Yucel Altunbasak,
John Apostolopoulos,
Philip A. Chou, and B.H. Juang

Realizing the Vision of Immersive Communication

We have witnessed tremendous progress in the buildup of communication networks in recent decades, driven by technologies that afford broad bandwidth and mobility to the user. While broadband and wireless technologies have greatly improved the delivery of information, packet-based connectivity promises the needed flexibility and efficiency for full dimensional remote collaboration. Delivery focuses on transporting information from one point to another, while full dimensional collaboration encompasses experiences in generating, granting, and receiving the information, together with all the parties involved, as in a face-to-face interaction session. By full-dimensional collaboration, we mean sensory aspects such as sight, sound, and touch as well as enhancements to cognitive capacity. We are at the onset of a revolution in telecommunication, from delivery of information to full-dimensional collaboration, as evidenced by the industry's strong push for the notion of telepresence (e.g., HP's HALO system or Cisco's Telepresence). In fact, there is an expectation of change of mode in telecommunication, from telephony and telepresence, towards the ultimate goal of immersive communication (IC), in which users communicate, interact, and collaborate over a distance with immersive and lifelike experiences.

These expectations come from two major driving forces: 1) the user's desire to be able to communicate and collaborate with other people in a most natural manner with uncompromised quality, and 2) the complexity of information involved in collaboration that needs intelligent support. Some historical perspective on telephony may illustrate the point. The history of the telephone literally began with the invention of the microphone amid the growing popularity of the telegraph in the second half of the 19th century. A.G. Bell's demonstration of being able to "talk with electricity" was achieved on 10 March 1876, the day that marked the birth of the telephone, which quickly surpassed the telegraph in its widespread use. The benefits of being able to communicate and interact (albeit only with voice at the time) with someone in a natural manner far outweighed those benefits that might have come with a telegraph system even with increased capacity. The importance of a natural sensory interface obviously extends to other modalities and the quality of communications. For example, few sound systems can be sold today without at least stereo reproduction and few visual displays without high definition. As frequent users of current teleconferencing systems, the quality of which is less than desirable, we long for the day when unimpaired, uninterrupted, and unfatiguing communication with our remote collaborators becomes attainable. Today, with the growth in technologies, we, as users of telecom systems, ask to realize the vision of "seeing is believing" and "sharing means understanding," to cope with the modern sophistication in human collaborative activities.

Thus, the vision of IC is to enable natural experiences and interactions among people, objects, and environments as if they were colocated, although they may be geographically distributed. Furthermore, in this time of high sensitivity for environmental and energy issues, the telecom industry has an increasing role to play in the new "green economy." A number of recent studies (e.g., the "Smart 2020 Report" and others [1], [2]) have concluded that information and communication technology (ICT) could contribute to the global reduction of CO_2 emission by 15% by 2020 as well as an energy efficiency savings of €500 billion. IC is an advanced set of ICT that would lead to further realization of these benefits through the reduction of unnecessary physical travels. Additional future promising applications of IC include education, entertainment, health care, and industrial design.

OVERVIEW OF THIS SPECIAL ISSUE

This special issue's goal is disseminating information on the challenges and recent advances that the various signal processing communities are achieving towards realizing the vision of IC. The articles in this issue examine the auditory, visual, and haptic aspects of IC.

To realize the immersive audio and acoustic experience in collaborative communications, traditional acoustic and audio signal processing needs to be extended to multichannel scenarios, spawning many new technical challenges related to voice communication services infrastructure. The first article, by Huang et al., presents a survey of the development of various immersive audio schemes in concert with the vision of IC.

Rendering of spatialized audio is a key aspect of IC. It can be done through headphones or through loudspeakers. The article by Algazi and Duda is a tutorial on headphone-based spatialized audio for IC. The article covers two scenarios. In the first scenario, separate "dry" or nonreverberant sources of audio

Digital Object Identifier 10.1109/MSP.2010.939043
Date of publication: 17 December 2010

are already available (e.g., from separately mic'd recordings) and the task is to make the listener perceive the sources coming from arbitrary points in space relative to the listener. That is, the audio landscape is synthetic. In the second scenario, the audio landscape in a real environment is sampled at a particular location and reproduced so as to make the listener perceive that he or she is listening in that location. Different techniques are required in the two scenarios.

