3D VISUAL COMMUNICATIONS

3D VISUAL COMMUNICATIONS

Guan-Ming Su
Dolby Labs, California, USA

Yu-Chi Lai
National Taiwan University of Science and Technology, Taiwan

Andres Kwasinski
Rochester Institute of Technology, New York, USA

Haohong Wang
TCL Research America, California, USA

A John Wiley & Sons, Ltd., Publication

This edition first published 2013

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Su, Guan-Ming.
3D visual communications / Guan-Ming Su, Yu-Chi Lai, Andres Kwasinski, Haohong Wang.
pages cm
Includes bibliographical references and index.
ISBN 978-1-119-96070-6 (cloth)
1. Multimedia communications. 2. Three-dimensional display systems. I. Lai, Yu-Chi. II. Kwasinski, Andres.
III. Wang, Haohong, 1973- IV. Title. V. Title: Three dimensional visual communications.
TK5105.15.S83 2013
006.7–dc23

2012031377

A catalogue record for this book is available from the British Library.

ISBN: 978-1-119-96070-6
Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India

Contents

Preface **ix**
About the Authors **xiii**

1 Introduction **1**
1.1 Why 3D Communications? 1
1.2 End-to-End 3D Visual Ecosystem 3
1.2.1 3D Modeling and Representation 5
1.2.2 3D Content Creation 6
1.2.3 3D Video Compression 7
1.2.4 3D Content Delivery 8
1.2.5 3D Display 9
1.2.6 3D QoE 9
1.3 3D Visual Communications 10
1.4 Challenges and Opportunities 11
References 15

2 3D Graphics and Rendering **17**
2.1 3DTV Content Processing Procedure 19
2.2 3D Scene Representation with Explicit Geometry – Geometry Based Representation 22
2.2.1 Surface Based Representation 23
2.2.2 Point Based Representation 37
2.2.3 Point Based Construction 38
2.2.4 Point Based Compression and Encoding for Transmission 38
2.2.5 Point Based Rendering: Splatting 39
2.2.6 Volumetric Representation 40
2.2.7 Volumetric Construction 40
2.2.8 Volumetric Compression and Encoding for Transmission 41
2.2.9 Volumetric Rendering 42
2.3 3D Scene Representation without Geometry – Image-Based Representation 43
2.3.1 Plenoptic Function 43
2.3.2 Single Texture Representation 46
2.3.3 Multiple Texture Representation 48
2.3.4 Image Based Animation 51

2.4 3D Scene Representation with Implicit Geometry – Depth-Image-Based Representation 51
2.4.1 History of Depth-Image-Based Representation 52
2.4.2 Fundamental Concept Depth-Image-Based Representation 53
2.4.3 Depth Construction 56
2.4.4 Depth-Image-Based Animation 57
References 57

3 3D Display Systems 63
3.1 Depth Cues and Applications to 3D Display 63
3.1.1 Monocular Depth Cues 63
3.1.2 Binocular Depth Cues 64
3.2 Stereoscopic Display 65
3.2.1 Wavelength Division (Color) Multiplexing 65
3.2.2 Polarization Multiplexing 69
3.2.3 Time Multiplexing 69
3.3 Autostereoscopic Display 71
3.3.1 Occlusion-Based Approach 71
3.3.2 Refraction-Based Approach 75
3.4 Multi-View System 78
3.4.1 Head Tracking Enabled Multi-View Display 79
3.4.2 Automultiscopic 79
3.5 Recent Advances in Hologram System Study 83
References 84

4 3D Content Creation 85
4.1 3D Scene Modeling and Creation 85
4.1.1 Geometry-Based Modeling 86
4.1.2 Image-Based Modeling 86
4.1.3 Hybrid Approaches 87
4.2 3D Content Capturing 87
4.2.1 Stereo Camera 87
4.2.2 Depth Camera 88
4.2.3 Multi-View Camera 88
4.2.4 3D Capturing with Monoscopic Camera 89
4.3 2D-to-3D Video Conversion 101
4.3.1 Automatic 2D-to-3D Conversion 103
4.3.2 Interactive 2D-to-3D Conversion 111
4.3.3 Showcase of 3D Conversion System Design 112
4.4 3D Multi-View Generation 125
References 126

