yang, S.-C. Chen, and M.-L. Shyu, "Temporal Multiple Correspondence Analysis for Big Data Mining in Soccer Videos," *The First IEEE International Conference on Multimedia Big Data (BigMM)*, 2015.

[11] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "Gate: A Framework and Graphical Development Environment for Robust NLP Tools and Applications," *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 168-175, 2002.

[12] Princeton University. Wordnet, A Lexical Database for English. http://wordnet.princeton.edu/, July 2011.

[13] Y. Yang, H.-Y. Ha, F. C. Fleites, S.-C. Chen, and S. Luis, "Hierarchical Disaster Image Classification for Situation Report Enhancement," *The 12th IEEE International Conference on Information Reuse and Integration (IRI)*, pp. 181-186, 2011.

Yimin Yang is a Ph.D. candidate at the School of Computing and Information Sciences (SCIS), Florida International University (FIU), Miami. She received her M.S. degree in Computer Science from FIU in 2012. Her research interests include multimedia data mining, multimedia systems, image and video processing, and information retrieval.



Shu-Ching Chen is an Eminent Scholar Chaired Professor in Computer Science at the School of Computing and Information Sciences, Florida International University, Miami since August 2009. Prior to that, he was an Assistant/Associate Professor in SCIS at FIU from 1999. He received his Ph.D. degree in Electrical and Computer Engineering in 1998, and Master's degrees in Computer Science, Electrical Engineering, and Civil Engineering in 1992, 1995, and 1996, respectively, all from Purdue University, West Lafayette, IN, USA. His main research interests include content-based image/video retrieval, multimedia data mining, multimedia systems, and Disaster Information Management. Dr. Chen was named a 2011 recipient of the ACM Distinguished Scientist Award. He received the best paper award from 2006 IEEE International Symposium on Multimedia. He was awarded the IEEE Systems, Man, and Cybernetics (SMC) Society's Outstanding Contribution Award in 2005 and was the co-recipient of the IEEE Most Active SMC Technical Committee Award in 2006.

Smart and Interactive Mobile Healthcare Assisted by Big Data

Yin Zhang[1] *and Min Chen*[2]

[1] School of Information and Safety Engineering,
Zhongnan University of Economics and Law, China
(*yin.zhang.cn@ieee.org*)

[2] School of Computer Science and Technology,
Huazhong University of Science and Technology, China
(*minchen@ieee.org*)

1. Introduction

With the rapid growth of smart phones and wireless technology, mobile terminals and applications in the world are growing rapidly. The advanced mobile computing and communications greatly enhance the user's experience by the notion of "carrying small while enjoying large", which have brought a huge impact to all aspects of people's lifestyles in terms of work, social, and economy [1]. The proliferation of mobile devices and their enhanced onboard sensing capabilities are some of the major forces that drive the explosion of mobile sensing data. Furthermore, advances in social networking and cyber-physical systems are making mobile data "big" and increasingly challenging for storage and processing. As a whole, mobile data has unique characteristics, e.g., mobile sensing, moving flexibility, noise, and a large amount of redundancy. Recently, new research on mobile analysis has been started in different fields [2-4]. Since the research on mobile big data is just started, we only introduce some of our recent works in this paper.

2. Systems for Mobile Healthcare Assisted by Big Data

As a novel means of generating data, crowd sensing has already come to the center stage of mobile computing. In order to extract maximum values through effective analysis of crowd sensing data, typically, various techniques such as machine learning, information transmission, social networking, video streaming [5] and graph clustering methods can be utilized [6]. By analyzing crowd sensing data, individual behavior patterns can be extracted, which could be useful for guiding and improving people's daily life. In [7], Yin et al. present a novel community-centric framework named CAP for event prediction based on crowd sensing big data analysis, which is illustrated in Figure 1. With a case study using a real dataset captured over a 15-month period, the proposed scheme has been validated that mobile users' activities can be predicted through big data analysis.