Rendering of spatialized audio on loudspeakers typically involves the production of directional sound fields. These can provide customized audio for targeted people, and reduced noise pollution or cross-talk for people nearby. Directional sound may be achieved through a number of approaches including conventional speaker arrays and ultrasonic speakers. The ultrasonic speaker has the advantage of highly directional sound projection without requiring a large speaker array. The article by Gan et al. presents an overview of ultrasonic speaker-based directional sound and how signal processing is key to achieve high-quality directional sound by overcoming a variety of problems, including current limitations in nonlinear acoustics and ultrasonic transducer technology.

Immersion and interactivity capabilities in visual communication will be greatly enhanced by recent developments in three-dimensional (3-D) sensing and computing. With 3-D representation of natural videos available, users can interactively control the viewpoint and generate new views of a dynamic scene as if each of them had a virtual camera. This is exactly equivalent to virtual reality, but with real video as opposed to synthetic views. The next two articles present two different approaches towards 3-D free-viewpoint videos. The first captures 3-D information directly whereas the second uses computational methods.

The article by Do et al. uses an emerging class of fast and low-priced cameras for measuring depth. These so-called "depth cameras" measure the time delay between transmission of a light pulse and detection of the reflected signal on an entire frame at once. However, the resolution of depth images from depth cameras is often low as compared to traditional color cameras, hence the need arises for fusing depth and color information in rendering free-viewpoint videos. The article presents methods of depth-image based rendering (DIBR) techniques as well as quantitative analysis of the rendering quality. Practical DIBR methods with general-purpose parallel computing platforms such as graphics processing units are also illustrated.

[THE VISION OF IC IS TO ENABLE NATURAL EXPERIENCES AND INTERACTIONS AMONG PEOPLE, OBJECTS, AND ENVIRONMENTS AS IF THEY WERE COLOCATED, ALTHOUGH THEY MAY BE GEOGRAPHICALLY DISTRIBUTED.]

The article by Tanimoto et al. presents a systems-approach to free-viewpoint TV (FTV). The article reviews a real-time system for FTV, including the complete chain of operation from image capture to display. FTV uses several cameras to capture images from different viewpoints. It then makes use of ray-space interpolation to synthesize any viewpoint. This article also briefly discusses international standardization efforts for FTV systems within the Moving Picture Experts Group.

The article by Cooperstock explores the communication requirements for tightly coupled activities in a "shared reality" environment, where distributed users are free to move about their physical spaces and interact with one another as if physically copresent. The article uses a distributed opera performance as one of the most demanding IC test beds. However, the discussion of video, audio, and haptics technologies along with the challenges associated with their use in IC is fairly general and can easily be extended to distributed collaborative human activities such as surgical planning, simulation training, and musical performances.

True IC involves more than just hearing and vision. Also required are haptics, comprising both the tactile sense (such as touch, pressure, and pain) and kinaesthetics (the sense of muscle movement and joint position). The last article of this special issue, by Steinbach et al., shows how haptic signals can be sampled, quantized, coded, transmitted, predicted, and used to drive a teleoperator to interact with a remote environment, and how forces or other information sensed from the environment can be fed back to the user over the network. The article emphasizes the role of perceptual modeling and psychophysics on processing and representation. At this time, IC is an incipient field; papers on taste and smell for IC are awaited by a future issue.

In this special issue, we tried to provide overviews of the state of the art in signal processing for IC. This is a rapidly emerging area with many opportunities for signal processing to have significant impact in the future. We hope these articles help to convey to the reader the excitement of the emerging area of IC.

We would like to thank the authors for their valuable contributions as well as the large number of anonymous reviewers who helped ensure the quality of this special issue. We would also like to thank the large number of authors who submitted white paper proposals for articles. We express our regret that, due to limited space and the need for balanced coverage, not all of the high-quality submissions could be included. Finally, we would like to thank Dan Schonfeld, Shih-Fu Chang, Li Deng, and Geri Krolin-Taylor for their encouragement and support in organizing this special issue.

REFERENCES

[1] The Climate Group, GeSI (Global e-Sustainability Initiative). (2008). Smart 2020: Enabling the low carbon economy in the information age, The Smart 2020 Report [Online]. Available: http://www.smart2020.org/

[2] (2010). Europe 2020: A European strategy for smart, sustainable and inclusive growth, European Commission 2020 [Online]. Available: http://ec.europa.eu/eu2020/index_en.htm

[SP]