IEEE MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE

E-LETTER

Vol. 4, No. 7, August 2009

CONTENTS

Message from Editor-in-Chief 2
HIGHLIGHT NEWS & INFORMATION 3
IEEE ICC 2010 @ Cape Town, South Africa 3
Message from the ICC 2010 TPC Co-Chairs 5
Chengshan Xiao, Missouri University of S&T, USA 5
Jan C. Olivier, University of Pretoria, South Africa 5
How to Submit Papers to ICC 2010 6
TECHNOLOGY ADVANCES 7
Distinguished Position Paper Series 7
A Framework on Multimodal Telecommunications from a Human Perspective ... 7
Zhengyou Zhang (IEEE Fellow), Microsoft Research, USA 7
Trends in Mobile Graphics 11
Kari Pulli, Nokia Research Center Palo Alto, USA 11
Assessing 3D Displays: Naturalness as a Visual Quality Metric 15
Wijnand A. IJsselsteijn, Eindhoven University of Technology, The Netherlands 15
Three-Dimensional Video Capture and Analysis 19
Cha Zhang, Microsoft Research, USA 19
3D Visual Content Compression for Communications 22
Karsten Müller,FraunhoferHeinrich Hertz Institute, Germany 22
Interactive 3D Online Applications 25
Irene Cheng, University of Alberta, Canada 25
Editor's Selected Paper Recommendation 34
Focused Technology Advances Series 35
Distributed Algorithm Design for Network Optimization Problems with Coupled Objectives 35
Jianwei Huang, the Chinese University of Hong Kong, China 35
MMTC COMMUNICATIONS & EVENTS 39
Call for Papers of Selected Journal Special Issues 39
Call for Papers of Selected Conferences 40
E-Letter Editorial Board 41

IEEE COMMUNICATIONS SOCIETY

Message from Editor-in-Chief

Time flies, now ICC 2010 is calling for paper submissions and the submission deadline is around the corner. In this letter, we introduce Cape Town, the beautiful city in South Africa, which is hosting the ICC 2010 next May, as well as the exciting 2010 FIFA World Cup Soccer tournament (right after the ICC 2010). The TPC Co-Chairs of ICC 2010 (Drs. Chengshan Xiao and Jan C. Olivier) sent a warm message to our MMTC members on page 6 to encourage all of us to support this event by submitting papers and proposals before the coming deadlines. On page 7, detailed instructions on how to submit your papers to ICC 2010 are provided.

In this Issue, we feature a Special Issue on *3D Visual Communications*, which contains six position papers from world top scientists in the field. In addition, in the Editor's Selected Paper Recommendation column, our Column Editor, Dr. Guan-Ming Su (Marvell Semiconductors, USA) recommend the paper "*Reliable multimedia transmission over cognitive radio networks using fountain codes*" by H. Kushwaha et al. Please check out more details in the article.

Furthermore, in the focused technology column, Prof. Jianwei Huang (Chinese University of Hong Kong, China) introduces a new framework of designing distributed algorithms for maximizing the network utility functions, thus optimizing the networking system performance.

Now let us explore inside the Special Issue on *3D Visual Communications*, which covers all stages of 3D content lifecycle including content capturing, analysis, compression, rendering and communications, and 3D quality evaluation. The papers are contributed by researchers in the graphics, vision, signal processing, and psychological fields, thus they convey perspectives from different angles.

The Distinguished Position paper delivered by Dr. Zhengyou Zhang (Microsoft Research, USA) highlights a new framework for real-time immersive telecommunications system design, where the receiver's expectation and attention are modeled to drive the whole content lifecycle from capturing, processing, delivery to rendering.

In the second paper, Dr. Kari Pulli (Nokia Research, USA) discusses the future trends of mobile graphics applications, such as user interfaces, maps and navigation systems, games, web applications, and augmented reality systems, and so. Clearly 3D development in the mobile device has a long way to go. After that, Prof. Wijnand IJsselsteijn (Eindhoven University of Technology, The Netherlands) demonstrates the latest development on the 3D visual quality measurement efforts.

Dr. Cha Zhang (Microsoft Research, USA) introduces the state of the art and the major challenges in the 3D content capturing and analysis in the fourth paper, which is naturally followed by the paper of Dr. Karsten Müller, (Heinrich Hertz Institute, Germany), in which the 3D content compression approaches are overviewed with both graphics based and vision based method categories.

At the end, Dr. Irene Cheng (University of Alberta, Canada) reviews the current advances and future trends in developing interactive 3D technology for reality enhanced visual communication applications. The strategies designed in the 3D compression and transmissions to achieve such goals are also demonstrated in the paper.

Clearly there are many other challenges and issues that have not been fully covered in this Special Issue. Hopefully we will get them covered in our future Issues.

As always, I thank all Editors of the E-Letter, and our authors to make this issue successful.

Thank you very much.

Haohong Wang
Editor-in-Chief, MMTC E-Letter

HIGHLIGHT NEWS & INFORMATION

IEEE ICC 2010 @ Cape Town, South Africa

IEEE ICC 2010 is to be held during 23-27 May 2010 in Cape Town, South Africa. The city of Cape Town, being voted one of the most beautiful cities in the world, welcomes conference attendees from all over the world.

The theme of IEEE ICC 2010, "Communications: Accelerating Growth and Development," is specifically matched to the conference location, emphasizing the role of communications that is playing in the rapid development of Africa.

As shown in the map above, Cape Town is located at the northern end of the Cape Peninsula. Table Mountain forms a dramatic backdrop to the City Bowl, with its plateau over 1,000 m (3,300 ft) high; it is surrounded by near-vertical cliffs, Devil's Peak and Lion's Head. Sometimes a thin strip of cloud forms over the mountain, and owing to its appearance, it is colloquially known as the "tablecloth". The peninsula consists of a dramatic mountainous spine jutting southwards into the Atlantic Ocean, ending at Cape Point. There are over 70 peaks above 300m within Cape Town's official city limits.

Cape Town is not only the most popular international tourist destination in South Africa; it is Africa's main tourist destination even overtaking Cairo. This is due to its good climate, natural setting, and relatively well-developed infrastructure. Table Mountain is one the most notable attractions. Reaching the top of the mountain can be achieved either by hiking up, or by taking the Table Mountain Cableway. On the other hand, Chapman’s Peak Drive, a narrow road that links Noordhoek with Hout Bay, also attracts many tourists for the views of the Atlantic Ocean and nearby mountains. It is possible to either drive or hike up Signal Hill for closer views of the City Bowl and Table Mountain.

At the tip of the Cape Peninsula you will find Cape Point within the Table Mountain National Park. The expansive Table Mountain National Park stretches from Signal Hill and Table Mountain in the north to Cape Point in the south and encompasses the seas and coastline of the peninsula. Within Cape Point the treacherous cliffs forming the most southwestern tip of Africa are some of the highest in the world and mark the spot where the cold Beguela current on the West coast and the warm Agulhus current on the East coast merge.

Cape Town is also famous for the beaches due to the city’s unique geography. It is possible to visit several different beaches in the same day, each with a different setting and atmosphere. Beaches located on the Atlantic Coast, for example Clifton beach, tend to have very cold water from the Benguela current which originates from the Southern Ocean,

Cape Town is noted for its architectural heritage, with the highest density of Cape Dutch style buildings in the world. Cape Dutch style combines the architectural traditions of the Netherlands, Germany and France.

Cape Town is also a host city for the 2010 FIFA World Cup Soccer tournament, which starts just after the conference ends.

If you have not started to plan for your family’s vacation in 2010 yet, please do consider Cape Town. Remember to have your papers (or proposal) ready as soon as possible for ICC 2010 following the deadlines below:

Paper Submission:	**September 10**, 2009
Tutorial Proposal:	**September 10**, 2009
Panel Proposal:	**August 28**, 2009

See you in Cape Town next May!

IEEE ICC 2010 @ Cape Town, South Africa

Communications: Accelerating Growth and Development!!!

Message from the ICC 2010 TPC Co-Chairs

Chengshan Xiao, Missouri University of S&T, USA
Jan C. Olivier, University of Pretoria, South Africa

The flagship conference in communications, IEEE International Conference on Communications (IEEE ICC 2010), will be held in Cape Town, South Africa, from 23 to 27 May 2010, just before the FIFA World Cup Soccer tournament hosted by South Africa for 2010.

IEEE ICC 2010 will feature a comprehensive technical program including 11 symposia and a number of tutorials and workshops. IEEE ICC 2010 will also feature an attractive program including keynote speakers, business and technology panels, and industry forums. Prospective authors are invited to submit original papers by the deadline 10 September 2009 for publication in the IEEE ICC 2010 Conference Proceedings and presentations for the symposia. For detailed information about each symposium and submission guidelines, please visit our website at http://www.ieee-icc.org/2010/.

The technical program committee (TPC) including the symposium co-chairs and their invited TPC members is committed to provide a comprehensive peer review and selection process to further raise the quality of the conference. All the accepted and presented papers at the conference will be published online via IEEE Xplore. We encourage everyone in this community to participate in the conference by submitting papers, proposals for symposia, tutorials and workshops, and most importantly by attending the conference in beautiful and exotic Cape Town, where you will have a wide range of exciting options to add to your conference tour: hikes up the famous Table Mountain; beautiful white sandy beaches; whale watching; tours of the Cape Winelands, historic Robben Island where Nelson Mandela was held for 27 years; game reserves with Safaris and the big 5; and even Cape Point where the Atlantic ocean and the Indian oceans meet with its penguin inhabitants.

IEEE ICC 2010 will be the first Communications Society flagship conference to be held on the African Continent. The organizing committee has been working diligently to make sure that IEEE ICC 2010 will be a great and fantastic conference. See you in Cape Town.

Thank you and best regards,

Chengshan Xiao and Jan C. Olivier
IEEE ICC 2010 Technical Program Co-Chairs

How to Submit Papers to ICC 2010

For ICC 2010, **Multimedia Services, Communication Software and Service Symposium (MCS)** is the **ONLY** symposium that is fully sponsored by MMTC, hence we encourage all our members to submit your Multimedia related papers to MCS symposium, where a Co-Chair and many TPC members recommended from MMTC would handle all the review process of the multimedia related paper submissions.

Here are the few steps to submit your papers to ICC 2010:

(1) Go to EDAS website: http://edas.info/index.php and sign into your account;
(2) Click the "**Submit paper**" Tab on the top of the webpage;
(3) In the list of the conferences, find "**ICC'10**" and click the icon on the rightmost column;
(4) In the list of ICC'10 symposia, please select the "**ICC'10 MCS**" and click the icon on the rightmost column;
(5) Start the normal paper submission process.

TECHNOLOGY ADVANCES

Distinguished Position Paper Series

A Framework on Multimodal Telecommunications from a Human Perspective

Zhengyou Zhang (IEEE Fellow), Microsoft Research, USA
zhang@microsoft.com

In this position paper, I am arguing for researching and developing multimodal telecommunications systems from the perspective of users who are the communication creatures. Human's innate communications skills are the natural consequence of hundreds of thousands of years of evolution. We should follow and leverage our own skills rather than trying to change our behaviors.

One of the major goals of communications is to get messages across to people on the other side. There are a variety of messages in any face-to-face communication. However, one can basically find three elements, known as "3 Vs", behind each message:

- Verbal: Words, what you say;
- Vocal: Tone of voice, how you say the words;
- Visual: Facial expression, gaze, body language.

The second and third elements are sometimes simply referred to as non-verbal elements. Prof. Albert Mehrabian conducted experiments dealing with communications of feelings and attitudes (i.e., like-dislike) in 1960s and found that the non-verbal elements are particularly important for communicating feelings and attitude, especially when they are incongruent: if words and body language disagree, one tends to believe the body language (Mehrabian, 1981). More concretely, according to Mehrabian, when a person talks about their feelings or altitudes, the three elements account for 7%, 28% and 55%, respectively. The exact numbers might be arguable depending on experimental settings, but clearly, non-verbal messages are extremely important, as confirmed by other researchers (Argyle, Salter, Nicholson, Williams, & Burgess, 1970).

Communications across distances are much harder, and in our human history, many tools have been invented to overcome this difficulty:

- Letters: Verbal
- Telephony: Verbal + Vocal
- Videoconferencing: Verbal + Vocal + Visual

Videoconferencing is leveraging all 3 Vs. So, are people satisfying with the videoconferencing systems we have? Obviously, the answer is no.

Let's take a look at the visual aspect of some nonverbal behaviors in face-to-face communications:

- Facial expressions
- Eye gaze
- Eye gestures
- Head gaze
- Head gestures
- Arm gestures
- Body postures

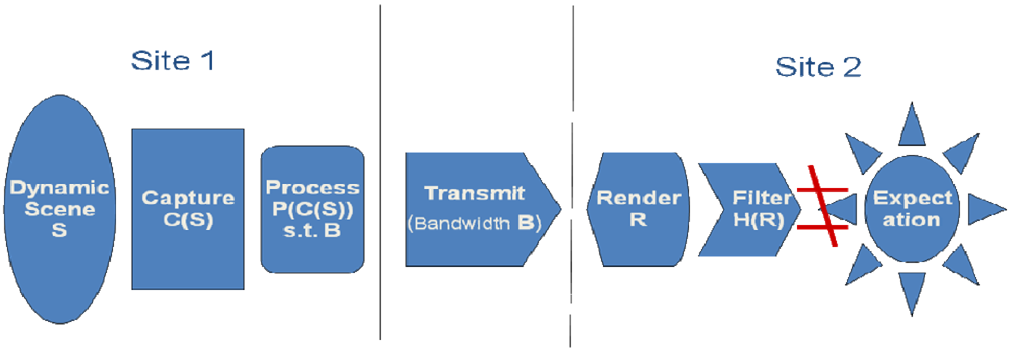
How much can a current videoconferencing system convey? Not much, mostly facial expressions and head gestures. Other behaviors are lost. Even if facial expressions and head gestures are displayed on the video, without other elements, say eye gaze, one may get an imprecise or even wrong perception (who is she really smiling to?).