5 3D Video Coding and Standards 129
5.1 Fundamentals of Video Coding 129
5.2 Two-View Stereo Video Coding 142
5.2.1 Individual View Coding 142
5.2.2 Inter-View Prediction Stereo Video Coding 143

5.3 Frame-Compatible Stereo Coding 144
5.3.1 Half-Resolution Frame-Compatible Stereo Coding 144
5.3.2 Full-Resolution Frame-Compatible Layer Approach 146
5.4 Video Plus Depth Coding 148
5.5 Multiple View Coding 156
5.6 Multi-View Video Plus Depth (MVD) Video 160
5.7 Layered Depth Video (LDV) 163
5.8 MPEG-4 BIFS and AFX 165
5.9 Free-View Point Video 166
References 167

6 Communication Networks 171
6.1 IP Networks 171
6.1.1 Packet Networks 171
6.1.2 Layered Network Protocols Architecture 172
6.2 Wireless Communications 174
6.2.1 Modulation 175
6.2.2 The Wireless Channel 177
6.2.3 Adaptive Modulation and Coding 191
6.3 Wireless Networking 193
6.4 4G Standards and Systems 193
6.4.1 Evolved Universal Terrestrial Radio Access Network (E-UTRAN) 195
6.4.2 Evolved Packet Core (EPC) 200
6.4.3 Long Term Evolution-Advance (LTE-A) 201
6.4.4 IEEE 802.16 – WiMAX 202
References 203

7 Quality of Experience 205
7.1 3D Artifacts 205
7.1.1 Fundamentals of 3D Human Visual System 205
7.1.2 Coordinate Transform for Camera and Display System 206
7.1.3 Keystone Distortion 211
7.1.4 Depth-Plane Curvature 212
7.1.5 Shear Distortion 212
7.1.6 Puppet-Theater Effect 213
7.1.7 Cardboard Effect 215
7.1.8 Asymmetries in Stereo Camera Rig 216
7.1.9 Crosstalk 217
7.1.10 Picket-Fence Effect and Lattice Artifacts 217
7.1.11 Hybrid DCT Lossy Compression Artifact 218
7.1.12 Depth Map Bleeding and Depth Ringing 219
7.1.13 Artifacts Introduced by Unreliable Communication Networks 219
7.1.14 Artifacts from New View Synthesis 219
7.1.15 Summary of 3D Artifacts 220
7.2 QoE Measurement 220
7.2.1 Subjective Evaluations 222
7.2.2 2D Image and Video QoE Measurement 226

	7.2.3 3D Video HVS Based QoE Measurement	235
	7.2.4 Postscript on Quality of Assessment	246
7.3	QoE Oriented System Design	247
	7.3.1 Focus Cues and Perceptual Distortions	247
	7.3.2 Visual Fatigue	249
	References	250
8	**3D Video over Networks**	**259**
8.1	Transmission-Induced Error	259
8.2	Error Resilience	267
8.3	Error Concealment	270
8.4	Unequal Error Protection	275
8.5	Multiple Description Coding	279
8.6	Cross-Layer Design	282
	References	286
9	**3D Applications**	**289**
9.1	Glass-Less Two-View Systems	289
	9.1.1 Spatially Multiplexed Systems	290
	9.1.2 Temporally Multiplexed Systems	290
9.2	3D Capture and Display Systems	291
9.3	Two-View Gaming Systems	294
9.4	3D Mobile	298
	9.4.1 HTC EVO 3D	298
	9.4.2 Mobile 3D Perception	299
9.5	Augmented Reality	302
	9.5.1 Medical Visualization	304
	9.5.2 Mobile Phone Applications	306
	References	309
10	**Advanced 3D Video Streaming Applications**	**313**
10.1	Rate Control in Adaptive Streaming	313
	10.1.1 Fundamentals of Rate Control	313
	10.1.2 Two-View Stereo Video Streaming	318
	10.1.3 MVC Streaming	318
	10.1.4 MVD Streaming	319
10.2	Multi-View Video View Switching	321
10.3	Peer-to-Peer 3D Video Streaming	325
10.4	3D Video Broadcasting	328
10.5	3D Video over 4G Networks	329
	References	331
Index		**335**