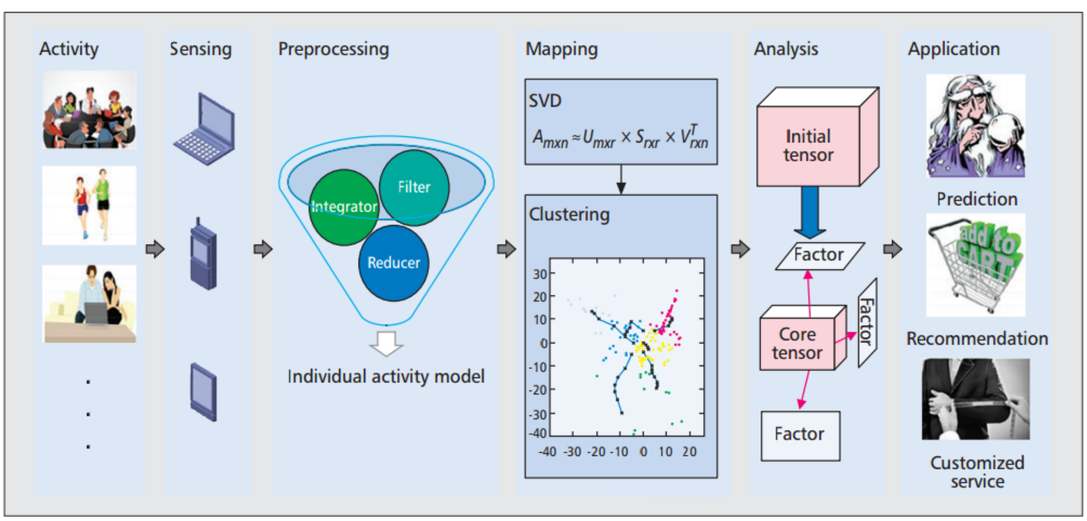


Figure 1: CAP Framework. [7]

Recently, the progress in wireless sensor, mobile communication technology, and stream processing enable people to build a body area network to have real-time monitoring of people's health. Generally, medical data from various sensors have different characteristics in terms of attributes, time and space relations [8], as well as physiological relations, etc. In addition, such datasets involve privacy and safety protection. In [9], Min et al. design a Robotics and Cloud-assisted Healthcare System (ROCHAS) to provide empty nester with situation-aware, human-centric, proactive and user-friendly healthcare services. In ROCHAS, a household low-cost robot serves a mobile terminal to collect physiological and environmental data for health and mental status analysis assisted by cloud computing and big data. The basic system architecture of ROCHAS is shown in Figure 2.

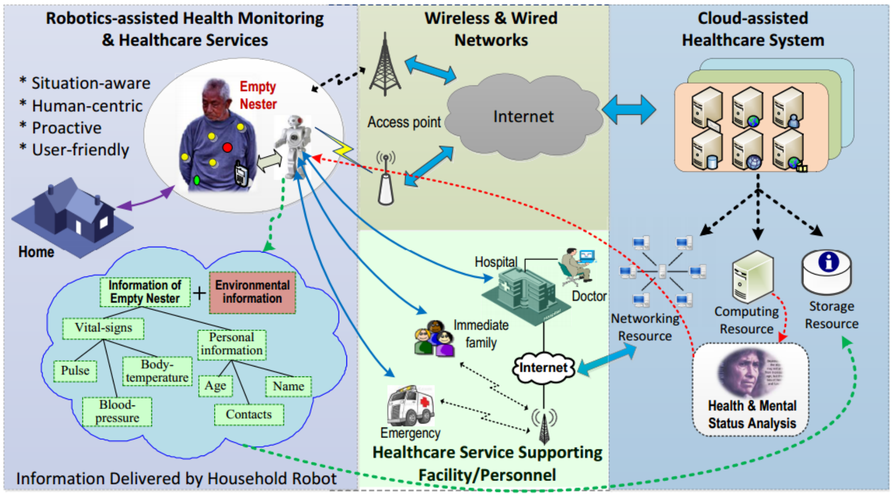


Figure 2: Architecture of ROCHAS. [9]

Furthermore, through the robot, various physiological and psychological data can be sensed and analyzed via affective computing. Assisted by the advanced techniques related to big data, emotion-aware mobile applications and services are available. For example, the Embedded and Pervasive Computing (EPIC) Laboratory at Huazhong University of Technology and Science develop an affective interaction named AIWAC to provide mobile user with emotion-aware services [10]. Figure 3 illustrates AIWAC architecture.

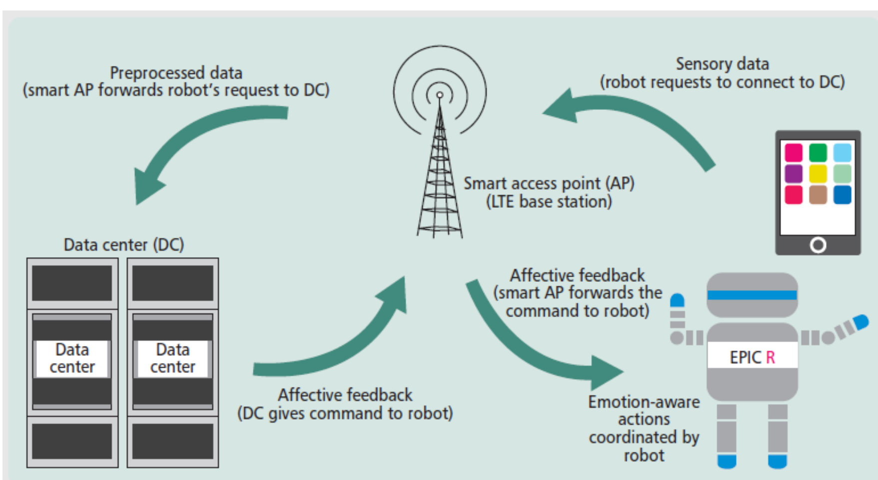


Figure 3: AIWAC Architecture. [10]

3. Conclusion

Nowadays, it is widely believed that big data greatly promotes the development of human society. In particularly, with mobile big data, anyone anywhere can be producer and participant of data other than consumer. Moreover, we can carry enhanced intelligence supported by mobile big data.