In the area of audio, there is also a lot to desire. Current conferencing systems (audio only or audio-visual) are essentially monaural. When multiple people are involved in conferencing, one major problem is that a participant at one end has difficulties in identifying who is talking at the other end and comprehending what is being discussed. The reason is that the voices of multiple participants are intermixed into a single audio stream. This is in sharp contrast to face-to-face communications where sounds are coming from all directions and human can perceive the direction and distance of a sound source using two ears. Human auditory systems exploit the knowledge of cues such as the interaural time difference (ITD) and the interaural level difference (ILD) due to distance between the two ears and the shadowing by the head (Blauert, 1983).

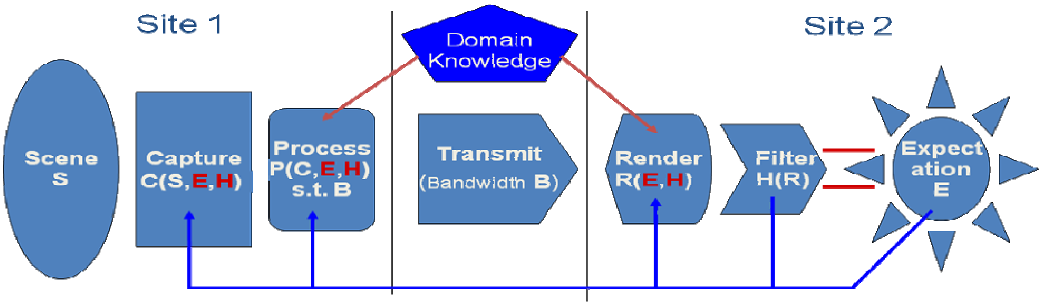
Another important aspect to consider is attention, which is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things. For example, humans have the ability to focus one's listening attention on the voice of a particular talker in an environment in which there are many others speaking at the same time, known as the cocktail party effect (Bregman, 1990). For an effective communication, the listener must pay attention and the speaker must react to where the listener's attention is. In a face-to-face communication, the bandwidth between people can be considered as boundless, but a person can only consume a certain amount of information due to his/her attention. In telecommunication, bandwidth is limited, and it is even more crucial to be aware of each participant's attention and deliver only the information important to the participants.

Decades ago, Herbert Simon already articulated very nicely: "...in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it" (Simon, 1971). When we build telecommunications systems, we should consider attention scarcity as a design requisite, and filter out unimportant or irrelevant information. With the rapid growth of bandwidth capacity and the decrease in cost and increase in quality of audio-visual capture devices, it is too easy to overload information.

Below is a diagram illustrating the working of current telecommunication systems. The information flows from the source (here Site 1) to the destination (here Site 2). Filter H(R) represents a person's attention and sensory limitation. The person at Site 2 has his/her expectation of what he/she wants. However, capture devices, processing algorithms and rendering equipment are pre-designed to meet what the designer expects the expectation is. The information the person at Site 2 actually obtains usually does not meet his/her expectation.



Below is a diagram illustrating a new framework I am proposing. The capture devices, processing algorithms and rendering equipment should be designed to reflect the receiver's expectation and attention in real-time.



There are a number of challenges in this framework:

- How to model users' expectation?
- How to determine users' attention?

- How to accommodate the deficiency of human perception?
- How to incorporate users' expectation and attention in each component?
- Real-Time! Real-Time!! Real-Time!!!

There have been several efforts along this direction. One example is the boardroom high-end video conferencing systems including HP Halo and Cisco TelePresence. My research group has been working on a number of projects. One is audio spatialization to leverage cocktail party effect. We have done audio spatialization over headphones (Chen & Zhang, 2009) as well as over loudspeakers (Zhang, Cai, & Stokes, 2008). For the latter, we have to address the issue of multichannel acoustical echo cancellation. The other is the Personal Telepresence Station (Knies, 2009), which provides the correct gaze cues in multiparty conferencing through video spatialization. We also work on 3D videoconferencing to make conferencing visually more immersive. A lot more, however, needs to be done, and this is an exciting area and the right time to work on.

Acknowledgment

Many ideas in this paper were presented in an ICME 2006 panel and in an ImmersCom 2009 panel.

Bibliography

[1] Argyle, M., Salter, V., Nicholson, H., Williams, M., & Burgess, P. (1970). The communication of inferior and superior attitudes by verbal and non-verbal signals. *British journal of social and clinical psychology* , *9*, 222-231.

[2] Baldis, J. (2001). Effects of spatial audio on memory, compression, and preference during desktop conferences. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 166-173). Seattle, USA: ACM Press.

[3] Blauert, J. (1983). *Spatial Hearing.* MIT Press.

[4] Bregman, A. (1990). *Auditory Scene Analysis.* MIT Press.

[5] Chen, W.-g., & Zhang, Z. (2009). Highly realistic audio spatialization for multiparty conferencing using headphones. *IEEE International Workshop on Multimedia Signal Processing (MMSP'09)* .

[6] Knies, R. (2009, March 9). *Making Virtual Meetings Feel Real.* Retrieved from Microsoft Research: http://research.microsoft.com/en-us/news/features/personaltelepresencestation-030909.aspx

[7] Mehrabian, A. (1981). *Silent messages: Implicit communication of emotions and attitudes (2nd ed.).* Belmont, California: Wadsworth.

[8] Moore, B. C. (2003). *An Introduction to the Psychology of Hearing* (Vol. 5th Edition). Academic Press.

[9] Simon, H. A. (1971). Designing Organizations for an Information-Rich World. In M. Greenberger, *Computers, Communication, and the Public Interest* (pp. 40-41). The Johns Hopkins Press.

[10] Zhang, Z., Cai, Q., & Stokes, J. (2008). Multichannel Acoustic Echo Cancelation in Multiparty Spatial Audio Conferencing With Constrained Kalman Filtering. *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC'08)* .

Zhengyou Zhang (SM'97–F'05) received the B.S. degree in electronic engineering from the University of Zhejiang, Hangzhou, China, in 1985, the M.S. degree in computer science from the University of Nancy, Nancy, France, in 1987, and the Ph.D. degree in computer science and the Doctor of Science (*Habilitation à diriger des recherches*) diploma from the University of Paris XI, Paris, France, in 1990 and 1994, respectively. He is a Principal Researcher with Microsoft Research, Redmond, WA, USA, and manages the multimodal collaboration research team. Before joining Microsoft Research in March 1998, he was with INRIA (French National Institute for Research in Computer Science and Control), France, for 11 years and was a Senior Research Scientist from 1991. In 1996-1997, he spent a one-year sabbatical as an Invited Researcher with the Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan. He has published over 200 papers in refereed international journals and conferences, and has coauthored the following books: *3-D Dynamic Scene Analysis: A Stereo Based Approach* (Springer-Verlag, 1992); *Epipolar Geometry in Stereo, Motion and Object Recognition* (Kluwer, 1996); and *Computer Vision* (Science Publisher, 1998, 2003, in Chinese). He has given a number of keynotes in international conferences.

Dr. Zhang is an IEEE Fellow, the Founding Editor-in-Chief of the *IEEE Transactions on Autonomous Mental Development*, an Associate Editor of the *International Journal of Computer Vision*, an Associate Editor of the *International Journal of Pattern Recognition and Artificial Intelligence*, and an Associate Editor of *Machine Vision and Applications*. He served as an Associate Editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* from 2000 to 2004, an Associate Editor of the *IEEE Transactions on Multimedia* from 2004 to 2009, among others. He has organized or participated in organizing numerous international conferences. More details are available at http://research.microsoft.com/~zhang/.

Trends in Mobile Graphics

Kari Pulli, Nokia Research Center Palo Alto, USA
kari.pulli@nokia.com

Mobile handheld devices such as mobile or cellular phones, PDAs, and PNDs (personal navigation devices) are the new frontier of computer graphics, bringing graphics to everyone and everywhere. Mobile phone is the most widely used electronic device of any kind, and the current high-end phones have as much processing power as a typical laptop had perhaps 7 years ago. It is the latest incarnation in the progression of more compact computing, starting from main frames and moving on to mini computers, PCs, laptops, and finally smartphones, phones where users can install new applications and that are almost general computers.

Around the turn of the millenium, when the first mobile graphics systems were being created, there were in practice only two target applications: games, and so-called screen savers showing simple animations. Games still remain an important application area, but there are others that are currently even more important [2]. The most important of them is graphical user interfaces (GUIs), followed by navigation applications. The demand from these applications drive development in new APIs and mobile graphics hardware, which again enable interesting new areas such as 3D web browsers and augmented reality (AR) applications.

User interfaces

The early user interfaces on mobile devices were simple and text-based. As the displays got better, more eye candy was added in form of application icons and simple animations. Now as touch displays allow most of the device’s front surface to be a display, so that less real estate needs to be reserved for buttons, people have come to expect beautiful graphics and fast transitions, as first demonstrated by the iPhone. Such large displays increasingly require dedicated graphics hardware, as software-based graphics engines are not fast enough, or are too power-hungry to drive real-time graphics on large displays. There are two main APIs for mobile GUIs, OpenVG [4] and OpenGL ES [11,8]. OpenVG was created both for presentation graphics, such as SVG (scalable vector graphics), PDF, Flash, and PowerPoint, and for user interfaces. Even though it is a 2D API, it can be used for 3D-like graphics, as long as the 3D surfaces consist of a relatively small number of flat polygons. Even though all 2D APIs allow the application to project the polygon corners from 3D to 2D, most 2D APIs do not support perspective transformations of images used to texture map the polygons; OpenVG does provide that support. The other possibility is to use OpenGL ES to draw the UI graphics, even if the UI is not really 3D. Currently most UIs are really 2D, and although they may have some 3D appearance, and show some 3D objects, the graphics hardware is really needed to provide smooth animation transitions. In fact, even though there have been several 3D UIs created by researchers, none of them have caught on, since 2D UIs are easier to use and navigate. OpenVG can be emulated on OpenGL ES 2.0 hardware, but not as power-efficiently as when the hardware design is originally targeted for OpenVG.

Maps and navigation

PNDs (personal navigation devices) have become ubiquitous as they have become cheaper. Even though it may be fun to view maps and plan trips on a desktop computer, one has the greatest need for navigation instructions while traveling, be that with a car, bicycle, or on foot. Most new smartphones also come with built-in GPS (global positioning system) receivers, maps, and navigation routing services, some even with a compass to facilitate orienting the maps correctly.

The first maps consisted of simple bitmap images. The current trend, however, is to move to using vector graphics such as SVG to render the maps. Maps are naturally viewed at different levels of resolution, especially on small displays, and with vector graphics every scale can be rendered neatly without scaling artifacts, and without having to store or transmit new pixel data at every different resolution level.

The next trend is to make the maps 3D [9]. The easiest 3D feature to add is topography, rendering geographic structures such as mountains and hills. The next step is to add 3D models of visible landmarks, such as the Eiffel tower in Paris or the Big Ben in London. As one continues to increase the realism of the displays, many buildings can be cleanly and compactly represented in a procedural form, such that only the ground extent, building height, and some style and color parameters have to be stored, and a 3D representation is created from the description on the fly. Full realism may not even be a desirable target in a navigation system as they may more distract the user than aid her. Symbolic representations with simple, clear lines, and clear audio instructions may be better choices, especially when one is driving a car and needs to make decisions about where to drive quickly.

Games

The first mobile games had very simple graphics. Snake was a very popular mobile phone game, where the snake grows in length as one captures more points, and becomes more difficult to maneuver without the head hitting a wall or the snake itself and thus ending the game. With color displays came various sports and rally games, and today games of almost any genre are available on mobile devices.

Before standard mobile 3D APIs games came each with their own proprietary software rendering engines. Even the launch of OpenGL ES 1.0 and 1.1 didn’t change the situation much, as a special engine is always faster than a generic one. As graphics hardware began to appear, the games began to use standard APIs, and soon proprietary engines will disappear as OpenGL ES 2.0, which allows a new level of visual richness, will be widely available.

While native games (written in C or C++) are very closely tied to a particular device family, and even to an individual model, mobile Java aims for application portability across different device manufacturers. M3G (JSR-184) was defined to be compatible with OpenGL ES so graphics hardware developed for OpenGL ES could be leveraged for M3G [11]. In addition to basic rendering, the API supports features useful in game engines such as scene graphs, sophisticated animation, and a compact file format for 3D content. While fast-paced graphics-rich action games tend to be native and use OpenGL ES, far more games have been written with the more portable M3G [1]. As OpenGL ES 2.0 allows shaders for native applications, M3G 2.0 (JSR-297) allows use of shaders from mobile Java applications.

Web and browsers

Mobile web used to, and had to be different from the desktop web experience, as the mobile data bandwidth and the screen sizes were both vastly smaller than on a desktop or laptop PC. This is now rapidly changing, and new smartphones have very complete web browsers providing a “one-web” experience, where the web is experienced in essentially the same way on any device. This is one of the new drivers for graphics performance on mobile phones.

Another recent development is re-birth of the idea of 3D graphics on the web. VRML, and later X3D attempted to be standard 3D descriptions for web content, but they never really took off. There were also proprietary 3D engines running on browsers, (e.g., Macromedia’s Shockwave), but they didn’t succeed either. Khronos group has a new working group for 3D Web, where they are creating essentially OpenGL ES 2.0 bindings to JavaScript. Most major browsers are represented in the working group, and the new standard may finally bring hardware-accelerated 3D graphics both to mobile and desktop browsers.

Augmented Reality

Augmented Reality (AR) enhances or augments user’s sensations of the real world around her, typically by rendering additional objects, icons, or text that relate to the real objects. Whereas traditional AR systems tended to have wearable displays, data goggles, the mobile AR systems capture the video stream from the camera on handheld device’s rear side, and post the augmented image on the display, providing a “magic lens”. A use case for an AR system could be a tourist guide helping the user to navigate to a destination and labeling objects of interest, or an educational system pointing to some device’s controls in a correct sequence, together with operation instructions. AR is a class of graphics applications that makes much more sense in a mobile setting than at one’s desk with a desktop PC.

