Introduction

Oliver Schreer, Peter Kauff and Thomas Sikora

The migration of immersive media towards telecommunication applications continues to advance. Impressive progress in the field of media compression, media representation, and the larger and ever-increasing bandwidth available to the customer, will foster the introduction of these services in the future. It is widely accepted that this trend towards immersive media is going to have a strong impact on our daily life.

The ability to evoke a state of 'being there' and/or of 'being immersed' into media applications will no longer remain the domain of the flight simulators, CAVE systems, cyberspace applications, theme parks or IMAX theatres. It will arise in offices, venues and homes and it has the potential to enhance quality of life in general.

First steps in this direction have already been observed during the last few years:

- Video conferencing has become more and more attractive for various lines of business. Today video conferencing enhances distributed collaboration in an emerging global market. It is therefore regarded as a high-return investment for decision-making processes. Today, high-end videoconferencing systems already offer telepresence capabilities to achieve communication conditions as natural as possible. This business sector will benefit from the future advent of immersive systems providing improved realism of scene reproduction.
- The market of team collaboration systems grows drastically. The first synchronous collaboration tools are being sold today. They meet the demands of an increasing competition in costs, innovation, productivity and development cycles. Most of them still rely on the screen-sharing principle and suffer from a lack of natural communication between the collaborating partners. Emerging teleimmersion systems will go beyond these limitations of conventional collaborative team software. They will employ collaborative virtual environments (CVE) with intuitive interaction and communication capabilities.
- In the entertainment sector we are now beginning to see the viable economics of high-definition broadcasting of live events in sports or culture to cinemas, halls and large

3D Videocommunication — Algorithms, concepts and real-time systems in human centred communication
Edited by O. Schreer, P. Kauff and T. Sikora © 2005 John Wiley & Sons, Ltd

group venues. Applications such as e-theatres, d-cinemas, home theatres and immersive televisions are envisioned and/or being investigated by many R&D departments around the world. Television, computer games, sports arenas, live events or cinema as we know them today will inevitably develop into new immersive applications to satisfy consumer demands during the coming decades.

It is very difficult to predict developments in the field of immersive media beyond the topics discussed today. But the examples pointed out above already indicate a shift of paradigms in the way we will capture, transmit and consume media information in the future. Due to falling prices and advancing quality, large-screen displays, audio-visual 3D scene representation and intuitive human–machine interfaces will become more and more established in daily use, especially in offices and in home environments. Immersive systems will leave its experimental state and immersive portals will become ubiquitous in business and entertainment. The impact for consumers as well as for business processes and value chains will be drastic.

The development from two-dimensional (2D) towards three-dimensional (3D) audiovisual communications is generally seen as one of the key components for the envisioned applications. Scientific challenges in this field are manifold. They range from high-quality 3D analysis of audio and video and arbitrary view and sound synthesis to encoding of 3D audio and video. Understanding of real-time implementation issues, as well as system architectures and network aspects will be essential for the success of these applications. The introduction of many of these services will require new standards for the representation and coding of 3D audiovisual data. Since many of these services will change the way of how we consume and interact with media applications, it is important to take human factors research into account. The ultimate goal is to develop applications with sufficient service quality and user acceptance. The presence research community contributes to many aspects of this kind of user-centred communications.

This book presents a comprehensive overview of the principles and concepts involved in the fascinating field of 3D audiovisual communications. It offers a practical step-by-step walk through the various challenges, concepts, components and technologies involved in the development of applications and services. Researchers and students interested in the field of 3D audiovisual communications will find this book a valuable resource, covering a broad overview of the current state of the art. Practical engineers from industry will find this book useful in envisioning and building innovative applications.

The book is divided in four major parts. The first part introduces to the challenging field of 3D video communications by presenting the most important applications in this domain, namely 3D television, free view point video and immersive videoconferencing. The second part covers the theoretical aspects of 3D video and audio processing. Following the logical order of a common signal processing chain, the third part is related to 3D reproduction of audio-visual content. In the last part, several aspects of 3D data sensors are discussed.

The aim of Section I *Applications of 3D Videocommunication* is to give a comprehensive overview on the state of the art, the challenges and the potential of 3D videocommunication. This part opens with a chapter on *History of Telepresence* by W.A. IJsselsteijn. It presents the foundation and justification for this new field of research and development. A historical review describes how the term tele-presence emerged. The following chapter on *3D TV Broadcasting* by C. Fehn presents in detail one of the key applications in the field of 3D videocommunications. The history of television, the concept for a next generation television system and an elaborate description of the end-to-end stereoscopic video chain are

discussed. These new emerging technologies also have a drastic impact on content creation and postproduction. To this end the chapter on *3D in Content Creation and Post-production* by O. Grau discusses new trends and perspectives in this domain. Since the ultimate goal of 3D audiovisual communication is to provide the user with a free view point, the chapter *Free Viewpoint Systems* by M. Tanimoto describes the challenges in this area and presents first results of experimental systems. Bidirectional 3D videocommunication is embedded in the concept of immersive videoconferencing. A chapter on *Immersive Videoconferencing* by P. Kauff and O. Schreer presents the history and the current state of the art of such systems for telepresence. Visionary approaches implemented in prototypes of immersive videoconferencing systems and immersive portals are outlined and discussed.