Preface

As the Avatar 3D movie experience swept the world in 2010, 3D visual content has become the most eye-catching spot in the consumer electronics products. This 3D visual wave has spread to 3DTV, Blu-ray, PC, mobile, and gaming industries, as the 3D visual system provides sufficient depth cues for end users to acquire better understanding of the geometric structure of the captured scenes, and nonverbal signals and cues in visual conversation. In addition, 3D visual systems enable observers to recognize the physical layout and location for each object with immersive viewing experiences and natural user interaction, which also makes it an important topic for both academic and industrial researchers.

Living in an era of widespread mobility and networking, where almost all consumer electronic devices are endpoints of the wireless/wired networks, the deployment of 3D visual representation will significantly challenge the network bandwidth as well as the computational capability of terminal points. In other words, the data volume received in an endpoint required to generate 3D views will be many times that of a single view in a 2D system, and hence the new view generation process sets a higher requirement for the endpoint's computational capability. Emerging 4G communication systems fit very well into the timing of 3D visual communications by significantly improving the bandwidth as well as introducing many new features designed specifically for high-volume data communications.

In this book, we aim to provide comprehensive coverage of major theories and practices involved in the lifecycle of a 3D visual content delivery system. The book presents technologies used in an end-to-end 3D visual communication system, including the fundamentals of 3D visual representation, the latest 3D video coding techniques, communication infrastructure and networks in 3D communications, and 3D quality of experience.

This book targets professionals involved in the research, design, and development of 3D visual coding and 3D visual transmission systems and technologies. It provides essential reading for students, engineers, and academic and industrial researchers. This book is a comprehensive reference for learning all aspects of 3D graphics and video coding, content creation and display, and communications and networking.

Organization of the book

This book is organized as three parts:

- principles of 3D visual systems: 3D graphics and rending, 3D display, and 3D content creation are all well covered

- visual communication: fundamental technologies used in 3D video coding and communication system, and the quality of experience. There are discussions on various 3D video coding formats and different communication systems, to evaluate the advantages of each system
- advances and applications of 3D visual communication

Chapter 1 overviews the whole end-to-end 3D video ecosystem, in which we cover key components in the pipeline: the 3D source coding, pre-processing, communication system, post-processing, and system-level design. We highlight the challenges and opportunities for 3D visual communication systems to give readers a big picture of the 3D visual content deployment technology, and point out which specific chapters relate to the listed advanced application scenarios.

3D scene representations are the bridging technology for the entire 3D visual pipeline from creation to visualization. Different 3D scene representations exhibit different characteristics and the selections should be chosen according to the requirement of the targeted applications. Various techniques can be categorized according to the amount of geometric information used in the 3D representation spectrum; at one extreme is the simplest form via rendering without referring to any geometry, and the other end uses geometrical description. Both extremes of the technology have their own advantages and disadvantages. Therefore, hybrid methods, rendering with implicit geometries, are proposed to combine the advantages and disadvantages of both ends of the technology spectrum to better support the needs of stereoscopic applications. In Chapter 2, a detailed discussion about three main categories for 3D scene representations is given.

In Chapter 3, we introduce the display technologies that allow the end users to perceive 3D objects. 3D displays are the direct interfaces between the virtual world and human eyes and these play an important role in reconstructing 3D scenes. We first describe the fundamentals of the human visual system (HVS) and discuss depth cues. Having this background, we introduce the simplest scenario to support stereoscopic technologies (two-view only) with aided glasses. Then, the common stereoscopic technologies without aided glasses are presented. Display technologies to support multiple views simultaneously are addressed to cover the head-tracking-enabled multi-view display, occlusion-based and reflection-based multi-view system. At the end of this chapter, we will briefly discuss the holographic system.