A key enabler for AR is image recognition capability, so the system knows what is visible, and also where in the field of view of the camera the object is located, so that it can be annotated or augmented. A secondary tool is efficient tracking, so that the labels, arrows, icons, etc., remain attached to the correct objects even when the user moves the camera around [5, 12]. Recent advances in computer vision algorithms, together with increased capabilities of camera phones equipped with other sensors such as GPS and compass, allow simple mobile AR applications running in real time. There are demos and startups that provide information overlay using only GPS and compass [6, 7], and others that also use cameras to recognize targets [10]. Some simple AR games only require simple motion tracking: in Mosquito Hunt from Siemens virtual mosquitos drawn over the background of a camera image seem to attack the user, who tries to get them to a cross hair and zap them first.

GPUs and GPGPU

The driving force of desktop graphics processing units (GPU) has been increasing performance in terms of speed. While speed is important also with mobile GPUs, even more important is low power consumption. Their design differs from desktop GPUs, for example mobile GPUs can't afford to get performance by simply adding more parallel processing. The first generation of mobile GPUs was designed for the fixed functionality of OpenGL 1.X, the second generation allows use of vertex and fragment shaders of OpenGL 2.0.

As desktop GPUs got more powerful, programmers wanted to use them also for other calculations. General purpose processing on GPUs, also known as GPGPU, uses graphics hardware for generalized number crunching. GPGPU is also now becoming attractive on mobile devices. Many image processing and computer vision algorithms can be mapped to OpenGL ES 2.0 shaders, benefiting from the speed and power advantages of GPUs. Unfortunately the second generation of mobile GPUs was only designed for graphics processing and does not support efficient transmission of data back to CPU from GPU. Careful scheduling of processing may allow data readback from a previous image to proceed while the next image is being processed, reducing the transfer overhead. With the advent of OpenCL (Open Computing Language) [3], a GPGPU API similar to NVidia's CUDA except that it is vendor-independent, and can run on CPUs and DSPs in addition to GPUs, the third generation of mobile GPUs will have better support for efficient data transfer. OpenCL will allow better portability for many image processing algorithms both on desktop and on mobile devices.

REFERENCES

[1] M. Callow, P. Beardow, and D. Brittain, "Big Games, small screens", *ACM Queue*, Nov./Dec., pp. 2-12, 2007.

[2] T. Capin, K. Pulli, and T. Akenine-Möller, "The State of the Art in Mobile Graphics Research", *IEEE Computer Graphics and Applications*, vol. 28, no. 4, pp. 74 - 84, 2008.

[3] Khronos OpenCL Working Group, A. Munshi Ed., "The OpenCL Specification, Version 1.0, Rev. 43", Khronos Group, USA, May 2009. http://www.khronos.org/opencl/

[4] Khronos OpenVG Working Group, D. Rice and R. Simpson Eds., "OpenVG Specification, Version 1.1", Khronos Group, USA, Dec. 2008. http://www.khronos.org/openvg/

[5] G. Klein, "PTAM + AR on an iPhone", http://www.youtube.com/watch?v=pBI5HwitBX4

[6] Layar: http://layar.eu/

[7] MARA: http://www.technologyreview.com/Biztech/17807/

[8] A. Munshi and D. Ginsburg, "OpenGL ES 2.0 Programming Guide", *Addison-Wesley Professional*, 2008.

[9] A. Nurminen, "Mobile 3D City Maps", *IEEE Computer Graphics and Applications*, vol. 28, no. 4, pp. 20-31, 2008.

[10] Point&Find: http://pointandfind.nokia.com

[11] K. Pulli, T. Aarnio, V. Miettinen, K. Roimela, and J. Vaarala, "Mobile 3D Graphics with OpenGL ES and M3G", *Morgan Kauffman*, 2006.

[12] D. Wagner, "High-speed natural feature tracking", http://studierstube.icg.tu-graz.ac.at/handheld_ar/highspeed_nft.php

Kari Pulli received his Ph.D. in Computer Science from University of Washington in Seattle, WA, USA and an MBA from University of Oulu, Finland. He is currently a Research Fellow at Nokia Research Center in Palo Alto, CA http://research.nokia.com/people/kari_pulli/. He has been an active contributor to several Khronos standards, Research Associate at Stanford University, Adjunct Professor (Docent) at University of Oulu, Visiting Scientist at MIT, and briefly worked at Microsoft, SGI, and Alias|Wavefront.

Assessing 3D Displays: Naturalness as a Visual Quality Metric

Wijnand A. IJsselsteijn, Eindhoven University of Technology, The Netherlands
w.a.ijsselsteijn@tue.nl

Binocular vision is based on the fact that objects and environments are seen from two separate vantage points, created by the interocular distance between the two eyes. This horizontal separation causes an interocular difference in the relative projections of monocular images onto the left and right retinas. When points from one eye's view are matched to corresponding points in the other eye's view, the retinal disparity variation across the image provides information about the relative distances and depth structure of objects. Stereopsis thus acts as a strong depth cue, particularly at short distances [1].

Stereoscopic display techniques are based on the principle of taking two images and displaying them in such a way that the left view is seen only by the left eye, and the right view seen only by the right eye. There are a number of ways of achieving this [2, 3, 4]. Stereoscopic displays can be categorized based on the technique used to channel the right and left images to the appropriate eyes. A distinguishing feature in this regard is whether the display method requires a viewing aid (e.g., polarized glasses) to separate the right and left eye images. Stereoscopic displays that do not require such a viewing aid are known as autostereoscopic displays. They have the eye-addressing techniques completely integrated into the display itself. Other distinguishing features are whether the display is suitable for more than one viewer (i.e., allows for more than one geometrically correct viewpoint), and whether look-around capabilities are supported, a feature inherent to a number of autostereoscopic displays (e.g., holographic or volumetric displays), but which requires some kind of head-tracking when implemented in most other stereoscopic and autostereoscopic displays.

Potential benefits of stereoscopic displays include an improvement of subjective image quality, improved separation of an object of interest from its visual surrounding, and better relative depth judgment and surface detection. Such characteristics make the use of stereoscopic displays advantageous to a variety of fields, including camouflage detection (e.g., stereoscopic aerial photography), precision manipulation in teleoperation or reduced vision environments (e.g., minimally invasive telesurgery, underseas operations), as well as communication and entertainment environments through creating a greater sense of social or spatial presence [5].

The effectiveness of stereoscopic displays in supporting certain tasks can be assessed in a relatively straightforward manner, by taking certain quantitative outcome measures (e.g., number of errors, time taken to successfully complete a task) and comparing monoscopic and stereoscopic modes of operation in terms of such measures. Such a performance-oriented approach will not be applicable, however, when stereoscopic displays are being developed and deployed for entertainment purposes, such as television or digital games. In an appreciation-oriented context, the standard assessment criterion that is applied to assess the fidelity of image creation, transmission and display systems is image quality [6].

Engeldrum [7] has defined image quality as “the integrated set of perceptions of the overall degree of the excellence of the image” (p.1). Image quality research has traditionally focused on determining the subjective impact of image distortions and artifacts that may occur as a consequence of errors in image capture, coding, transmission, rendering, or display. In stereoscopic imaging, image quality research has focused on effects of (i) stereoscopic image compression, (ii) commonly encountered optical errors (image shifts, magnification errors, rotation errors, keystone distortions), (iii) imperfect filter characteristics (luminance asymmetry, color asymmetry, crosstalk), and (iv) stereoscopic disparities (in particular vertical disparities as a consequence of toed-in camera configurations). More recently, with the advent of the RGB+Z format [8], new image artifacts have been introduced that require more extensive quality evaluations.

The International Telecommunication Union (ITU) offers a series of recommendations (standards) for the optimal assessment of picture quality of television images. Two recommendations are of specific importance to us here. The first, ITU Recommendation BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, describes the proper application of several scaling techniques for assessing the quality of television images (e.g., double-stimulus continuous quality-scale (DSCQS) method, double-stimulus impairment scale (DSIS) method). The recommendation includes descriptions of the general setting, presentation of test materials, viewing parameters, observer characteristics, grading scales, and analysis of results.

The second recommendation of interest in this context is BT.1438, Subjective Assessment of Stereoscopic Television Pictures. This recommendation describes the assessment factors that are peculiar to stereoscopic television systems, including depth resolution, depth motion, puppet theatre effect, and cardboard effect. The same quality assessment method as described in BT.500-11 is recommended for use in the case of evaluating stereoscopic television pictures. These recommendations enable engineers to optimize and enhance their systems and to evaluate competitive systems.

Many studies report a clear preference for stereoscopic images over their 2D counterparts [5, 9-11]. So, the introduction of the third dimension clearly contains an added value. However, over recent years, several experiments have pointed in the direction that image quality may not be the most appropriate term to capture the evaluation processes associated with 3D images [5, 9, 12]. Tam et al. [9] reported a low correlation between image quality and depth, indicating that when observers are asked to assess perceived image quality of MPEG-2 compressed images, the scores are mainly determined by the introduced impairments and not by depth. More recently, Seuntiëns et al. [12] showed that JPEG coding artifacts degrade the image quality, but that different camera base distances (CBD) do not influence the image quality. As IJsselsteijn [5] has argued, it appears that the added value of depth in 3D images if compared to 2D images can clearly be identified by image quality evaluations as long as the pictures are not impaired with image artifacts such as noise, blockiness or blur. When the image is impaired, the quality assessment becomes more dependent on the grade of the impairment and much less on the depth dimension. Taken together, these results point towards the need for additional assessment criteria or terms that allow researchers to measure the added value of depth, even when images contain artifacts or distortions due to compression, conversion, rendering, or display.

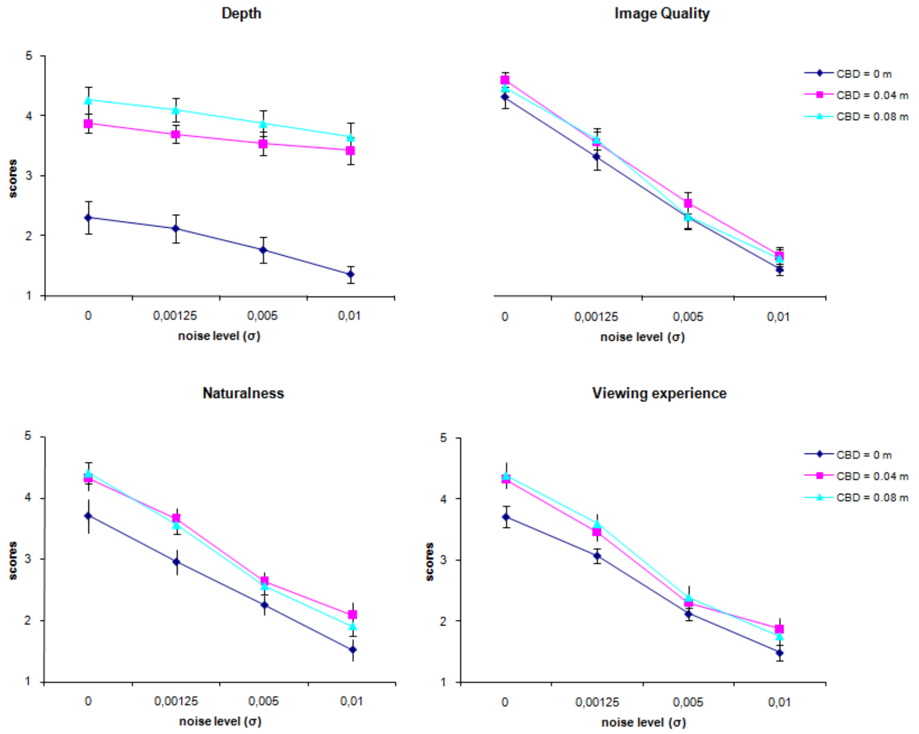
At the 3D Interaction, Communication and Experience (3D/ice) Lab of Eindhoven University of Technology, we are running an extensive research programme investigating a broad set of human factors issues associated with 3D displays, including work on assessing and minimizing visual (dis)comfort associated with (auto)stereoscopic displays, quantifying performance benefits of the use of 3D displays in various medical domains, and assessing the effectiveness of stereoscopic displays in supporting (non-verbal) communication in immersive teleconferencing. One of the longstanding lines of research at our lab has been the assessment of stereoscopic displays from an appreciation oriented perspective, which includes work on image quality and other candidate metrics that are potentially able to more fully characterize the visual experience associated with 3D displays (see, e.g., [13]). In the recent past, two specific evaluation terms, viewing experience and naturalness, have shown promising results [14].

We have recently completed a series of three experiments where we systematically investigated the sensitivity of image quality, naturalness and viewing experience in relation to the introduction of stereoscopic disparity (i.e., measuring the added value of depth) as well as the introduction of image artifacts. The experiments cover multiple display types (stereoscopic and autostereoscopic), multiple image distortions (blur and noise), and a variation in image content and stereoscopic disparity settings. The figure below shows some representative results from the series of experiments. Participants were asked to provide an assessment of a number of stereoscopic stills, which varied in disparity and image distortion. In the figure below, noise (4 levels of Gaussian white noise) is plotted along the x-axis, and the various assessment scores (Depth, Image Quality, Naturalness and Viewing Experience) are plotted along the y-axes in separate figures. Disparity is varied as a function of the camera-base distance (CBD) and plotted as three separate graphs in each figure (0, 4, and 8 cm). Note that a CBD of 0 cm equals the monoscopic viewing condition. The results demonstrate that, first, Image Quality, Naturalness and Viewing Experience appear to be equally (and predictably) sensitive to the introduction of an image artifact – in this case noise level. However, only Naturalness and Viewing Experience also show a differentiation between the monoscopic condition on the one hand and the two stereoscopic conditions on the other. Image Quality, although consistently sensitive to image artifacts, does not account for the added value that stereoscopic modes of viewing bring to the total visual experience. Whereas the results for Viewing Experience are a little more erratic, Naturalness consistently shows a stereo-advantage, in addition to sensitivity to image artifacts.