References

[1] K. Zheng, Y. Wang, W. Wang, M. Dohler, J. Wang, ``Energy-efficient wireless in-home: the need for interference-controlled femtocells'', *IEEE Wireless Communications*, vol.18, no.6, pp. 36-44, 2011.

[2] K. Zheng, F. Hu, W. Wang, W. Xiang, M. Dohler, ``Radio resource allocation in LTE-advanced cellular networks with M2M communications'', *IEEE Communications Magazine*, vol.50, no.7, pp. 184-192, 2012.

[3] Y. Li, T. Wu, P. Hui, D. Jin, and S. Chen, "Social-aware D2D communications: qualitative insights and quantitative analysis," *IEEE Communications Magazine*, Vol. 52, No. 6, pp. 150-158, Jun. 2014.

[4] C.F. Lai, M. Chen, J.S. Pan, C.H. Youn, H.C. Chao, ``A Collaborative Computing Framework of Cloud Network and WBSN Applied to Fall Detection and 3-D Motion Reconstruction'', *IEEE Journal of Biomedical and Health Informatics,* Vol. 18, No. 2, pp. 457-466, March 2014.

[5] J. He, Z. Xue, D. Wu, D. Wu, Y. Wen, ``CBM: Online Strategies on Cost-aware Buffer Management for Mobile Video Streaming'', *IEEE Transactions on Multimedia (IEEE TMM)*, vol. 16, No. 1, pp. 242 - 252, Jan. 2014.

[6] H.P. Chiang, C.F. Lai, Y.M. Huang, ``A Green Cloud-assisted Health Monitoring Service on Wireless Body Area Networks'', *Information Sciences*, Vol. 284, No. 10, pp. 118–129, Nov. 2014.

[7] Y. Zhang, M. Chen, S. Mao, L. Hu, V. Leung, ``CAP: Crowd Activity Prediction Based on Big Data Analysis'', *IEEE Network*, Vol. 28, No. 4, pp. 52-57, July 2014.

[8] Y. Li, C. Song, D. Jin, S. CHen, ''A dynamic graph optimization framework for multi-hop D2D communication underlaying cellular networks'', *IEEE Wireless Communications*, Vol. 21, No. 5, pp. 52-61, Dec. 2014.

[9] M. Chen, Y. Ma, S. Ullah, W. Cai, E. Song, "ROCHAS: Robotics and Cloud-assisted Healthcare System for Empty Nester'', *BodyNets 2013*, Boston, USA, Sep. 2013.

[10] M. Chen, Y. Zhang, Y. Li, M. Hassan, A. Alamri,"AIWAC: Affective Interaction through WearAble Computing and Cloud Technology'', *IEEE Wireless Communications Magazine*, Vol. 22, No. 1, pp. 20-27, 2015.

Yin Zhang is with the School of Information and Safety Engineering, Zhongnan University of Economics and Law. He was a post-doctoral fellow at the School of Computer Science and Technology at Huazhong University of Science and Technology (HUST) from 2013 to 2015. He has co-authored over 30 technical journal and international conference papers. He is a guest editor for New Review of Hypermedia and Multimedia.

Min Chen is a professor in School of Computer Science and Technology at Huazhong University of Science and Technology (HUST). He is Chair of IEEE Computer Society (CS) Special Technical Communities (STC) on Big Data. He was an assistant professor in School of Computer Science and Engineering at Seoul National University (SNU) from Sep. 2009 to Feb. 2012. He was R&D director at Confederal Network Inc. from 2008 to 2009. He worked as a Post-Doctoral Fellow in Department of Electrical and Computer Engineering at University of British Columbia (UBC) for three years. Before joining UBC, he was a Post-Doctoral Fellow at SNU for one and half years. He received Best Paper Award from IEEE ICC 2012, and Best Paper Runner-up Award from QShine 2008. He has more than 180 paper publications, including 85 SCI papers. He is a Guest Editor for IEEE Network, IEEE Wireless Communications Magazine, etc. He is Co-Chair of IEEE ICC 2012-Communications Theory Symposium, and Co-Chair of IEEE ICC 2013-Wireless Networks Symposium. He is General Co-Chair for the 12th IEEE International Conference on Computer and Information Technology (IEEE CIT-2012) and Mobimedia 2015. He is General Vice Chair foTridentcom 2014. His research focuses on Internet of Things, Machine to Machine Communications, Body Area Networks, Body Sensor Networks, E-healthcare, Mobile Cloud Computing, Cloud-Assisted Mobile Computing, Ubiquitous Network and Services, Mobile Agent, and Multimedia Transmission over Wireless Network, etc. He is an IEEE Senior Member since 2009.