In Chapter 4, we look at 3D content creation methods, from 3D modeling and representation, capturing, 2D to 3D conversion and, to 3D multi-view generation. We showcase three practical examples that are adopted in industrial 3D creation process to provide a clear picture of how things work together in a real 3D creation system.

It has been observed that 3D content has significantly higher storage requirements compared to their 2D counterparts. Introducing compression technologies to reduce the required storage size and alleviate transmission bandwidth is very important for deploying 3D applications. In Chapter 5, we introduce 3D video coding and related standards. We will first cover the fundamental concepts and methods used in conventional 2D video codecs, especially the state-of-the-art H.264 compression method and the recent development of next generation video codec standards. With common coding knowledge, we first introduce two-view video coding methods which have been exploited in the past decade. Several methods, including individual two-view coding, simple inter-view prediction stereo video coding, and the latest efforts on frame-compatible stereo coding, are

presented. Research on the depth information to reconstruct the 3D scene has brought some improvements and the 3D video coding can benefit from introducing depth information into the coded bit stream. We describe how to utilize and compress the depth information in the video-plus-depth coding system. Supporting multi-view video sequence compression is an important topic as multi-view systems provide a more immersive viewing experience. We will introduce the H.264 multiple view coding (MVC) for this particular application. More advanced technologies to further reduce the bit rate for multi-view systems, such as the multi-view video plus depth coding and layered depth video coding system, are introduced. At the end of this chapter, the efforts on the 3D representation in MPEG-4, such as binary format for scenes (BIFS) and animation framework extension (AFX), are presented. The ultimate goal for 3D video system, namely, the free viewpoint system, is also briefly discussed.

In Chapter 6, we present a review of the most important topics in communication networks that are relevant to the subject matter of this book. We start by describing the main architecture of packet networks with a focus on those based on the Internet protocol (IP) networks. Here we describe the layered organization of network protocols. After this, we turn our focus to wireless communications, describing the main components of digital wireless communications systems followed by a presentation of modulation techniques, the characteristics of the wireless channels, and adaptive modulation and coding. These topics are then applied in the description of wireless networks and we conclude with a study of fourth generation (4G) cellular wireless standards and systems.

To make 3D viewing systems more competitive relative to 2D systems, the quality of experience (QoE) shown from 3D systems should provide better performance than from 2D systems. Among different 3D systems, it is also important to have a systematic way to compare and summarize the advances and assess the disadvantages. In Chapter 7, we discuss the quality of experience in 3D systems. We first present the 3D artifacts which may be induced throughout the whole content life cycle: content capture, content creation, content compression, content delivery, and content display. In the second part, we address how to measure the quality of experience for 3D systems subjectively and objectively. With those requirements in mind, we discuss the important factors to design a comfortable and high-quality 3D system.

Chapter 8 addresses the main issue encountered when transmitting 3D video over a channel: that of dealing with errors introduced during the communication process. The chapter starts by presenting the effects of transmission-induced errors following by a discussion of techniques to counter these errors, such as the error resilience, error concealment, unequal error protection, and multiple description coding. The chapter concludes with a discussion of cross-layer approaches.

Developing 3D stereoscopic applications has become really popular in the software industry. 3D stereoscopic research and applications are advancing rapidly due to the commercial need and the popularity of 3D stereoscopic products. Therefore, Chapter 9 gives a short discussion of commercially available products and technologies for application development. The discussed topics include commercially available glass-less two-view systems, depth adaptation capturing and displaying systems, two-view gaming systems, mobile 3D systems and perception, and 3D augmented reality systems.