Although traditionally image quality has been suggested to be a sufficient overall fidelity metric, published literature as well as the findings from our experiments suggest that this concept does not adequately address the added value of depth introduced by the stereoscopic 3D effect. Our experimental work has been aimed at investigating the sensitivity of potential alternatives to image quality as an indication of “overall degree of the excellence of the image” [7]. The robust finding emerging throughout our work, including the experiment discussed in this paper, is that the naturalness concept is more likely to be sensitive to the 3D effect than image quality.

Our findings are crucial for a fair evaluation of 3D displays against 2D benchmarks, and to improve our understanding of 2D versus 3D tradeoffs. Stereoscopic display technology frequently employs techniques for image separation (e.g., temporal multiplexing, spatial multiplexing) which come at a cost of monoscopic image quality. In order to make a convincing case for stereoscopic displays, one needs to be able to quantify what the added value of a 3D stereoscopic image is, over and above the monoscopic rendition of the same image, even when the image quality of each monoscopic image has been compromised in order to render the image in stereo.

In order to make a fair comparison, we need a quality assessment metric that is sensitive to both ‘traditional’ image distortions and artifacts, as well as the perceptual cost and/or benefit associated with stereoscopic, rather than monoscopic, image presentation. We believe that the naturalness concept offers significant promise in this respect.



ACKNOWLEDGEMENT

Wijnand IJsselsteijn gratefully acknowledges support from Philips Research as well as the European Commission through the MUTED (http://www.muted3d.eu/) and HELIUM3D (http://www.helium3d.eu/) projects. Work reported here has been performed in collaboration with Marc Lambooij, André Kuijsters, Frans van Vugt, and Ingrid Heynderickx.

REFERENCES

[1] J. Cutting, and P. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in W. Epstein & S. Rogers (eds), *Perception of Space and Motion*, Academic Press, San Diego, CA, pp. 69-117, 1995.

[2] T. Okoshi, "Three dimensional displays," *Proceedings of the IEEE, vol. 68*, pp. 548-564, 1980.

[3] S. Pastoor, "3-D displays: A review of current technologies," *Displays, vol. 17*, no. 2, pp. 100-110, 1997.

[4] I. Sexton, and P. Surman, “Stereoscopic and autostereoscopic display systems,” *IEEE Signal Processing Magazine, vol. 16*, no. 3, pp. 85-99, 1999.

[5] W.A. IJsselsteijn, “*Presence in Depth*,” Ph.D. thesis, Eindhoven University of Technology, The Netherlands.

[6] J. A. J. Roufs, “Perceptual image quality: Concept and measurement,” *Philips Journal of Research 47*, pp. 35–62, 1992.

[7] P.G. Engeldrum, “*Psychometric Scaling: a Toolkit for Imaging Systems Development*,” Imcotek Press, Winchester, USA, 2002.

[8] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV,” *Proceedings of the SPIE, 5291*, pp 93-104, 2004.

[9] W. Tam, L. Stelmach, and P. Corriveau, “Psychovisual aspects of viewing stereoscopic video sequences,” *Proceedings of the SPIE, 3295*, pp. 226–235, 1998.

[10] A. Berthold, “The influence of blur on the perceived quality and sensation of depth of 2D and stereo images,” Technical report, ATR Human Information Processing Research Laboratories, 1997.

[11] J. Freeman and S. Avons, “Focus group exploration of presence through advanced broadcast services,” *Proceedings of the SPIE, 3959*, pp. 530–539, 2000.

[12] P.J.H. Seuntiëns, L.M.J. Meesters and W.A. IJsselsteijn, “Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation,” *ACM Transactions on Applied Perception, 3*(2), pp. 95-109, 2006.

[13] W.A. IJsselsteijn, D.G. Bouwhuis, J. Freeman, and H. de Ridder, “Presence as an experiential metric for 3-D display evaluation,” *SID Digest 2002*, pp. 252-255, 2002.

[14] P. Seuntiëns, I. Heynderickx, and W.A. IJsselsteijn, “Capturing the added value of 3D-TV: Viewing experience and naturalness of stereoscopic images,” *The Journal of Imaging Science and Technology, 52*(2), 2008.

Wijnand A. IJsselsteijn is Associate Professor at the Human-Technology Interaction group of Eindhoven University of Technology in The Netherlands. He has a background in psychology and artificial intelligence, with an MSc in cognitive neuropsychology from Utrecht University, and a PhD in media psychology/HCI from Eindhoven University of Technology on the topic of telepresence. His current research interests include social digital media, immersive and stereoscopic media technology, and digital gaming. His focus is on conceptualizing and measuring human experiences, including social perception and self-perception, in relation to these advanced media. Wijnand is significantly involved in various industry and government sponsored projects on awareness systems, enriched communication media, digital games, and autostereoscopic display technologies. He is associate director of the Media, Interface, and Network Design labs, and co-founder and co-director of the Game Experience Lab (http://www.gamexplab.nl), and director of the 3D/ice Lab at Eindhoven University of Technology. He has published over 120 peer-reviewed journal and conference papers, and co-edited five books.

Three-Dimensional Video Capture and Analysis

Cha Zhang, Microsoft Research, USA

chazhang@microsoft.com

Three-dimensional movies have re-gained a lot of interest recently, with companies like Walt Disney and DreamWorks Animation all investing heavily in making 3D films. While it is relatively easy to render stereoscopic images from graphical models, for many other applications, such as 3D TV broadcasting, free viewpoint 3D video and 3D teleconferencing, it is necessary to capture 3D video contents using camera arrays. In this short article, we briefly review techniques required for 3D video capture and analysis, which serve as the front end for many 3D video related applications.

Considering that most 3D displays today (including 3D films) need only two slightly different views to be sent to the left and right eyes of the user, the simplest 3D video capturing front end is a stereo camera pair. To create satisfactory 3D images, one needs to be careful on a few things. First, the cameras need to be synchronized. It used to be that camera synchronization has to be conducted through a common external trigger. Nowadays, when the number of cameras is small, one can simply daisy chain a few 1394 FireWire cameras on a common bus, and these cameras will be automatically synchronized. The second important thing is camera calibration and image rectification [1]. For human eyes to perceive 3D contents comfortably, it is necessary that the two video streams are rectified. Rectification is a transformation process that project images from multiple cameras onto a common imaging plane, such that the epipolar lines are aligned horizontally. In other words, when a point in the 3D scene is projected to the two cameras, the projected pixels must lie on the same horizontal line. Image rectification requires careful calibration of the cameras. The most popular camera calibration algorithm today is invented by Zhang in [2]. Bouguet had a nice MATLAB implementation of the same algorithm [3], which also has a routine to rectify an image pair.

The data format for stereoscopic displays is not limited to image pairs. As part of the European Information Society Technologies (IST) project “Advanced Three-Dimensional Television System Technologies” (ATTEST), Fehn et al. [4] proposed to use image plus depth as the capture and transmission format for stereoscopic videos, which is now the default 3D video format for some 3D display manufactures such as Philips. Image plus depth is an evolutionary format that has low overhead in terms of bitrate (because the depth map can be compressed as a grayscale image) and is backward compatible to 2D video standards. The depth map can be manually created for legacy videos, or computed from a stereo camera pair, or directly captured by depth sensors. The main challenge in using the image plus depth format is that the left and right views to be displayed still need to be synthesized. For scenes with a lot of occlusions, such a task is non-trivial since some of the occluded regions in the synthesized views are not visible in the 2D image. Various algorithms have been proposed in literature for this hole-filling challenge [5]. Fortunately, for most users, this is not a concern as it is dealt internally by the 3D display manufacturers.

It is worth spending a few additional words on depth sensors. Triangulation and time-of-flight are two of the most popular mechanisms for depth sensing. In triangulation, a stripe pattern is projected onto the scene, which is captured by a camera positioned at a distance from the projector. To avoid contaminating the color and texture of the scene, the projector can periodically switch among a few stripe patterns in order to create a uniform illumination by temporal averaging. Another possibility is to use invisible spectrum lights such as infrared light sources for the stripe pattern. Time-of-flight depth sensors can be roughly divided into two main categories: pulsed wave and continuous modulated wave. Pulsed wave sensors measure the time of delay directly, while continuous modulated wave sensors measure the phase shift between the emitted and received laser beams to determine the scene depth. One example of pulsed wave sensors is the 3D terrestrial laser scanner systems manufactured by Riegl (http://www.riegl.com/). Continuous modulated wave sensors include SwissRanger (http://www.swissranger.ch/), ZCam Depth Camera from 3DV Systems (http://www.3dvsystems.com/), etc.

One limitation of many stereoscopic 3D displays today is the lack of motion parallax. Indeed, in the real world the relative position of objects changes as we move our head. To replicate that experience, we need not only to provide different views to each eye, but the view has to change as the viewer's position changes. Alternatively, one may like to have some interactive control over the virtual viewpoint where the images are synthesized. Another possibility is when one has a multi-view autostereoscopic 3D display, which displays multiple views in order to increase the range of sweep spot for 3D perception. These multiple views need to be synthesized from a given camera array (often with smaller number of cameras than the number of views). To fulfill these tasks, one needs to synthesize novel views from a wide range of possible viewing positions, which creates a set of new challenges for 3D video capture and analysis.

The state-of-the-art approach to creating free viewpoint 3D video is to use large scale camera arrays [6][7], or to use a small scale camera array with prior knowledge of the objects in the scene [8], or to use hybrid depth camera and regular camera arrays [9]. The cameras still need to be synchronized by external triggers. Large camera arrays typically involve huge amount of data, which are typically handled by many computers or dedicated video capture/compression card. Calibration of large camera arrays is a very tedious task, but in principle they can still be handled by Zhang's method mentioned above. Another interesting issue is color calibration. Color inconsistency across cameras may cause incorrect view-dependent color variation during rendering. The problem is very interesting but rarely explored in literature, except the recent work by Joshi et al. [10].

The main challenge in free viewpoint 3D video is view synthesis, namely, how to render the virtual views specified by the user from the array of images captured. This is far from a solved problem, although there has been an extensive literature on this topic, commonly referred as image-based rendering. Since a detailed review of image-based rendering is out of the scope of this article, we refer the readers to [11] and [11] for two detailed survey papers of the topic.

In summary, although 3D video capture requires careful consideration of issues such as geometry/color calibration and data storage, it is largely an engineering problem, and there have been many existing systems ranging from 2 cameras to over a hundred cameras reported in literature. On the other hand, analyzing the captured content, such as hole filling, depth reconstruction and novel view synthesis, is still far from mature. For Multimedia Communication researchers, in addition to the traditional issues of 3D video compression and network delivery, the interplay between these traditional topics and 3D video analysis could be a very interesting topic to work upon [12].

Reference

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[2] Z. Zhang, "A flexible new technique for camera calibration", *IEEE Trans. on PAMI*, 22(11):1330-1334, 2000.

[3] J.-Y. Bouguet, Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/.

[4] C. Fehn, P. Kauff, M. Op De Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek and I. Sexton, "An Evolutionary and Optimised Approach on 3D-TV", *Proc. of International Broadcast Conference*, 2002.

[5] C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)", *Proc. of VIIP 2003*.

[6] B. Wilburn, M. Smulski, H.-H. K. Lee, M. Horowitz, "The light field video camera", *Proc. Media Processors* 2002, SPIE Electronic Imaging, 2002.

[7] M. Tanimoto, "Overview of free viewpoint television", *EURASIP Signal Process. Image Commun.*, 21(6), 454-461, 2006.

[8] J. Carranza, C. Theobalt, M. Magnor and H.-P. Seidel, "Free-viewpoint video of human actors", *Proc. SIGGRAPH*, 2003.

[9] E. Tola, C. Zhang, Q. Cai and Z. Zhang, "Virtual view generation with a hybrid camera array", *Tech. report, CVLAB-REPORT-2009-001*, EPFL.

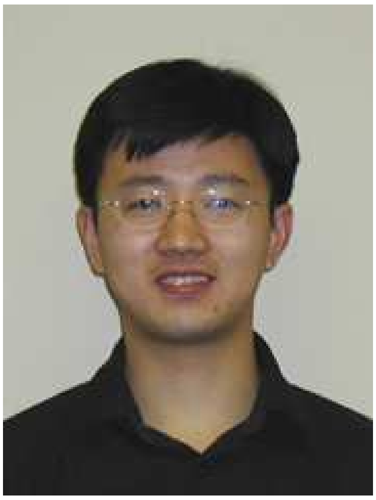
[10] N. Joshi, B. Wilburn, V. Vaish, M. Levoy and M. Horowitz, "Automatic color calibration for large camera arrays", *Tech. report, CS2005-0821*, UCSD CSE.

[11] H.-Y. Shum, S. B. Kang and S.-C. Chan, "Survey of image-based representations and compression techniques", *IEEE Trans. CSVT*, 13(11), 1020-1037, 2003.

[12] C. Zhang and T. Chen, "A survey on

image-based rendering – representation, sampling and compression”, *EURASIP Signal Process. Image Commun.*, 19(1), 1-28, 2004.

[13] D. Florencio and C. Zhang, “Multiview video Compression and Streaming Based on Predicted Viewer Position”, *Proc. ICASSP 2009*.



Cha Zhang received his B.S. and M.S. degrees from Tsinghua University, Beijing, China in 1998 and 2000, respectively, both in Electronic Engineering, and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University, in 2004. He then joined Microsoft Research as a researcher in the Communication and Collaboration Systems Group.

Dr. Zhang was the Publicity Chair for the International Packet Video Workshop in 2002, the Probram Co-Chair for the first Immersive Telecommunication Conference in 2007, the Steering Committee Co-Chair for the second Immersive Telecommunication Conference in 2009, and the Program Co-Chair for ACM Workshop on Media Data Integration (in conjunction with ACM Multimedia) 2009. He is an Associate Editor for the Journal of Distance Education Technologies and on the Editorial Board of the ICST Trans. on Immersive Telecommunications. He served as a Technical Program Committee member for numerous conferences such as ACM Multimedia, CVPR, ICCV, ECCV, ICME, ICPR, ICWL, etc. He won the Best Paper Award at ICME 2007.