Call for Papers

Symposium on Signal Processing in Mobile Multimedia Communication Systems

Orlando, Florida, USA, December 14-16, 2015

http://2015.ieeeglobalsip.org/symp_mobilecom.html

The IEEE Global Conference on Signal and Information Processing (GlobalSIP) is a recently launched flagship conference of the IEEE Signal Processing Society. The conference will focus broadly on signal and information processing with an emphasis on up-and-coming signal processing themes. The conference will feature world-class speakers, tutorials, exhibits, and technical sessions consisting of poster or oral presentations.

The Signal Processing in Mobile Multimedia Communication System Symposium will focus on the signal processing challenges in delivering real-time multimedia data over wireless links to from servers to mobile devices or between mobile devices. In particular, the symposium will address the signal processing research issues related to seamless delivery of real time multimedia data to mobile devices. The symposium would like to address the following questions. How to design and implement an energy-efficient multimedia coding techniques on battery-operated mobile devices? How could efficiently transmit huge amount of multimedia data between servers and mobile devices? How could best preset multimedia data to mobile devices with limited display size? How to integrate different signal processing techniques that all compete for limited resources, e.g. CPU, memory, battery, display etc., on mobile devices? How to achieve cross-layer optimal resource allocation to maximize the overall quality of service of multimedia services?

Technical paper submissions are solicited in the interest topics which may include, but are not limited to:

- Multimedia signal enhancement
- Multimedia content analysis and event detection
- Multimedia security and forensics
- Joint multimodal processing and analysis
- Multimedia indexing and retrieval
- Distributed/centralized source coding
- Scalable and low delay source coding
- Error/loss resilient source coding
- Resource constraint multimedia transmission
- Multimedia transmission over MIMO antennas
- Distributed multimedia compression
- Mobile multimedia for learning
- Media fusion for communication and presentation
- Audio/video analysis, modeling, processing and transformation
- Image analysis, modeling, and recognition
- Communication and cooperation through mobile multimedia
- Scalable multimedia big data management

Submission of Papers: Prospective authors are invited to submit full-length papers, with up to four pages for technical content including figures and possible references, and with one additional optional 5th page containing only references. Manuscripts should be original (not submitted/published anywhere else) and written in accordance with the standard IEEE double-column paper template. All paper submissions should be carried out through EDAS system (http://edas.info). A selection of best papers and best student papers will be made by the GlobalSIP 2015 best paper award committee upon recommendations from Technical Committees.

Program Committee

General Chair
Honggang Wang, University of Massachusetts, Dartmouth

Technical Chairs
Qing Yang, Montana State University, Bozeman, USA
Zheng Yuan, RealCommunications Inc., San Jose, USA
Shiwen Mao, Auburn University, Auburn, USA

Publicity Chair
Zhaohui Wang, Michigan Technological University, Houghton, USA

Important dates
Paper submission deadline: **May 25, 2015**
Review results announced: **June 30, 2015**
Camera-ready papers due: **September 5, 2015**

Call for Papers

IEEE Conference on Standards for Communications and Networking (CSCN 2015)

Tokyo, Japan, 28-30 October 2015

http://www.ieee-cscn.org

Standards play a key role in the success of the communications industry, as enablers of global systems inter-operability. IEEE CSCN aims for closing the gap between researchers, scientists and standards experts from academia, industry and different standardization bodies. It will serve as a platform for presenting and discussing standards-related topics in the areas of communications, networking and related disciplines, facilitating standards development as well as cooperation among the key players.

IEEE CSCN is inviting submission of high quality technical as well as visionary papers, which will be reviewed and selected by an international Technical Program Committee (TPC) representing both academia and industry, with a strong standardization background. Topics of interest include, but are not limited to, enhancements to existing systems and communication protocols developed by standards bodies such as ITU-T, IEEE, IETF, 3GPP, ETSI, OMA, GSMA, Broadband Forum, Metro Ethernet Forum, oneM2M, ONF, among others. Visionary papers on hot topics, such as Advanced 5G Radio Access and Network Infrastructure, Converged Access Networks, Optical Networks, Twisted Pair and Coaxial Access Networks, Software Defined Networks and Services, Network Functions Virtualization (NFV), and other works in progress being currently discussed by the standardization bodies will be included.