In the final chapter, we introduce the state-of-the-art technologies for delivering compressed 3D content over communication channels. Subject to limited bandwidth

constraints in the existing communication infrastructure, the bit rate of the compressed video data needs to be controlled to fit in the allowed bandwidth. Consequently, the coding parameters in the video codec need to be adjusted to achieve the required bit rate. In this chapter, we first review different popular 2D video rate control methods, and then discuss how to extend the rate control methods to different 3D video streaming scenarios. For the multi-view system, changing the viewing angle from one point to another point to observe a 3D scene (view switching) is a key feature to enable the immersive viewing experience. We address the challenges and the corresponding solutions for 3D view switching. In the third part of this chapter, we discuss the peer-to-peer 3D video streaming services. As the required bandwidth for 3D visual communication service poses a heavy bandwidth requirement on centralized streaming systems, the peer-to-peer paradigm shows great potential for penetrating the 3D video streaming market. After this, we cover 3D video broadcasting and 3D video communication over 4G cellular networks.

Acknowledgements

We would like to thank a few of the great many people whose contributions were instrumental in taking this book from an initial suggestion to a final product. First, we would like to express our gratitude to Dr. Chi-Yuan Yao for his help on collecting and sketching the content in Sections 9.1 and 9.2 and help with finishing Chapter 9 in time. We also thank him for his input on scene representation because of his deep domain knowledge in the field of computer geometry. We would like to thank Dr. Peng Yin and Dr. Taoran Lu for their help in enriching the introduction of HEVC. We also thank Mr. Dobromir Todorov for help in rendering figures used in Chapters 2 and 9. Finally, the authors appreciate the many contributions and sacrifices that our families have made to this effort. Guan-Ming Su would like to thanks his wife Jing-Wen's unlimited support and understanding during the writing process; and also would like to dedicate this book to his parents. Yu-Chi Lai would like to thank his family for their support of his work. Andres Kwasinski would like to thank his wife Mariela and daughters Victoria and Emma for their support, without which this work would not have been possible. Andres would also like to thank all the members of the Department of Computer Engineering at the Rochester Institute of Technology. Haohong Wang would like to thank his wife Xin Lu, son Nicholas and daughter Isabelle for their kind supports as always, especially for those weekends and nights that he had to be separated from them to work on this book at the office. The dedication of this book to our families is a sincere but inadequate recognition of all their contributions to our work.

About the Authors

Guan-Ming Su received the BSE degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 1996 and the MS and PhD degrees in Electrical Engineering from the University of Maryland, College Park, U.S.A., in 2001 and 2006, respectively. He is currently with Dolby Labs, Sunnyvale, CA. Prior to this he has been with the R&D Department, Qualcomm, Inc., San Diego, CA; ESS Technology, Fremont, CA; and Marvell Semiconductor, Inc., Santa Clara, CA. His research interests are multimedia communications and multimedia signal processing. He is the inventor of 15 U.S. patents and pending applications. Dr Su is an associate editor of Journal of Communications; guest editor in Journal of Communications special issue on Multimedia Communications, Networking, and Applications; and Director of review board and R-Letter in IEEE Multimedia Communications Technical Committee. He serves as the Publicity Co-Chair of IEEE GLOBECOM 2010, International Liaison Chair in IEEE ICME 2011, Technical Program Track Co-Chair in ICCCN 2011, and TPC Co-Chair in ICNC 2013. He is a Senior member of IEEE.

Yu-Chi Lai received the B.S. from National Taiwan University, Taipei, R.O.C., in 1996 in Electrical Engineering Department. He received his M.S. and Ph.D. degrees from University of Wisconsin–Madison in 2003 and 2009 respectively in Electrical and Computer Engineering. He received his M.S. and Ph.D. degrees from University of Wisconsin–Madison in 2004 and 2010 respectively in Computer Science. He is currently an assistant professor in NTUST. His research focus is on the area of computer graphics, computer vision, multimedia, and human-computer interaction. Due to his personal interesting, he is interested in industrial projects and he currently also cooperates with IGS to develop useful and interesting computer game technologies and NMA to develop animation technologies.