Dr. Zhang’s current research interest include immersive telecommunication, audio/video processing for multimedia collaboration, pattern recognition and machine learning, etc. Previously, he has also worked on image-based rendering (compression / sampling / rendering), 3D model retrieval, active learning, peer-to-peer networking, automated lecture rooms, etc.

3D Visual Content Compression for Communications

Karsten Müller,FraunhoferHeinrich Hertz Institute, Germany
Karsten.mueller@hhi.fraunhofer.de

The recent popularity of 3D media is driven by research and development projects worldwide, targeting all aspects from 3D content creation and capturing, format representation, coding and transmission as well as new 3D display technologies and extensive user studies. Here, efficient compression technologies are developed that also consider these aspects: 3D production generates different types of content, like natural content, recorded by multiple cameras or synthetic computer-generated content. Transmission networks have different conditions and requirements for stationary as well as mobile devices. Finally, different end user devices and displays are entering the market, like stereoscopic and auto-stereoscopic displays, as well as multi-view displays. From the compression point of view, visual data requires most of the data rate and different formats with appropriate compression methods are currently investigated. These visual methods can be clustered into vision-based approaches for natural content, and (computer) graphics-based methods for synthetic content.

A straight-forward representation for natural content is the use of N camera views as N color videos. For this, compression methods derive technology from 2D video coding, where spatial correlations in each frame, as well as temporal correlations within the video sequence are exploited. In multi-view video coding (MVC) [1], also correlations between neighboring cameras are considered and thus multi-view video can be coded more efficiently than coding each view individually.

The first types of 3D displays, currently hitting the market, are stereoscopic displays. Therefore, special emphasis has been given to 2-view or stereoscopic video. Here, three main coding approaches are investigated. The first approach considers conventional stereo video and applies individual coding of both views as well as multi-view video coding. Although the latter provides better coding efficiency, some mobile devices require reduced decoding complexity, such that individual coding might be considered. The second approach considers data reduction prior to coding. This method is called mixed resolution coding and is based on the binocular suppression theorem. This states, that the overall visual perception is close to that of the original high image resolution of both views, if one original view and one low-pass filtered view (i.e. the upsampled smaller view) is presented to the viewer. The third coding approach is based on one original view and associated per-pixel depth data. The second view is generated from this data after decoding. Usually, depth data can be compressed better than color data, therefore coding gains are expected in comparison to the 2 view-color-only approaches. This video+depth method has another advantage: The baseline or virtual eye distance can be varied and thus adapted to the display and/or adjusted by the user to change the depth impression. However, this method also has disadvantages, namely the initial creation/provision of depth data by limited-capability range cameras or error-prone depth estimation. Also, areas only visible in the second view are not covered. All this may lead to a disturbed synthesized second view. Currently, studies are carried out about the suitability of each method, considering all aspects of the 3D content delivery chain.

Both approaches, multi-view video coding and stereo video coding, still have limitations, such that newer 3D video coding approaches focus on advanced multi-view formats. Here, the most general format is multi-view video + depth (MVD) [2]. It combines sparse multi-view color data with depth information and allows rendering infinitely dense intermediate views similar for any kind of autostereoscopic N-view displays. The methods, described above are a subset of this format. For possible better data compression, also variants of the MVD format are considered, like layered depth video (LDV). Here, one full view and depth is considered with additional residual color and depth information from neighboring views. The latter covers the dissoccluded areas in these views. For these advanced formats, coding approaches will be considered with respect to best compression capability, but also for down-scalability and backward compatibility to existing approaches, like stereo video coding. The new approaches are more than simply extending existing methods: Classical video coding is based on rate-distortion optimization, where the best compression at the best possible quality is sought. This objective quality is measured by pixel-wise mean squared error (MSE), comparing the lossy decoded and original picture. In MVD or LDV, new intermediate views are synthesized, where no original views are available. Therefore, a classical MSE is not suited for synthesized images. Therefore, new objective quality measures are required. Also, depth data compression has to be evaluated by the quality of intermediate color views, not by comparison with the uncompressed depth data. Further research is required for finding the optimal bit rate distribution between color and depth data, such that a number of open questions need to be addressed.

A completely different branch of 3D coding approaches was developed for synthetic content. This content is represented by color and scene geometry data in the most general case and may contain a number of further attributes, like reflection models. The most important, color and geometry data, have there analogies in natural scene content representation, i.e. color and depth data. In graphics-based approaches, 3D scene color data is treated as multi-texture or multi-view video data, similar to the methods, described above. In its most general form, 3D scene geometry is treated as dynamic 3D wireframes with 3D points or vertices and associated connectivity. The latter tells which points are connected and the most widely used is a triangular connectivity, where the wireframe consists of planar triangles. For temporal changes, the wireframe is “animated”, i.e. the spatial positions of the vertices may change over time. Coding of such scene geometry is carried out by coding the vertex positions and connectivity of the first frame or static mesh (e.g. by 3D Mesh Compression), followed by coding of the dynamic part, which consists of the vertex position changes or 3D motion vectors of the following frames [3]. Here, the connectivity is assumed constant and thus only needs to be coded for the first frame. For this, methods like dynamic 3D or frame-animated mesh compression have been introduced. Spatial correlation between motion vectors of neighboring vertices, as well as temporal correlation between subsequent meshes are exploited. Further special formats for 3D synthetic content exist, where the scene geometry takes a special form due to prior knowledge of scene objects, like terrain data and human faces/bodies. Examples for this are mesh grid and facial/bone-based animation respectively.

Both approaches, Computer Vision and Computer Graphics based coding, exist in parallel. In future work, more emphasis will be put on interchangeability to provide a continuum between Computer Vision and Computer Graphics also for coding methods and to allow easy format conversion between both coded representations.

References

[1] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding”, Doc. N9978, Hannover, Germany, July 2008.

[2] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, “Multi-view Video plus Depth Representation and Coding”, Proc. IEEE International Conference on Image Processing (ICIP'07), San Antonio, TX, USA, pp. 201-204, Sept. 2007.

[3] K. Müller, P. Merkle, and T. Wiegand, “Compressing 3D Visual Content”, IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 58-65, November 2007.

Karsten Müller received the Dipl.-Ing. degree in Electrical Engineering and Dr.-Ing. degree from the Technical University of Berlin, Germany, in 1997 and 2006 respectively.

In 1993 he spent one year studying Electronics and Communication Engineering at Napier University of Edinburgh/Scotland, including a half-year working period at Integrated Communication Systems Inc. in Westwick/England.

In 1996 he joint the Heinrich-Hertz-Institute (HHI) Berlin, Image Processing Department, where he is working on projects focused on motion and disparity estimation, incomplete 3D surface representation and coding, combined 2D and 3D object description, viewpoint synthesis and 3D scene reconstruction.

He has been actively involved in MPEG activities, standardizing the multi-view description for MPEG-7 and view-dependent multi-texturing methods and dynamic mesh compression for the 3D scene representation in MPEG-4 AFX (Animation Framework eXtension). He contributed to the multi-view coding in MPEG-4 MVC (Multi-View Video Coding) and co-chairs the MPEG 3D Video activities.

In recent projects he was involved in research and development of traffic surveillance systems and visualization of multiple-view video, object segmentation and tracking and interactive user navigation in 3D environments. Currently, he is a Project Manager for European projects in the field of 3D video technology and multimedia content description.

He is Senior Member of the IEEE (M'98-SM'07).

Interactive 3D Online Applications

Irene Cheng, University of Alberta, Canada
lin@cs.ualberta.ca

Abstract

Interactive 3D content has become an integral part in online applications. Examples can be found in Second Life, Google 3D City scene and Microsoft 3D reconstruction using multiple 2D photos. One advantage attributable to 3D content is the extra dimension, which provides better visual communication than 2D to represent structures and explain concepts. More importantly, human computer interaction makes the user feel engaged and in control, and thus obtain higher satisfaction. The increasing popularity of Blackberry, iPhone and cell-phones makes us constantly anticipate what main stream interactive 3D online applications will be available next on mobile devices. These promises also lead us to think what new strategies should be employed to overcome obstacles such as competing bandwidth, signal interference and limited processing power on mobile devices. Despite the advancements in the last two decades, providing interactive 3D online content in high quality without jeopardizing interactivity is still a challenge. This article reviews current and potential trend, the issues involved, as well as the compression and transmission strategies designed to address these challenges.

Keywords: Interactive 3D, Visual Communication, Online Applications, 3D compression and transmission, Level-of-detail, Perceptual Quality

1. Introduction

Spontaneous system response and high quality display are two main user expectations from interactive 3D online applications. In the old days, due to slow rendering, insufficient bandwidth and low processing power, using interactive 3D content was not possible in online applications. In order to support fast real time rendering, display quality had been compromised, where coarse 3D models mapped with low quality texture were used. With the advancement in technology, the qualities of 3D graphics and animations have improved significantly in the last ten years, particularly in the online entertainment industry. 3D content provides vibrant and realistic sensation, which is more appealing than still images. Augmented reality incorporating 3D objects in video streams enhances human perception by strengthening the depth perspective. Virtual 3D scene presented in a panoramic setting brings the viewers to a totally immersive environment. Moreover, the emerging 3D television and display (3DTV) technology has changed the landscape relating to how digital content is delivered, presented and appreciated by human observers. While some 3DTV technology still uses eye-wear to simulate immersion, more advanced stereoscopic displays have been launched providing multiple and free-view point entertainment without using eye-wear. This trend will have great impact on home entertainment.

Visual communication no doubt has evolved from the traditional 2D media to a 3D virtual world and beyond. Although there is a general acceptance of interactive 3D content and, to a certain extent, an addiction to online games, the capacities and benefits of interactive 3D technology are still not fully explored as they should have been. The goal of this article is to inspire thoughts to develop more effective systems and robust algorithms to support interactive 3D online applications.

The rest of this paper is organized as follows; Section 2 reviews some interactive 3D online applications; Section 3 looks at some 3D compression and transmission strategies; Section 4 highlights the emerging development in 3DTV; and Section 5 gives the summary.

2. Evolution of Visual Communication towards Interactive 3D

2.1 Interactive 3D Online Learning, Testing and Authoring

Online education, which includes K-12, university degrees and diplomas, has been used to complement classroom teaching. Other than making education more accessible, the adaption of interactive 3D content in online education aims to provide more semantically and visually appealing information that cannot be provided by using traditional media. For example, instead of using text and image in a multiple choice format to ask a question, interactive 3D can be embedded in Math and Chemistry questions –

students drag and drop the numbered 3D objects into different baskets in order to make the sum equal in each basket (Figure 1 & 2). Students pick the appropriate atoms from the Periodic table and put bonds between them to construct molecules (Figure 3). By manipulating and visualizing the molecules in 3D space, students can understand the molecular structures better.

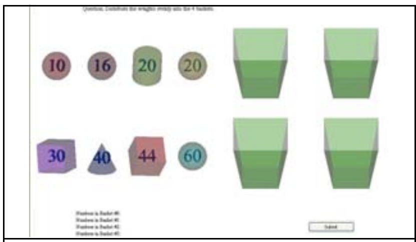


Figure 1: A 3D Math item: before distributing the numbered objects

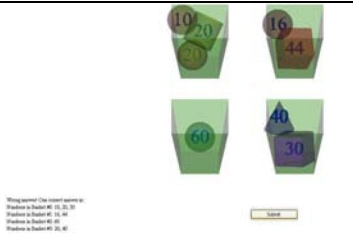
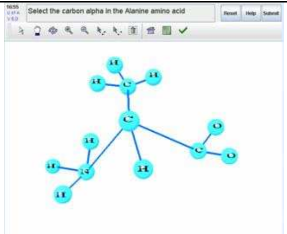


Figure 2: A 3D Math item: after distributing the numbered objects



Select the carbon alpha in the Alanine amino acid

Figure 3: Construct an interactive 3D molecules

Figure 4: An interactive 3D education game to test a Physic topic – Moment.

Figure 5: 3D video – Augmented reality combining real and virtual characters.

Figure 6: Interactive 3D content on a cell phone.

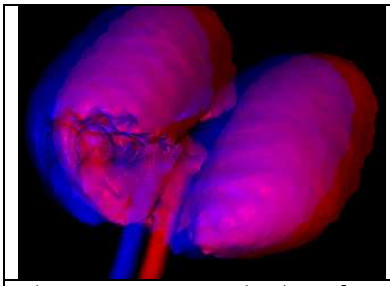


Figure 7: Stereoscopic view of a lung implemented for visualization using red-blue glasses.

Figure 8: Animated 3D character driven by music.

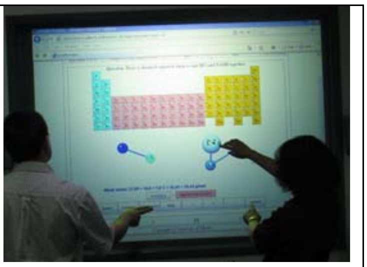


Figure 9: 3D content manipulated on a multi-touch display.

Research studies find that interactive 3D content can induce student engagement and improve learning performance [12]. Interactive 3D content can also be used to test different cognitive skills, such as music [14]. Moreover, 3D online content can support adaptive testing and effective student modeling, enabling students to work at their own pace and obtain personal guidance depending on their learning abilities. A student's skill level can be evaluated using Item Response Theory (IRT) [19]. Other benefits associated with interactive 3D content includes: Automatic estimation of the difficulty level of a question item using a Parameters Based Model [15], and scoring partial marks using a Graph Edit Algorithm [40].

There is a major concern relating to providing interactive 3D online content in education. While multiple choice questions can be easily created using a standard template, the implementation of interactive 3D question items can be unattainable by teachers who do not have sufficient programming skills. This necessitates the provision of an easy to use authoring tool, as well as an efficient client-server communication setup to support online adaptive testing. Such support is essential in order to secure a smooth transition from traditional teaching to interactive 3D online education.