The conference will also solicit papers on the relationship between innovation and standardization, technology governance of standards, the history of standardization, tools and services related to any or all aspects of the standardization lifecycle, and compatibility and interoperability, including testing methodologies and certification to standards.

Accepted and presented papers will be published in the IEEE CSCN Conference Proceedings and submitted to IEEE Xplore® as well as other Abstracting and Indexing (A&I) databases. The conference's best papers will be recommended for publication at the IEEE Communications Magazine's supplement on Communications Standards.

IEEE CSCN will also include several panel sessions and keynotes focusing on the broad issue impacting standards directions in the telecommunications sector. Tutorials on topics of critical interest across multiple SDOs will be also considered.

IEEE CSCN will accommodate the following special industry sessions, allowing companies to explain their latest offering that embodies some specific new standards:

- Advances in Vehicular Communications
- Software Defined Sensors Networks and IoT: perspective and proposals for new standardization activities
- Optical Wireless Communication

IEEE CSCN 2015 will take place in Tokyo immediately preceding and located adjacent to IETF's 94th Plenary in Yokohama, Japan.

Steering Committee:
Tarik Taleb, Aalto University, Finland (Chair)
Chih-Lin I, China Mobile, China
Bruno Chatras, Orange, France
Bernard Barani, European Commission, Belgium
Henrik Abramowitz, Ericsson, Sweden
Terje Tjelta, Telenor, Norway
Hermann Brand, ETSI, France
Diego Lopez-Garcia, Telefonica, Spain
David Soldani, Huawei European Research Centre, Germany
Robert S. Fish, NETovations Group LLC, ComSoc VP of Standards Activities, USA
Luis M. Correia, IST – University of Lisbon, Portugal
Alexander D. Gelman, IEEE ComSoc and NETovations, USA

Honorary General Chairs:
Hiroshi Esaki, University of Tokyo, Japan
Takehiro Nakamura, NTT DOCOMO, Japan
Kei Sakaguchi, Osaka University, Japan

General Chair:
Tarik Taleb, Aalto University, Finland

TPC Chairs:
Adlen Ksentini, INRIA Rennes / University of Rennes 1, France
Michiaki Hayashi, KDDI Labs, Japan
Athul Prasad, Nokia, Finland
Anass Benjebbour, NTT DOCOMO, Japan
JaeSeung Song, Sejong University, Korea
Tuncer Baykas, Istanbul Medipol University, Turkey

Important Dates
Paper Submissions: **June 15, 2015**
Notifications: **August 15, 2015**
Camera-ready: **September 5, 2015**

Call for Papers

IEEE International Conference on Cloud Computing Technology & Science (CLOUDCOM)

Vancouver, BC, Canada, November 30-December 3, 2015

http://2015.cloudcom.org

CloudCom is the premier conference on Cloud Computing worldwide, attracting researchers, developers, users, students and practitioners from the fields of big data, systems architecture, services research, virtualization, security and privacy, high performance computing, always with an emphasis on how to build cloud computing platforms with real impact. The conference is co-sponsored by the Institute of Electrical and Electronics Engineers (IEEE), is steered by the Cloud Computing Association, and draws on the excellence of its world-class Program Committee and its participants. The conference proceedings are published by IEEE CS Press (IEEE Xplore) and indexed by EI and ISSN.

The conference this year solicits research articles in various areas including, but not limited to:

Architecture
* Intercloud architecture models
* Cloud federation & hybrid cloud infrastructure
* Cloud services delivery models, campus integration & "last mile" issues
* Networking technologies
* Programming models & systems/tools
* Cloud system design with FPGAs, GPUs, APUs
* Storage & file systems
* Scalability & performance
* Resource provisioning, monitoring, management & maintenance
* Operational, economic & business models
* Green data centers
* Dynamic resource provisioning

Services & Applications
* Cloud services models & frameworks
* Cloud services reference models & standardization
* Cloud-powered services design
* Business processes, compliance & certification
* Data management applications & services
* Application workflows & scheduling
* Application benchmarks & use cases
* Cloud-based services & protocols
* Fault-tolerance & availability of cloud services and applications
* Application development and debugging tools
* Business models & economics of cloud services

IoT & Mobile in the Cloud
* IoT cloud architectures & models
* Cloud-based dynamic composition of IoT
* Cloud-based context-aware IoT
* Mobile cloud architectures & models
* Green mobile cloud computing
* Resource management in mobile cloud environments
* Cloud support for mobility-aware networking protocols
* Multimedia applications in mobile cloud environments
* Cloud-based mobile networks and applications