Section II of the book addresses the question of how 3D audiovisual data may be represented and processed. Our prime goal is to provide the reader with a complete overview on all aspects related to processing of audio and video. The chapter *Fundamentals of Multiple-view Geometry* by S. Ivekovic, A. Fusiello and E. Trucco outlines the theory relevant for understanding the imaging process of a 3D scene onto a single camera. The chapter focuses on the pinhole camera model, a stereo camera system by explaining the key issues of epipolar geometry and finally the three-view geometry based on the trifocal tensor. Several aspects of rectification and reconstruction are covered as well. This chapter provides the theoretical foundation for the following chapters of the section.

Stereo analysis provides implicit depth information from few camera images of the same scene. The chapter *Stereo Analysis* by N. Atzpadin and J. Mulligan illustrates in detail current approaches in stereo processing using two or three cameras. The fundamental challenges of disparity analysis are discussed and an overview on different algorithms is presented. More precise 3D models can be generated based on multiple views of a scene or an object. The chapter *Reconstruction of Volumetric 3D Models* by P. Eisert focuses on the relevant approaches in this domain — important for many new 3D multimedia services.

A next important step in the 3D processing chain consists of rendering novel views. In Chapter 9 *View Synthesis and Rendering Methods*, R. Koch and J.-F. Evers-Senne provide a classification of existing rendering methods. The subdivision in methods without geometry information, methods with implicit and explicit geometry gives a comprehensive insight into recently developed approaches in this new field of research.

The chapter *3D Audio Capture and Analysis* by M. Schwab and P. Noll covers aspects of the 3D acquisition process of human speech. This includes echo-control, noise reduction and 3D audio source localization to support convincing rendering of audiovisual scene content.

Standardization is a key issue for interoperability of systems from different vendors. Chapter 11 *Coding and Standardization* by A. Smolic and T. Sikora outlines the basic coding strategies frequently used for audio, image and video storage and transmission. International coding standards such as ITU, MPEG-2/4 and MP3 are discussed in this context. The most recent standardization activity relevant to 3D videocommunications is the MPEG-4 AdHoc-Group work on 3D Audio/Visual. The reader is provided with a detailed overview on these activities.

Section III covers different aspects of 3D reproduction of audiovisual content. It is introduced by a chapter on *Human Factors of 3D Displays* by W.A. IJsselsteijn, P.J.H. Seuntiens and L.M.J. Meesters. The authors discuss several aspects of stereoscopic viewing. The human factors aspect is essential for the development of convincing 3D video systems. The basics of human depth perception are presented since knowledge in this domain is fundamental for stereoscopic viewing. Principles of stereoscopic reproduction and the impact on

stereoscopic image quality are discussed. The chapter *3D Displays* by S. Pastoor discusses core technologies for existing 3D displays. Aided viewing as well as autostereoscopic viewing approaches are addressed and discussed in the context of existing display prototype systems and products.

In conjunction with mixed reality applications, head-mounted displays (HMD) play an important role for visualization of virtual content in a real scene. Developments in this field are outlined in Chapter 14 on *Mixed Reality Displays* by S. Pastoor and C. Conomis. After a brief description of challenges of mixed reality displays and some aspects of human spatial vision in this field, a comprehensive overview of different technologies and systems is given.

Even in situations where vision is dominant, the auditory sense helps to analyse the environment and creates a feeling of immersion. The correct or at least plausible reproduction of spatial audio becomes an important topic. T. Sporer and S. Brix present the fundamentals of *Spatialized Audio and 3D Audio Rendering* in Chapter 15.

Section IV covers the field of active 3D data sensors. Active techniques enable the creation of very detailed 3D models with high accuracy. In Chapter 16 *Sensor-based Depth Capturing* by J.G.M. Goncalves and V. Sequeira, various active techniques for the capture of range images are outlined. The authors discuss limitations, accuracies and calibration aspects of these methods.

Chapter 17 *Tracking and User Interface for Mixed Reality* by Y. Abdeljaoued, D. Marimon; Sanjvan and T. Ebrahimi completes the book. The authors discuss the tracking of objects for the purpose of accurate registration between the real and virtual world. Furthermore, the importance of interaction technologies for the creation of convincing mixed reality systems is emphasized.