2.2 Online Games, Home Entertainment and Mobile Applications

3D online games have become a part of life for most from teens to adults. New 3D content, better graphics quality and more interactive animations continue to be the targets of game developers. In fact the education community has also adopted the role-play scenario (edutainment) in order to engage students in learning [32]. An example of using a game to test a student's physics skill is shown in Figure 4 (applying moments on a beam to balance opposite forces by catching the randomly falling snow balls, each of which corresponds to a weight). Instead of spending time on non-academic games, the goal of edutainment is to make use of the engaging and rewarding factors of interactive 3D content, so that students can learn while enjoying playing online games.

Online games provide home entertainment, which provides far more user interactivity than video-on-demand. Recent trend shows that the delivery of 3D digital content will soon be extended to a larger scale from the Internet-connected PCs to household TVs. Instead of sitting in a cinema, 3D movies can be viewed on TV screens at home. Not only virtual reality, but also augmented reality [20] can be delivered to the audience (Figure 5).

In the age of globalization and mobility, the discussion of online applications cannot be completed without including mobile devices. The high usage of mobile devices by all age groups has pushed wireless networks to a new era. Subsequent to the launch of music on iPod and text messaging on Blackberry, researchers and developers now focus on downloading streamlined versions of Internet applications on mobiles. An example of using interactive 3D on a cell phone to test English is shown in Figure 6.

A major consideration in providing interactive 3D mobile content is the speed of real time delivery and rendering. User interactivity can be facilitated by pre-fetching or caching some basic data. However, this may not be feasible for some applications, e.g. adaptive testing in education, where each response from the mobile has to be assessed on the server which then selects the next appropriate question and communicates it to the mobile device. This can jeopardize the interactivity drastically.

2.3 Computer Aided Diagnosis (CAD), Surgical Planning and Training

For online shopping, manipulating a product in 3D instead of seeing 2D pictures provides consumers better information to help make shopping decision. Interactive 3D visualization has also gained increasing attention in medical research, and has become a part of many state-of-the-art CAD procedures, such as Tele-health and Tele-surgery. Detection and segmentation of anatomical 3D structures and abnormalities, e.g. brain tumors and TB cavities, are examples of algorithms used for screening false positives and helping the clinicians to prioritize their limited resources. Instead of displaying 3D objects on a flat computer screen, a new approach based on stereoscopic viewing is introduced to improve the clarity of visualizing abnormal structures (Figure 7) [2]. Due to the shortage of medical personnel, remote discussion enables specialists with different expertise participate in surgical planning collaboratively, as well as medical students attend lab sessions online. Interactive 3D content provides a virtual operation room for such collaboration to take place.

2.4 Other Interactive 3D Online Applications

Interactive 3D online applications can be designed for single user operation or multi-users collaboration. Collaborative Web3D allows multiple participations to complete a task together [11]. Collaboration is not restricted to artists crafting 3D artwork. It can involve users participating in different types of projects in education, business and medicine.

From a multimodal perspective, 3D characters animated by sound or music (Figure 8) [36] can be extended to an online collaborative environment. Another potential development is to extend the current emerging touch-based display technology, e.g. Apple iPhone and Microsoft Surface, for remote collaboration. An example of collaborative multi-touch in education is shown in Figure 9 [25].

In order to supply 3D content, the size of 3D repositories has increased enormously, which in turn has inspired the study of 3D objects similarity match and retrieval. Instead of comparing the entire object which is computationally expensive, a compact representation of the 3D object (skeleton) is often used for matching (Figure 14) [41].

3. Compression and Transmission Strategies for Interactive 3D Content

In order to optimize available resources, such as bandwidth and network reliability, to support interactive 3D online applications, many compression and transmission strategies have been proposed in the literature. We first look at some common issues encountered in reliable networks, before discussing the additional challenges in unreliable networks.

3.1 Reliable Networks

A main challenge in data transmission is limited resources, in particular bandwidth. Although fiber optics and broadband cables have improved the effectiveness of information communication, the increasing demand for more high quality multimedia content still outweighs the capacities of high speed networks. As a result, many state-of-the-art compression schemes were proposed. For interactive 3D content, there are three components to consider: 3D mesh, surface material and motion data. Depending on the data type, different compression and transmission schemes have to be applied accordingly. The question to address is what proportion of the available bandwidth should be allocated to each type of data, which leads to the evaluation of perceptual quality on 3D models.

Raw 3D geometry captured by scanning devices is often too dense to be transmitted for real time applications. Meshes at different level-of-detail (LOD) are generated and appropriate levels are chosen taking both application requirements and available resources into consideration. The traditional metrics used to evaluate visual qualities of compressed 3D meshes are based on geometric measurement; *i.e.*, by measuring the deviation of the surface between the original and the coarse mesh. In the discrete LOD approach, a number of compressed meshes are pre-generated. Even if the bandwidth capacity can support a mesh in between two pre-generated levels, the server can only deliver the one with a lower quality. In the continuous LOD approach, coarse meshes can be generated on the fly by removing vertices or faces one at a time, making full use of the available bandwidth. Many robust LOD generation algorithms have been developed in the last fifteen years. During transmission, vertices are grouped into hierarchical layers based on their importance to visual quality. The bottom layer gives the basic shape of the 3D object while higher layers refine the object progressively using the available bandwidth [30].

Perceptually motivated evaluation metrics emerged when researchers realized that geometry based compression techniques have become saturated [24]. Considerable efforts were expended on verifying geometric deviation with perceptual evaluation experiments in order to achieve higher visual fidelity of 3D display. The metrics can be categorized into view-dependent and view-independent. For interactive 3D applications, the latter is preferred because not only one view but the quality of the entire object is judged by the user. Instead of applying the same resolution to the entire rendered view, progressive compression [4, 33] applies declining resolution from the centre to the peripheral. The centre is the fovea or the target of the viewer. Compression dictated by foveal attention was introduced in [5, 38].

JPEG and JPEG2000 are efficient compression schemes for the surface texture on 3D objects. However, when both geometry and texture need to be transmitted, the challenge is how to allocate the bandwidth so that the optimal visual quality can be achieved. How the human visual system (HVS) responds to decreasing mesh and texture resolution was analyzed in [28]. Psychophysical experimental findings show that increasing mesh resolution does not improve visual quality after reaching a certain threshold, while increasing texture resolution continues to improve the overall visual quality of a 3D object. A perceptually optimized mathematical model was then proposed based on this finding to allocate the available bandwidth between geometry and texture data [9].

Thus far we have assumed a fixed bandwidth throughout the transmission. But in practice, network bandwidths tend to fluctuate. One approach to secure a consistent quality of service (QoS) is to actively monitor the available bandwidth and adapt the data to a target size that can be transmitted within a given time limit [44]. The algorithm uses a probabilistic model to estimate the optimal amount of bandwidth that can be sacrificed to estimate the current network throughput, while using the remaining bandwidth to transmit application data. Since low texture data has bigger adverse effect on visual quality [21], a Harmonic Time Compensation Algorithm was discussed in [8], which adaptively adjusts the texture quality in subsequent packets if unfavorable throughput prevent the transmission of data with satisfactory quality in earlier packets.

As Google, Microsoft and other influential research labs have started to expand their horizon to include large scale outdoor 3D scenes, maintaining mesh connectivity becomes computationally expensive and in some cases difficult to construct accurately from raw scanned data. A point cloud representation is seen as more manageable and effective. In general, transmission algorithms applicable to meshes can be applied to point clouds with the luxury of throwing away connectivity. Point clouds can also be divided into different LODs [29] and transmitted hierarchically starting from the bottom layer. The coarse point cloud is progressively refined using the available bandwidth [13]. The QSplat technique compresses the data further by grouping neighboring 3D points into clusters based on a pre-defined Euclidean distance threshold [34, 35]. While the advantage is fast rendering, the original model cannot be fully reconstructed and the degraded visual quality can be visible when refining a coarse cluster.

Image-based instead of model-based rendering also attracts a lot of attention for panoramic outdoor scenes in 3D graphics because of its potential of bypassing the difficulties of both modeling real objects and of rendering photorealistic scene illuminations. A cylindrical warping procedure using multiple photos to visualize 3D effect is shown in Figure 10. The multi-resolution features in JPEG2000 are used to support the different zoomed levels [45].

Figure 10: An example of cylindrical warping using 20 photos to generate more vivid 3D depth perspective.

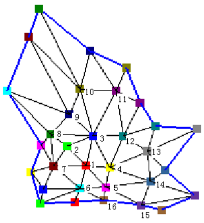


Figure 11: Adjacent vertices are distributed to different transmission packets.

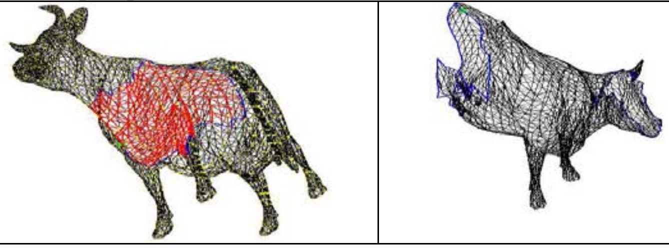


Figure 12: Following the valence driven algorithm, (left) the colored patch shows the neighbourhood after 400 vertices are distributed to different packets. (Right) shows the partly reconstructed mesh after 3,000 vertices are retrieved from the packets.

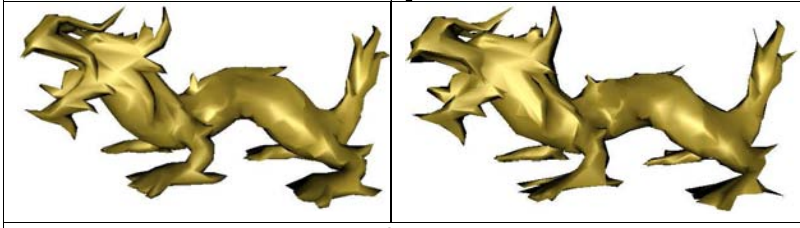


Figure 13: Visual quality is satisfactorily preserved by the curvature-driven probabilistic transmission strategy: (left) original 3D model and (right) reconstructed model using a network packet loss rate of 55%.

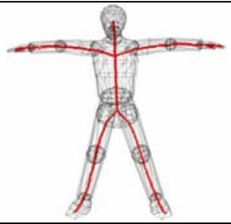


Figure 14: A 3D model with its medial axis (skeleton).

3.2 Unreliable Networks and Mobile Devices

While reliable networks guarantee delivery, supporting interactive 3D applications on wireless networks has to overcome the problem of packet loss which is often bursty. Some wireless protocols proposed in the last decade [3] include Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Indirect-TCP (I-TCP), and so on. For wireless networks, where packet loss occurs as a result of unreliable links and route changes, the TCP strategy leads to further delays and degradation in transmission quality. Approaches for robust transmission of mesh over wireless networks have been proposed [1, 10], which assume that some parts of the mesh can be transmitted without loss, allowing progressive refinement to give good results. Other researchers developed models for packet loss probability to protect critical data via retransmission [23]. A view-independent joint texture-mesh transmission strategy was proposed in [28]. Different from other approaches, this strategy does not need to guarantee delivery of certain packets in order to make other packets useful. The basic concept is to distribute neighboring vertices (mesh data) and texels (texture data) into separate packets to reduce the probability of losing adjacent vertices or texels creating big voids, so that the lost patch can be reconstructed by interpolation. An extension of this interleaving strategy to irregular 3D meshes is described in [18], which applies a valence driven algorithm [37] to traverse the mesh from a starting face. Adjacent vertices are stripped off following a circular path from the starting face and distributed into separate packets (Figure 11). An example of encoding and decoding a cow object is shown in Figure 12. The visual quality of the transmitted 3D objects is further improved by probabilistically setting higher protection priority to high curvature surface structures [17] (Figure 13).

3.3 Geometric and Perceptual Metrics – Just-Noticeable-Difference (JND) in 3D

Visual Quality Metrics (VQM) [31], Structural Similarity (SSIM) index [39, 43] and other quality evaluation methods on image and video were reviewed in the May issue of the MMTC E-Letter. This article focuses on visual quality on 3D textured mapped meshes.

The impacts of perceptual factors on mesh simplification techniques have been discussed by many researchers [27, 42, 46]. Scale-space filtering is a technique commonly used on 2D data and was first applied to generate 3D level-of-details incorporating human perceptual factors [6]. This work differs from the traditional geometry-based approaches, in applying JND to further reduce redundant data, making transmission more efficient. Instead of assuming that visual quality decreases as the number of vertices decreases, experimental results showed that degradation followed a step like function – multiple vertices need to be removed before a noticeable visual impact was generated [7]. The work was extended by considering relative change measured from the medial axis of the object, instead of using absolute change to quantify the magnitude of visual stimuli [16]. The evidence of associating relative change with 3D shape perception was verified through psychophysical experiments [22].

4. Emerging Interactive 3D Technology

While interactive 3D brings the virtual world closer to reality than traditional 2D displays, a longer term goal is to provide more prominent 3D sensation, e.g. touch and smell, and support more flexible viewing angles.

4.1 3D Stereoscopic, Multiple and Free Viewpoint

Emerging 3D television and display (3DTV) technology has changed the landscape relating to how digital content is presented and how human observers appreciate digital content. Interactive 3D content has become more visually appealing by using stereoscopic effects. Unlike traditional displays, 3DTV stereoscopic techniques stimulate a viewing perception in which virtual objects and scenes appear to stand out from the 2D display. Instead of sitting in front of desktop computers, home users are soon likely to enjoy a wide range of interactive stereoscopic 3D online entertainment on TV screens. When entertainment is delivered to homes, it is necessary to allow family members view together. To achieve this goal, it is necessary to evolve communication technology to support the transmission of not only single view stereo video [26] but also multi-view 3D stereoscopic data, as well as to support mobile services. As the prices of such multi-view displays go down, the demand from the general public will certainly increases.

5. Summary

The development of interactive 3D techniques is far from mature. There are many interesting research topics to study. To date, there is still inadequate study to address the perceptual quality optimization issue taking the transmission of mesh, texture and motion data into consideration. By highlighting the potential of interactive 3D online applications, and reviewing the compression and transmission strategies in this article, the goal is to attract more novel ideas to advance visual communication to a new milestone.