Virtualization
* Computational resources, storage & network virtualization
* Resource monitoring
* Virtual desktops
* Resilience, fault tolerance, disaster recovery
* Modeling & performance evaluation
* Disaster recovery
* Energy efficiency

Big Data
* Machine learning
* Data mining
* Approximate & scalable statistical methods
* Graph algorithms
* Querying & search
* Data lifecycle management
* Frameworks, tools & their composition
* Dataflow management & scheduling

HPC in the Cloud
* Load balancing
* Middleware solutions
* Scalable scheduling
* HPC as a Service
* Programming models
* Use cases & experience reports
* Cloud deployment systems

Security & Privacy
* Accountability & audit
* Authentication & authorization
* Cloud integrity
* Cryptography for & in the cloud
* Hypervisor security
* Identity management & security as a service
* Prevention of data loss or leakage
* Secure, interoperable identity management
* Trust & credential management
* Trusted computing
* Usable security

Important Dates
Paper Submissions: **June 15, 2015**
PhD Consortium Paper Submissions: **July 15, 2015**
Notifications: **August 15, 2015**
Camera-ready: **September 15, 2015**

Call for Papers

European Conference on Ambient Intelligence (AmI 2015)

Athens, Greece, 11-13 November 2015

http://www.ami-conferences.org/2015

The European Conference on Ambient Intelligence (AmI) is the prime venue for research on Ambient Intelligence with an international and interdisciplinary character. It brings together researchers from the fields of science, engineering, and design working towards the vision of Ambient Intelligence which represents a future where we shall be surrounded by invisible technological means, sensitive and responsive to people and their behavior, deliver advanced functions, services and experiences. Ambient Intelligence combines concepts of ubiquitous technology, intelligent systems and advanced user interfaces putting the human in the center of technological developments.

AmI 2015 welcomes innovative, high quality research contributions advancing the state of the art in Ambient Intelligence. While the conference covers a breadth of AmI-related themes, this year's event pays special attention to the following themes each attracting a growing community of Ambient Intelligence researchers:

- AmI & Healthcare
- AmI & Well-being
- AmI & Social Robots
- AmI & Evaluation
- AmI & City
- AmI & Other Applications
- New & Emerging Topics

Program Committee

Honorary Chair
Emile Aarts, Eindhoven Univ. of Technology, NL

Program Chairs
Boris De Ruyter, Philips Research, NL
Periklis Chatzimisios, Alexander TEI of Thessaloniki, GR

Workshop Chairs
Andreas Komninos, University of Strathclyde, UK
Vassilis Koutkias, INSERM, FR

Thematic Chairs:
Massimo Zancanaro, Bruno Kessler Foundation, IT
Reiner Wichert, Fraunhofer, DE
Vanessa Evers & Gwenn Englebienne, University of Twente, NL
Vassilis Kostakos, Univ. of Oulu, FI
Dimitris Charitos, National Kapodistrian Univ. of Athens, GR
Christos Goumopoulos, University of the Aegean, GR
Ioannis Chatzigiannakis, University of Rome La Sapienza, IT

Conference co-Chairs
Achilles Kameas, Hellenic Open University, GR
Irene Mavrommati, Hellenic Open University, GR

Poster & Demos Chairs
Petros Nikopolitidis, Aristotle University of Thessaloniki, GR
Javed Vassilis Khan, Breda Univ. of Applied Sciences, NL

Academic & Professional Society Liaison
Athanassios Skodras, University of Patras, GR

Industrial Liaison
Kieran Delaney, Cork Institute of Technology, IE

Publicity Chair
Norbert A. Streitz, Smart Future Initiative, DE

The following Workshops will be organized in the context of AmI 2015 serving as a meeting point, aiming for intensive networking and scientific debate as well as shaping visions of the future:

- WS1: Industrial Human-Computer-Interaction
- WS2: Aesthetic Intelligence
- WS3: Discovery, Exploration and Understanding of Urban Social Context
- WS4: Affective Interaction with Avatars
- WS5: Designing for Ambient Intelligent Lighting

Important Dates (Full and short papers)

- Submission Deadline: **1 June 2015**
- Author notification: **1 July 2015**
- Camera ready deadline: **20 July 2015**

Call for Papers

IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2015)

September 7-9, 2015, Guildford, UK

http://www.ieee-camad.org

Scope

IEEE CAMAD 2015 will be held as a stand-alone event in the University of Surrey in the UK. The University of Surrey is home of the 5G Innovation Centre and is dedicated to conduct research and development in future mobile communication technology. IEEE CAMAD will continue to focus on 5G technologies this year. IEEE CAMAD will be hosting several special sessions, and will bring together scientists, engineers, manufacturers and service providers to exchange and share their experiences and new ideas focusing on research and innovation results under 5G. In addition to contributed papers, the conference will also include keynote speeches, panel and demo sessions.