Andres Kwasinski received in 1992 his diploma in Electrical Engineering from the Buenos Aires Institute of Technology, Buenos Aires, Argentina, and, in 2000 and 2004 respectively, the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Maryland, College Park, Maryland. He is currently an Assistant Professor at the Department of Computer Engineering, Rochester Institute of Technology, Rochester, New York. Prior to this, he was with the Wireless Infrastructure group at Texas Instruments Inc., working on WiMAX and LTE technology, and with the University of Maryland, where he was a postdoctoral Research Associate. Dr. Kwasinski is a Senior Member of the IEEE, an Area Editor for the IEEE Signal Processing Magazine and Editor for

the IEEE Transactions on Wireless Communications. He has been in the Organizing Committee for the 2010 IEEE GLOBECOM, 2011 and 2012 IEEE ICCCN, 2012 ICNC and 2013 IEEE ICME conferences. Between 2010 and 2012 he chaired the Interest Group on Distributed and Sensor Networks for Mobile Media Computing and Applications within the IEEE Multimedia Communications Technical Committee. His research interests are in the area of multimedia wireless communications and networking, cross layer designs, cognitive and cooperative networking, digital signal processing and speech, image and video processing for signal compression and communication, and signal processing for non-intrusive forensic analysis of speech communication systems.

Haohong Wang received the B.S. degree in computer science and the M.Eng. degree in computer applications both from Nanjing University, China, the M.S. degree in computer science from University of New Mexico, and the Ph.D. degree in Electrical and computer engineering from Northwestern University, Evanston, USA. He is currently the General Manager of TCL Research America, TCL Corporation, at Santa Clara, California, in charge of the overall corporate research activities in North America with research teams located at fourplaces. Prior to that he held various technical and management positions at AT&T, Catapult Communications, Qualcomm, Marvell, TTE and Cisco. Dr. Wang's research involves the areas of multimedia processing and communications, mobile sensing and data mining. He has published more than 50 articles in peer-reviewed journals and International conferences. He is the inventor of more than 40 U.S. patents and pending applications. He is the co-author of 4G Wireless Video Communications (John Wiley & Sons, 2009), and Computer Graphics (1997).

Dr. Wang is the Editor-in-Chief of the Journal of Communications, a member of the Steering Committee of IEEE Transactions on Multimedia, and an editor of IEEE Communications Surveys & Tutorials. He has been serving as an editor or guest editor for many IEEE and ACM journals and magazines. He chairs the IEEE Technical Committee on Human Perception in Vision, Graphics and Multimedia, and was the Chair of the IEEE Multimedia Communications Technical Committee. He is an elected member of the IEEE Visual Signal Processing and Communications Technical Committee, and IEEE Multimedia and Systems Applications Technical Committee. Dr. Wang has chaired more than dozen of International conferences, which includes the IEEE GLOBECOM 2010 (Miami) as the Technical Program Chair, and IEEE ICME 2011 (Barcelona) and IEEE ICCCN 2011 (Maui) as the General Chair.

1

Introduction

1.1 Why 3D Communications?

Thanks to the great advancement of hardware, software, and algorithms in the past decade, our daily life has become a major digital content producer. Nowadays, people can easily share their own pieces of artwork on the network with each other. Furthermore, with the latest development in 3D capturing, signal processing technologies, and display devices, as well as the emergence of 4G wireless networks with very high bandwidth, coverage, and capacity, and many advanced features such as quality of service (QoS), low latency, and high mobility, 3D communication has become an extremely popular topic. It seems that the current trend is closely aligned with the expected roadmap for reality video over wireless, estimated by Japanese wireless industry peers in 2005 (as shown in Figure 1.1), according to which the expected deployment timing of stereo/multi-view/hologram video is around the same time as the 4G wireless networks deployment. Among those 3D video representation formats, the stereoscopic and multi-view 3D videos are more mature and the coding approaches have been standardized in Moving Picture Experts Group (MPEG) as "video-plus-depth" (V+D) and the Joint Video Team (JVT) Multi-view Video Coding (MVC) standard, respectively. The coding efficiency study shows that coded V+D video only takes about 1.2 times bit rate compared to the monoscopic video (i.e., the traditional 2D video). Clearly, the higher reality requirements would require larger volumes of data to be delivered over the network, and more services and usage scenarios to challenge the wireless network infrastructures and protocols.