6. References

1. G. Alregib, Y. Altunbasak and J. Rossignac, “Error-Resilient Transmission of 3D Models,” ACM Trans. On Graphics, vol.24, no.2, 182-208, 2005.
2. H. Azari, I. Cheng and A. Basu, “Stereo 3D Mouse (S3D-Mouse): Measuring Ground Truth for Medical Data in a Virtual 3D Space,” IEEE Engineering in Medicine and Biology Conference, 2009.
3. A. Bakre and B.R. Badrinath, “I-TCP: Indirect TCP for mobile hosts,” Int. Conf. Distributed Computing Systems, 136-143, 1995.
4. P. Baudisch, D. DeCarlo, A. Duchowski and W. Geisler, “Focusing on the Essential: Considering Attention in Display Design,” ACM Communication, vol.46, no.3, 60-66, Mar 2003.
5. A. Basu and K. Wiebe, “Video-Conferencing using Spatially Varying Sensing with Multiple and Moving Foveas,” Proc. IEEE Int. Conf. on Pattern Recognition, 1994.
6. I. Cheng and P. Boulanger, “Feature Extraction on 3D TexMesh Using Scale-space Analysis and Perceptual Evaluation,” IEEE Trans. On circuits and Systems for Video Technology, vol. 15, no.10, 2005.
7. I. Cheng and P. Boulanger, “A 3D Perceptual Metric Using Just-Noticeable-Difference,” EUROGRAPHICS 2005.
8. I. Cheng and P. Boulanger, “Adaptive Online Transmission of 3D TexMesh Using Scale-space and Visual Perception Analysis,” IEEE Trans. On Multimedia, vol.8, no.3, 550-563, 2006.
9. I. Cheng and A. Basu, “Perceptually Optimized 3D Transmission over Wireless network,” IEEE Trans. On Multimedia, vol.9, no.2, 386-396, Feb 2007.
10. Z. Chen, B. Bodenheimer and J.F. Barnes, “Robust Transmission of 3D Geometry Over Wireless Networks,” Proc. Int. Conf. on 3D Web Technology.
11. I. Cheng, M. Bates and A. Basu, “Collaborative Online 3D Editing,” Siggraph Webgraphics, 27-31, 2003.
12. I. Cheng, A. Basu and R. Goebel, “Interactive Multimedia for Adaptive Online Education,” IEEE Multimedia Magazine, Jan-Mar, 16-24, 2009.
13. F. Chen, I. Cheng and A. Basu, “Packet Loss Resilient Transmission of 3D Point Clouds,” IEEE Int. Conf. on Multimedia, 2009.
14. I. Cheng, C. Kerr and W. Bischof, “Assessing Rhythm Recognition skills in a Multimedia Environment,” IEEE Int. Conf. on Multimedia, 2008.
15. I. Cheng, R. Shen and A. Basu, “An Algorithm for Automatic Difficulty Level Estimation of Multimedia Mathematical Test Items,” IEEE Int. Conf. on Adv. Learning Technologies, 2008.
16. I. Cheng, R. Shen, X. Yang and P. Boulanger, “Perceptual Analysis of Level-of-Detail: The JND Approach,” IEEE Int Sym. On Multimedia, 533-540, 2006.
17. I. Cheng, L. Ying and K. Daniilidis, “A Curvature-Driven Probabilistic Strategy for Transmission of Arbitrary 3D Meshes over Unreliable Networks,” Int. Sym. On 3DPVT, 2008.
18. I. Cheng, L. Ying, K. Daniilidis and A. Basu, “Robust and Scalable Transmission of Arbitrary 3D Models over Wireless Networks,” EURASIP Journal on Image and Video Processing IVP/890482, 2008.
19. S.E. Embretson and S.P. Reise, “Item Response Theory for Psychologists”, Lawrence Erlbaum Associates, 2000.
20. F. Chen, I. Cheng and A. Basu, “Integrating 3D Point Clouds with Multi-Viewpoint Video,” IEEE 3DTV Conf. 2009.
21. D. Koller, M. Turitzin, M. Levoy, M. Tarini, g. Croccia, P. Cignoni and R. Scopigno, “Protected Interactive 3D Graphics via Remote Rendering,” ACM Trans. On Graphics, vol.23, no.3, 695-703, 2004.
22. E. Johnson, “Systematic Distortions of Shape from Stereopsis,” Vision Research, 31: 1351-1360, 1991.
23. K.V. Lee and S.T. Chanson, “Packet Loss Probability for Real-Time Wireless Communications,” IEEE Trans. On Vehicular Technology, vol.51, no.6, 1569-1575, 2002.
24. D. Luebke, “A Developer’s Survey of Polygonal Simplification Agorithms,” IEEE Computer Graphics and Applications, vol.21, no.3, 24-35, May/Jun 2001.

25. D. Michel, A. Argyros, D. Grammrenos, X. Zabulis and T. Sarmis, “building a Multi-Touch Display Based on Computer Vision Techniques,” IAPR Conf. on Machine Vision Applications (MVA) 2009.
26. W. Matusik and H. Pfister, “3DTV: A Scalable System for Real-time Acquisition, Transmission and Autostereoscopic display of Dynamic scenes,” ACM Trans. On Graphics, 23(3): 814-824, 2004.
27. C. O’Sullivan, S. Howlett, R. Mcdonnell, Y. Moryan and K. O’Conor, “Perceptually Adaptive Graphics,” EUROGRAPHICS, State of the Art Report, 2004.
28. Y. Pan, I. Cheng and A. Basu, “Quality Metric for Approximating Subjective Evaluation of 3D Objects,” IEEE Trans. On Multimedia, vol.7, no.2, 269-279, 2005.
29. M. Pauly, L. Kobbelt nd M. Gross, “Point-based Multiscale Surface Representation,” ACM Trans. On Graphics, 25(2):177-193, 2006.
30. R. Pajarola and J. Rossignac, “Compressed Progressive Meshes,” IEEE Trans. On Visualization and Computer Graphics, vol.6, no.1, 79-93, 2000.
31. M.H. Pinson and S. Wolf, “A new Standardized method for Objectively Measuring Video Quality,” IEEE Trans. On Broadcasting, vol.50, 312-322, 2004.
32. S. Rodriguez, I. Cheng and A. Basu, “Multimedia Games for Learning and Testing Physics,” IEEE Int. Conf. on Multimedia, 1838-1841, 2007.
33. M. Reddy, “Perceptually Optimized 3D Graphics,” Applied Perception, vol.21, 68-75, Sep/Oct 2001.
34. S. rusinkiewicz and M. Levoy, “QSplat: A Multiresolution point rendering system for large meshes,” Proc. ACM SIGGRAPH 2000, 343-352.
35. S. Rusinkiewicz and M. Levoy, “Streaming QSplat: A Viewer for Networked Visualization of Large, Dense Models,” ACM Symposium on Interactive 3D Graphics, 2001, 63-83.
36. R. Taylor, P. Boulanger and P. Olivier, “An Audience-Participatory Performance System,” Int. Sym. On Smart Graphics, 2009.
37. C. Touma and C. Gotsman, “Triangle Mesh Compression,” Graphics Interface, 26-34, 1998.
38. K. Wiebe and A. Basu, “Improving Image and Video Transmission over ATM,” Proc. IEEE Int. Conf. on Pattern Recognition, 1996.
39. Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structured Similarity,” IEEE Trans. On Image Processing, vol.13, no.4, 600-612, 2004.
40. G. Wu and I. Cheng, “An Interactive 3D Environment for Computer Based Education,” IEEE Int. Conf. on Multimedia, 1834-1837, 2007.
41. T. Wang and I. Cheng, “Generation of Unit-Width Curve Skeletons based on Valence Driven Spatial Median (VDSM),” Int. Sym. On Visual Computing (ISVC), 2008
42. N. Williams, D. Luebke, J. Cohen, M. Kelley and B. Schubert, “Perceptually Guided Simplification of Lit, Textured Meshes,” ACM Sym. On Interactive 3D Graphics, 113-121, 2003.
43. Z. Wang, E.P. Simoncelli and A.C. Bovik, “Multi-Scale Structural Similarity for Image Quality Assessment,” IEEE Conf. Sig. Syst., Comput., Asilomar, 2003.
44. Y. Yu, I. Cheng and A. Basu, “Optimal Adaptive Bandwidth Monitring for QoS based Retrieval,” IEEE Trans. On Multimedia, vol.5, no.3, 466-472, 2003.
45. P. Zanuttigh, G.M. Cortelazzo and L. MacDonald, “3D Online Multimedia: Processing, Transmission and Visualization,” Chapter 13, World Scientific Publishing Co Pte Ltd., 2009.
46. E. Zhang and G. Turk, “Visibility-Guided Simplification,” Proc. Conf. Visualization, 2002.

Irene Cheng is the scientific director of the iCORE-NSERC Multimedia Research Centre and an adjunct faculty in the Department of Computing Science, University of Alberta, Canada. Her research interests include incorporating human perception – Just-

Noticeable-Difference – with scale-space analysis following psychophysical methodology, to improve 3D simplification and transmission techniques. In 2006-7, she designed a multimedia framework incorporating innovative interactive 3D item types for education, and led the development team to implement the Computer Reinforced Online Multimedia Education (CROME) System (http://crome.cs.ualberta.ca). She received an Alumni Recognition Award in 2008 from the University of Alberta for her R&D contributions. She has received, or been offered, many scholarships and fellowships from NSERC, iCORE and others. Currently, she is involved with research in 3DTV and perceptually motivated technologies in multimedia, high-dimensional visualization and transmission, particularly for medical and education applications. She serves as program chair, and as TC member in numerous conferences. She is an IEEE Senior Member, the Chair of the IEEE Northern Canada Section, Engineering in Medicine and Biology Society Chapter, Board Member of the IEEE System, Man and Cybernetics (SMC) Society, Human Perception in Vision, Graphics and Multimedia TC, and Voting Member of the IEEE Communication Society, Multimedia Communications TC. Her organized workshops include ICCV-S3DV and ACM MM-iWAM in 2009, CVPR-S3D in 2008 and ISVC special tracks from 2006-09. She is a guest editor for IJDMB special issue on Adv. in 3DTV. She has over eighty publications including two books, journals and conference papers.

Editor's Selected Paper Recommendation

H. Kushwaha, Y. Xing, R. Chandramouli, and H. Heffes, "Reliable multimedia transmission over cognitive radio networks using fountain codes," *Proceedings of the IEEE*, vol. 96, no. 1, pp. 155-165, Jan. 2008

Deploying wireless real-time multimedia applications encounters many challenges, such as considerable bandwidth requirement, stringent delay constraints, and reliable link transmission. The end-to-end final received multimedia quality is mainly dominated by the link capacity and link quality. However, radio spectrum is a limited resource and a major portion of the spectrum has been already allocated. Recently, Federal Communications Commission (FCC) proceedings propose the notion of secondary spectrum access to improve spectrum efficiency by allowing dynamic access to the vacant parts of the spectrum owned by the primary user (PU) to become accessible temporarily by a secondary user (SU). This dynamic access of spectrum by secondary users can be realized by adopting cognitive radios. Cognitive radio is a wireless communication paradigm in which either the network or the wireless node itself changes particular transmission or reception parameters to execute its tasks efficiently without interfering with the licensed users or other cognitive radios.

In cognitive radio networks, a multimedia transmission application can select a set of subchannels (SC) from different PUs to establish a communication link. Note that a link can consist of multiple different SCs at different frequencies. The advantages are (1) to have higher overall throughput contributed from multiple SCs; (2) to add path diversity from different SCs in the network to facilitate error protection during streaming the multimedia content. On the other hand, the inherent problem of distributing media over multiple SCs is the coordination required between the SCs. In order to efficiently utilize the limited spectral resource, the multimedia applications need to carefully coordinate the packet scheduling. Besides, the applications also need to overcome the interference from PUs. Thus, error protection mechanism needs to be brought into scenes to ensure the successful content reconstruction.

The aforementioned two problems occurred in cognitive radio networks can be alleviated by adopting fountain codes [1][2]. An ideal fountain code has the property that an infinite amount of parity packets from k information packets can be generated and the entire source message can be reliably reconstructed from any n ($n>k$) received encoding packets. In this paper, the authors propose to apply fountain codes to multimedia transmission over cognitive networks scenario to achieve two goals simultaneously. First, the transmitters can spread the packets over different SCs with no need of coordination between them since the receiver cares about the amount (instead of order or type) of correctly received packets. Secondly, it works as a channel code to resist the packet loss effect from PU interference and other channel conditions.

More specifically, the authors study the problem of optimizing the spectral resources in the secondary usage scenario for multimedia applications with respect to the number of available SCs and primary user occupancy of the SCs. Analysis and simulations detail the optimal number of SCs and optimal fountain codes to achieve optimal spectral efficiency.

Researches on multimedia over cognitive radio networks are gaining more and more interests. Authors in [3] study the channel selection problem for multimedia over cognitive networks. The scenario of multimedia multicast over cognitive networks is studied in [4].

[1] J. Wagner, J. Chakareski, and P. Frossard, "Streaming of scalable video from multiple servers using rateless codes", in *IEEE ICME*, Jul. 2006.
[2] H. Jenkac; T. Stockhammer, and W. Xu, "Asynchronous and reliable on-demand media broadcast", *IEEE Network Magazine*, vol. 20, no. 2, pp.14-20, March/April 2006..
[3] H. P. Shiang and M. van der Schaar, "Queuing-based dynamic channel selection for heterogeneous multimedia applications over cognitive radio networks," *IEEE Trans. Multimedia*, Vol. 10, no.5, pp. 896~909, Aug. 2008
[4] D. Hu; S. Mao; J.H. Reed, "On video multicast in cognitive radio networks", *INFOCOM 2009*. pp.2222–2230.