Topics of interest include, but are not limited to, the following:

* Autonomic Communication Systems and Self-Organized Networks
* Body Area Networks and Applications
* Cloud Computing, Network Virtualization and Software Defined Networks
* Cognitive Radio and Network Design
* Cross-Layer & Cross-System Protocol Design
* Design of Content Aware Networks and Future Media Internet
* Design of Satellite Networks
* Design, Modeling and Analysis of Wireless, Mobile, Ad hoc and Sensor Networks
* Design, Modeling and Analysis of Network Services and Systems
* Future Service Oriented Internet Design
* Green Wireless Communication Design
* Modeling and Simulation Techniques for Integrated Communication Systems
* Modeling and Simulation Techniques for IoT and M-to-M Communications
* Modeling and Simulation Techniques for Social Networks
* Next Generation Mobile Networks
* Network Monitoring and Measurements
* Network Optimization and Resource Provisioning
* Next Generation Internet
* Optical Communications & Fiber Optics
* Network Management, Middleware Technologies and Overlays
* Quality of Experience: Framework, Evaluation and Challenges
* Seamless Integration of Wireless, Cellular and Broadcasting Networks with Internet
* Fast Simulation Techniques for Communication Networks
* Simulation Techniques for Large-Scale Networks
* Smart Grids: Communication, Modeling and Design
* Test Beds and Real Life Experimentation
* Traffic Engineering, Modeling and Analysis
* Validation of Simulation Models with Measurements

Important Dates

- Paper Submission (main track): **May 25, 2015 (FIRM)**
- Paper Submission (special sessions and short papers): **June 15, 2015**
- Notification of Acceptance: **July 1, 2015**
- Camera Ready Papers: **August 1, 2015**

Call for Papers

IEEE MASS 2015 Workshop on Content-Centric Networking

in conjunction with IEEE MASS 2015
Dallas, TX, USA, October 19, 2015

http://www.eng.auburn.edu/~szm0001/ccn2015/index.html

Scope

With the exponential growth of content in recent years (e.g., videos) and the availability of the same content at multiple locations (e.g., same video being hosted at Youtube, Dailymotion), users are interested in fetching a particular content and not where that content is hosted. Also, the ever-increasing numbers of mobile devices that lack fixed addresses call for a more flexible network architecture that directly incorporates in-network caching, mobility and multipath routing, to ease congestion in core networks and deliver content efficiently. By treating content as first-class citizen, Content-Centric Networking (CCN) aims to evolve the current Internet from a host-to-host communication based architecture to a content-oriented one where named objects are retrieved in a reliable, secure and efficient manner. CCN has been under active exploration over the past few years, resulting in both clean-slate and overlay architectures and solutions. This workshop will provide researchers and practitioners to meet and discuss the latest developments in this field. The outcomes of this workshop include 1) investigating and understanding some of the challenges in CCN; 2) fostering collaboration among researchers interested in CCN.

In recent years, rapid progress has been made in CCN; multiple initial architectural designs sharing common goals of in-network caching, mobility support and multipath routing have been proposed and prototypes have been implemented. Challenges related to caching and routing of content has received attention. Research areas focusing on what content to cache, how to route for content have been explored, but areas such as security, privacy and economic models for CCN have received limited attention.

The goal of this workshop is to bring together researchers from academia and industry and investigate the architectural issues and challenges in CCN. We invite submissions describing new research contributions including but not limited to the following topics:

- Content-oriented routing protocols
- Content naming
- Scalability issues in CCN
- CCN Architecture design and evaluation
- Security issues in CCN
- Privacy in CCN
- Content centric wireless networks
- Mobility management
- Evaluation of in-network caching techniques
- Limits and limitations of CCN architectures
- Economics and business models
- CCN specific transport protocols
- Specific implementations of CCN architectures

Important Dates

Paper submission: **July 1, 2015**
Paper Acceptance: **July 27, 2015**
Camera-ready paper: **Aug 1, 2015**

Submission Guidelines

Please follow the author instructions at http://www.eng.auburn.edu/~szm0001/ccn2015/index.html

All workshop papers will be included in the IEEE Proceedings.

Steering Committee

Jim Kurose

University of Massachusetts, Amherst, USA
http://www-net.cs.umass.edu/personnel/kurose.html

Workshop Organizers

Anand Seetharam

School of Computing and Design
California State University Monterey Bay, USA
Email: aseetharam@csumb.edu
http://itcdland.csumb.edu/~aseetharam/index.htm

Shiwen Mao

Department of Electrical and Computer Engineering
Auburn University, USA
Email: smao@ieee.org
http://www.eng.auburn.edu/~szm0001

Call for Papers

IEEE Transactions on Cloud Computing

Special Issue on "Mobile Clouds"

Mobile cloud computing represents one of the latest developments in cloud computing advancement. In particular, mobile cloud computing extends cloud computing services to the mobile domain by enabling mobile applications to access extern al computing and storage resources available in the cloud. Not only mobile applications are no longer limited by the computing and data storage limitations within mobile devices, nevertheless adequate offloading of computation intensive processes also has the potential to prolong the battery life.