From a 3D point of view, reconstructing a scene remotely and/or reproducibly as being presented face-to-face has always been a dream through human history. The desire for such technologies has been pictured in many movies, such as *Star Trek*'s Holodeck, *Star War*s' Jedi council meeting, *The Matrix*'s matrix, and *Avatar*'s Pandora. The key technologies to enable such a system involve many complex components, such as a capture system to describe and record the scene, a content distribution system to store/transmit the recorded scene, and a scene reproduction system to show the captured scenes to end users. Over the past several decades, we have witnessed the success of many applications, such as television broadcasting systems in analog (e.g., NTSC, PAL) and digital (e.g., ATSC, DVB) format, and home entertainment system in VHS, DVD, and Blu-ray format.

3D Visual Communications, First Edition. Guan-Ming Su, Yu-Chi Lai, Andres Kwasinski and Haohong Wang.

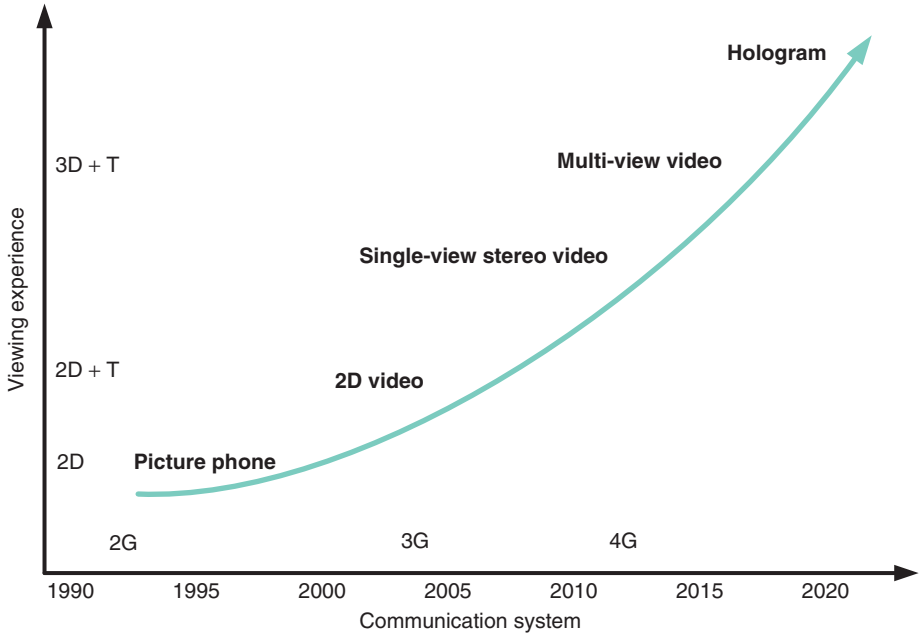

Figure 1.1 Estimated reality video over wireless development roadmap.

Although those systems have served for many years and advanced in many respects to give better viewing experiences, end users still feel that the scene reconstruction has its major limitation: the scene presentation is on a 2D plane, which significantly differs from the familiar three-dimensional view of our daily life. In a real 3D world, humans can observe objects and scenes from different angles to acquire a better understanding of the geometry of the watched scenes, and nonverbal signals and cues in visual conversation. Besides, humans can perceive the depth of different objects in a 3D environment so as to recognize the physical layout and location for each object. Furthermore, 3D visual systems can provide immersive viewing experience and higher interaction. Unfortunately, the existing traditional 2D visual systems cannot provide those enriched viewing experiences.