** The Column Editor recommending this paper is Dr. Guan-Ming Su*

Focused Technology Advances Series

Distributed Algorithm Design for Network Optimization Problems with Coupled Objectives

Jianwei Huang, the Chinese University of Hong Kong, China
jwhuang@ie.cuhk.edu.hk

In order to optimize the performance of various communication and networking systems, we often model it as a Network Utility Maximization (NUM) problem [1]. In the NUM formulation, we associate each network entity (user) with a utility function representing its performance/satisfaction/happiness, and thus maximizing the total network utility is equivalent to maximizing the network performance. Due to the distributed and heterogeneous nature of today's network, one of the key challenges is how to design distributed algorithms that can achieve the global optimal NUM solution.

The difficulty in distributed algorithm design often lies in the coupling nature of the NUM problem. There exist two kinds of coupling: coupled constraints (e.g., limited total network resource) and coupled objective functions (e.g., due to interferences or collisions among wireless nodes). The coupled constraints can be decomposed using the dual or primal decompositions which have been widely used and studied (see [1] [2] and numerous references therein). The coupled objective functions, however, are more difficult to deal with and less studied. One recent innovative approach of dealing with coupled objective functions is to use "consistency pricing" [3], which is mainly suitable for strictly convex NUM formulations, involves significant amount of message passing among users, and requires updating the variables using small step-sizes which often leads to slow convergence.

Here we establish a new framework of designing distributed algorithm for NUM with coupled objective functions. The key idea is to "reverse-engineer" the algorithm based on the KKT condition of the NUM problem, which involves "localizing the global objective function" and designing suitable physical message passing mechanism. A key feature of the corresponding algorithm is that it does involve use any small step-sizes, thus typically has much faster and more robust convergence compared with the consistence pricing approach. Moreover, the algorithm also works for certain cases where the strictly convexity of the NUM problem can not be directly proved using the standard optimization approach.

To be more concrete, we will consider a network of a set of K users. Each user k has a coupled utility function $U_k(x_k,\mathbf{x}_{-k})$ that depends on both his own local decision variable x_k and other users' decision variables $\mathbf{x}_{-k}=(x_1,\ldots,x_{k-1},x_{k+1},x_K)$. An example of the utility function would be the data rate achieved by a wireless user, which only depends on its own transmission power but also on the interference (and other other users' transmission power). User k can choose x_k from a feasible set $X_k=[X_k^{min},X_k^{max}]$. Denote $\mathbf{x}$ as the vector containing all users' decision variables and $\mathbf{X}$ as the set containing all users' feasible sets.

We want to solve the following NUM problem:

$$\max_{\mathbf{x}\in\mathbf{X}} \sum_{k=1}^{K} U_k\left(x_k,\mathbf{x}_{-k}\right). \tag{1}$$

We assume that the utility function $U_k(x_k,\mathbf{x}_{-k})$ is increasing and strictly concave in x_k but not necessarily concave in $\mathbf{x}$. This means that Problem (1) is not necessarily a strictly concave maximization problem in variables $\mathbf{x}$, and thus might have more than one local/global optimal solution. Solving such a problem is difficult in general even through centralized computation. Our target is to solve the problem in a distributed fashion under certain technical conditions.

The starting point of our algorithm design is the Karush–Kuhn–Tucker (KKT) conditions of Problem (1), which are the necessary conditions for a global optimal solution. A KKT point $\mathbf{x}^*$ needs to satisfy the following for each k,

$$\frac{\partial U_k\left(x_k^*,\mathbf{x}_{-k}^*\right)}{\partial x_k^*}+\sum_{j\neq k}\frac{\partial U_j\left(x_k^*,\mathbf{x}_{-k}^*\right)}{\partial x_k^*}=\lambda_k^*-\mu_k^*, \tag{2}$$

$$\lambda_k^*(x_k^*-X_k^{\max})=0,\ \mu_k^*(X_k^{\min}-x_k^*)=0, \tag{3}$$

$$\lambda_k^*,\mu_k^*\geq 0. \tag{4}$$

Our task is to design a distributed algorithm that is guaranteed to converge to the KKT set. If the KKT set is a singleton set, then the corresponding element must be the unique global optimal solution and our algorithm converges to such a solution.

We design distributed algorithm by letting each user solve a local optimization problem that is defined based on local observation and limited message passing among users. We will first reverse engineer the "local objective function" for each user such that the KKT conditions can be satisfied if all users maximize their local objective functions properly.

In particular, each user k can choose a local objective function $Y_k(x_k,\boldsymbol{x}_{-k})$ that is increasing and strictly concave in x_k and satisfies

$$\frac{\partial Y_k\left(x_k,\mathbf{x}_{-k}\right)}{\partial x_k}=\frac{\partial U_k\left(x_k,\mathbf{x}_{-k}\right)}{\partial x_k}+\sum_{j\neq k}\frac{\partial U_j\left(x_k,\mathbf{x}_{-k}\right)}{\partial x_k}. \tag{5}$$

Thus if user k solves the following local optimization problem for the optimal choice of $\boldsymbol{x}^*_{-k}$,

$$\max_{x_k\in X_k} Y_k\left(x_k,\mathbf{x}^*_{-k}\right). \tag{6}$$

The obtained optimal solution x^*_k satisfies (2)-(4).

One seemingly obvious candidate of $Y_k(x_k,\boldsymbol{x}_{-k})$ is the total network utility,

$$Y_k\left(x_k,\mathbf{x}_{-k}\right)=\sum_{j=1}^{K}U_j\left(x_k,\mathbf{x}_{-k}\right). \tag{7}$$

This choice, however, is typically not practical since it is often too strong to assume that a user knows the exact forms of other users' utility functions and decision variables.

A more practical approach is to construct a local function that only depends on user k's local observations of the network and some limited information passing among users. These messages are functions of users' current decisions variables x_k's. Since users are iteratively update x_k's aiming at finding the global optimal solution x_k's, the values of the messages will also change over time. This means that the function $Y_k(x_k,\boldsymbol{x}_{-k})$ can also be time-varying, i.e., user k solves different version of local optimization problem (6) at different time instances until the system converges. We assume that each user k can announce a locally computable message m_k as a function of $\boldsymbol{x}$. Define $\boldsymbol{m}_{-k}=(m_1,\ldots,m_{k-1},m_{k+1},m_K)$, and thus $\boldsymbol{m}$ is the vector of all users' messages. We will construct a new local function $Z_k(x_k,\boldsymbol{x}_{-k},\boldsymbol{m}_{-k})$ such that

$$\frac{\partial Z_k\left(x_k,\mathbf{x}_{-k},\mathbf{m}_{-k}\right)}{\partial x_k}=\frac{\partial U_k\left(x_k,\mathbf{x}_{-k}\right)}{\partial x_k}+\sum_{j\neq k}\frac{\partial U_j\left(x_k,\mathbf{x}_{-k}\right)}{l x_k}.$$

The key requirement is that user k can calculate function $Z_k(x_k,\boldsymbol{x}_{-k},\boldsymbol{m}_{-k})$ based on its local measurement (which can depend on $\boldsymbol{x}_{-k}$) and the messages announced by other users ($\boldsymbol{m}_{-k}$).

User k's local optimization problem is to calculate

$$x_k=q_k\left(\mathbf{x}_{-k},\mathbf{m}_{-k}\right)=\arg\max_{\overline{x}_k\in X_k} Z_k\left(\overline{x}_k,\mathbf{x}_{-k},\mathbf{m}_{-k}\right).$$

For message m_k, user k need to update it as

$$m_k=f_k(\mathbf{x}).$$

The functions q_k and f_k are derived based on specific problem structures and examples can be found in [4] [5].

We are now ready to describe the asynchronous distributed and algorithm. For each user k, let $T_{k,x}$ and $T_{k,m}$ be two unbounded sets of positive time instances at which user k updates its local decision variable x_k and message m_k, respectively.

<u>Asynchronous and Distributed Algorithm</u>

```
Let time $t = 0$.
for all user $k$ do
    Randomly initialize $x_k(0)$ and $m_k(0)$.
end for
repeat
    $t = t + 1$.
    for all user $k$ do
        if $t \in T_{k,x}$ then
            $x_k(t + 1) = q_k(\boldsymbol{x}_{-k}(t),\boldsymbol{m}_{-k}(t))$.
        end if
        if $t \in T_{k,m}$ then
            Announce $m_k(t + 1) = f_k(\boldsymbol{x}(t))$.
        end if
    end for
until $(\boldsymbol{x}(t), \boldsymbol{m}(t))$ converge
```

A key feature of the algorithm is that there are no small step-sizes involved, which is unlike the consistency price approach in [3]. The stepsize-free design typically leads to much faster convergence, but makes the analysis (e.g., proving convergence) more difficult. Mathematical tools that are useful for the convergence analysis include contraction mapping [6], monotone mapping [6], supermodular game [4], potential game [8], and congestion game [9-11].

REFERENCES

[1] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," Proceedings of the IEEE, vol. 95, no. 1, pp. 255–312, 2007.

[2] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," IEEE Journal on Selected Areas of Communications, vol. 24, no. 8, pp. 1439 – 1451, 2006.

[3] C. W. Tan, D. P. Palomar, and M. Chiang, "Distributed optimization of coupled systems with applications to network utility maximization," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, May 2006.

[4] J. Huang, R. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," IEEE Journal of Selected Areas in Communications, vol. 24, pp. 1074–1084, May. 2006.

[5] A. H. Mohsenian-Rad, J. Huang, M. Chiang, and V. W. S. Wong, "Utility-optimal random access: Reduced complexity, fast convergence, and robust performance," IEEE Transactions on Wireless Communications, vol. 8, no. 2, pp. 898–911, February 2009.

[6] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods. Prentice Hall, 1989.

[7] D. Monderer and L. Shapley, "Potential games," Games and Economic Behavior, vol. 14, pp. 124–143, 1996.

[8] G. Scutari, S. Barbarossa, and D. Palomar, "Potential Games: A Framework for Vector Power Control Problems with Coupled Constraints," ICASSP 2006, vol. 4, pp. 241–244, 2006.

[9] I. Milchtaich, "Congestion games with player-specific payoff functions," Games and Economic Behavior, 1996.

[10] B. Vocking, "Congestion games: Optimization in competition," Proceedings of the 2nd Algorithms and Complexity in Durham Workshop, Jan 2006.

[11] M. Liu and Y. Wu, "Spectum sharing as congestion games," Allerton conference, 2008.

Jianwei Huang is an Assistant Professor in the Department of Information Engineering at the Chinese University of Hong Kong. He received B.S. in Electrical Engineering from Southeast University (Nanjing, Jiangsu, China) in 2000, M.S. and Ph.D. in Electrical and Computer Engineering from Northwestern University (Evanston, IL, USA) in 2003 and 2005, respectively. He worked as a Postdoc Research Associate in the Department of Electrical Engineering at Princeton University during 2005-2007, and as a summer intern in Motorola (Arlington Heights, IL, USA) in 2004 and 2005.

Dr. Huang conducts research in the area of nonlinear optimization and game theoretical analysis of communication networks, with current focus on network economics, cognitive radio networks, broadband communication networks, and multimedia over wireless. He is the recipient of the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2009 and Walter P. Murphy Fellowship at Northwestern University in 2001.

Dr. Huang is an Associate Editor of (Elsevier) Journal of Computer & Electrical Engineering, the Lead Guest Editor of IEEE Journal of Selected Areas in Communications special issue on "Game Theory in Communication Systems", the Lead Guest Editor of Journal of Communications special issue on "Cognitive-Radio Enabled Communications and Networking", the Lead Guest Editor of Journal of Advances in Multimedia special issue on "Collaboration and Optimization in Multimedia Communications", and a Guest Editor of Journal of Advances in Multimedia special issue on "Cross-layer Optimized Wireless Multimedia Communications". He serves as the TPC co-chair of the International Conference on Game Theory for Networks 2009, the TPC co-chair of IEEE Globecom Wireless Communications Symposium 2010, and TPC member of various conferences such as INFOCOM, MobiHoc, ICC, and WCNC.

MMTC COMMUNICATIONS & EVENTS

Call for Papers of Selected Journal Special Issues

Journal of Communications
Special Issue on Multimedia Computing and Communications

Guest Editors: Fan Zhai, Homer Chen, Thomas Stockhammer, Touradj Ebrahimi
Paper Submission deadline: **September 1**, 2009
Target Publishing Issue: 2nd Quarter, 2010
CfP Weblink: http://www.academypublisher.com/jcm/si/jcmsi_mcc.html

International Journal of Digital Multimedia Broadcasting
Special Issue on Video Analysis, Abstraction, and Retrieval: Techniques and Applications

Guest Editors: Jungong Han, Ling Shao, Peter H.N. de With, Ling Guan
Paper Submission deadline: **September 1**, 2009
Target Publishing Issue: March 1, 2010
Cfp Weblink: http://www.hindawi.com/journals/ijdmb/si/varta.html

Ad Hoc Networks (Elsevier)
Special Issue on Multimedia Ad Hoc and Sensor Networks

Guest Editors: Tommaso Melodia, Martin Reisslein
Paper Submission deadline: **December 15**, 2009
Target Publishing Issue: 3rd Quarter, 2010
CfP Weblink: http://www.elsevierscitech.com/pdfs/cfp_adhoc0709.pdf

Multimedia System Journal
Special Issue on Wireless Multimedia Transmission Technology and Application

Guest Editors: Gabriel-Miro Muntean, Pascal Frossard, Haohong Wang, Yan Zhang, Liang Zhou
Paper Submission deadline: **Jan. 15**, 2010
Target Publishing Issue: 4th Quarter, 2010
Cfp Weblink: http://www.ifi.uio.no/MMSJ/CFP_SI_Wireless_MMSJ.pdf

Call for Papers of Selected Conferences

ICC 2010

Website:	http://www.ieee-icc.org/2010/
Dates:	May 23-27, 2010
Location:	Cape Town, South Africa
Submission Due:	**Sept. 10, 2009**

E-Letter Editorial Board

EDITOR-IN-CHIEF

EDITOR

MMTC Officers

CHAIR

VICE CHAIRS

SECRETARY