Besides, there is also an incentive for mobile devices to host foreign processes. This represents a new type of mobile cloud computing services. Ad-hoc mobile cloud is one instance that mobile users sharing common interest in a particular task such as image processing of a local happening can seek collaborative effort to share processing and outcomes. Vehicular cloud computing is another instance of mobile cloud computing that exploits local sensing data and processing of vehicles to enhance Intelligent Transportation Systems.

This Special Issue will collect papers on new technologies to achieve realization of mobile cloud computing as well as new ideas in mobile cloud computing applications and services. The contributions to this Special Issue may present novel ideas, models, methodologies, system design, experiments and benchmarks for performance evaluation. This special issue also welcomes relevant research surveys. Topics of interest include, but are not limited to:

- Trends in Mobile cloud applications and services
- Architectures for mobile cloud applications and services
- Mobile cloud computing for rich media applications
- Service discovery and interest matching in mobile cloud
- Collaboration in mobile clouds
- Process offloading for mobile cloud computing
- Mobile device virtualization
- Mobile networks for cloud computing Mobile cloud monitoring and management
- Security and privacy in mobile clouds
- Performance evaluation of mobile cloud computing and networks
- Scalability of mobile cloud networks
- Software defined systems for mobile clouds
- Self-organising mobile clouds
- Mobile vehicular clouds
- Disaster recovery in mobile clouds
- Economic, social and environmental impact of mobile clouds
- Mobile cloud software architecture

Important Dates

Paper submission: **July 30, 2015 (extended)**
First Round Decisions: September 15, 2015
Major Revisions Due: November 15, 2015
Final Decisions: December 15, 2015
Publication: 2016

Submission & Major Guidelines

This special issue invites original research papers that present novel ideas and encourages submission of "extended versions" of 2-3 Best Papers from the IEEE Mobile Cloud 2015 conference. Every submitted paper will receive at least three reviews and will be selected based on the originality, technical contribution, and scope. Submitted articles must not have been previously published or currently submitted for publication elsewhere. Papers should be submitted directly to the IEEE TCC at https://mc.manuscriptcentral.com/tcc, and must follow TCC formatting guidelines. For additional information, please contact Chuan Heng Foh (c.foh@surrey.ac.uk).

Editor-in-Chief

Rajkumar Buyya, the University of Melbourne, Australia

Guest Editors

- Chuan Heng Foh, University of Surrey, UK
- Satish Narayana Srirama, University of Tartu, Estonia
- Elhadj Benkhelifa, Staffordshire University, UK
- Burak Kantarci, Clarkson University, USA
- Periklis Chatzimisios, Alexander TEI of Thessaloniki, Greece
- Jinsong Wu, Alcatel Lucent, China

MMTC OFFICERS (Term 2014 — 2016)

CHAIR

Yonggang Wen
Nanyang Technological University
Singapore

STEERING COMMITTEE CHAIR

Luigi Atzori
University of Cagliari
Italy

VICE CHAIRS

Khaled El-Maleh (North America)
Qualcomm
USA

Maria G. Martini (Europe)
Kingston University,
UK

Liang Zhou (Asia)
Nanjing University of Posts &Telecommunications
China

Shiwen Mao (Letters & Member Communications)
Auburn University
USA

SECRETARY

Fen Hou
University of Macau, Macao
China

E-LETTER BOARD MEMBERS (Term 2014—2016)

Periklis Chatzimisios	Director	Alexander TEI of Thessaloniki	Greece
Guosen Yue	Co-Director	Broadcom	USA
Honggang Wang	Co-Director	UMass Dartmouth	USA
Tuncer Baykas	Editor	Medipol University	Turkey
Tasos Dagiuklas	Editor	Hellenic Open University	Greece
Chuan Heng Foh	Editor	University of Surrey	UK
Melike Erol-Kantarci	Editor	Clarkson University	USA
Adlen Ksentini	Editor	University of Rennes 1	France
Kejie Lu	Editor	University of Puerto Rico at Mayagüez	Puerto Rico
Muriel Medard	Editor	Massachusetts Institute of Technology	USA
Nathalie Mitton	Editor	Inria Lille-Nord Europe	France
Zhengang Pan	Editor	China Mobile	China
David Soldani	Editor	Huawei	Germany
Shaoen Wu	Editor	Ball State University	USA
Kan Zheng	Editor	Beijing University of Posts & Telecommunications	China