The earliest attempt to construct a 3D image was via the anaglyph stereo approach which was demonstrated by W. Rollmann in 1853 and J. C. D'Almeida in 1858 and patented in 1891 by Louis Ducos du Hauron. In 1922, the earliest confirmed 3D film was premiered at the Ambassador Hotel Theater in Los Angeles and was also projected in the red/green anaglyph format. In 1936, Edwin H. Land invented the polarizing sheet and demonstrated 3D photography using polarizing sheet at the Waldorf-Astoria Hotel. The first 3D golden era was between 1952 and 1955, owing to the introduction of color stereoscopy. Several golden eras have been seen since then. However, there are many factors affecting the popularity and success of 3D visual systems, including the 3D visual and content distribution technologies, the viewing experience, the end-to-end ecosystem, and competition from improved 2D systems. Recently, 3D scene reconstruction algorithms have achieved great improvement, which enables us to reconstruct a 3D scene from a 2D one and from stereoscope images, and the corresponding hardware can support the heavy computation

at a reasonable cost, and the underlying communication systems have advanced to provide sufficient bandwidth to distribute the 3D content. Therefore, 3D visual communication systems have again drawn considerable attention from both academia and industry.

In this book, we discuss the details of the major technologies involved in the entire end-to-end 3D video ecosystem. More specifically, we address the following important topics and the corresponding opportunities:

- the lifecycle of the 3D video content through the end-to-end 3D video communication framework,
- the 3D content creation process to construct a 3D visual experience,
- the different representations and compression formats for 3D scenes/data for content distribution. Each format has its own advantages and disadvantages. System designers can choose the appropriate solution for given the system resources, such as computation complexity and communication system capacity. Also, understanding the unequal importance of different syntaxes, decoding dependencies, and content redundancies in 3D visual data representation and coding can help system designers to adopt corresponding error resilient methods, error concealment approaches, suitable unequal error protection, and customized dynamic resource allocation to improve the system performance,
- the advanced communication systems, such as 4G networks, to support transmission of 3D visual content. Being familiar with those network features can help the system designer to design schedulers and resource allocation schemes for 3D visual data transmission over 4G networks. Also, we can efficiently utilize the QoS mechanisms supported in 4G networks for 3D visual communications,
- the effective 3D visual data transmission and network architectures to deliver 3D video services and their related innovative features,
- the 3D visual experience for typical users, the factors that impact on the user experiences, and 3D quality of experience (QoE) metrics from source, network, and receiver points of view. Understanding the factors affecting 3D QoE is very important and it helps the system designer to design a QoE optimized 3D visual communications system to satisfy 3D visual immersive expectations,
- the opportunities of advanced 3D visual communication applications and services, for example, how to design the source/relay/receiver side of an end-to-end 3D visual communication system to take advantage of new concepts of computing, such as green computing, cloud computing, and distributed/collaborated computing, and how to apply scalability concepts to handle 3D visual communications given the heterogeneous 3D terminals in the networks is an important topic.

1.2 End-to-End 3D Visual Ecosystem

As shown by the past experience and lessons learned from the development and innovation of visual systems, the key driving force is all about how to enrich the user experiences, or so-called QoE. The 3D visual system also faces the same issues. Although a 3D visual system provides a dramatic new user experience after traditional 2D systems, the QoE concept has to be considered at every stage of the communication system pipeline during system design and optimization work to ensure the worthwhileness of moving from 2D to

3D. There are many factors affecting the QoE, such as errors in multidimensional signal processing, lack of information, packet loss, and optical errors in display. Improperly addressing QoE issues will result in visual artifacts (objectively and subjectively), visual discomfort, fatigue, and other things that degrade the intended 3D viewing experiences.

An end-to-end 3D visual communication pipeline consists of the content creation, 3D representation, data compression, transmission, decompression, post-processing, and 3D display stages, which also reflects the lifecycle of a 3D video content in the system. We illustrate the whole pipeline and the corresponding major issues in Figure 1.2. In addition, we also show the possible feedback information from later stages to earlier stages for possible improvement of 3D scene reconstruction.

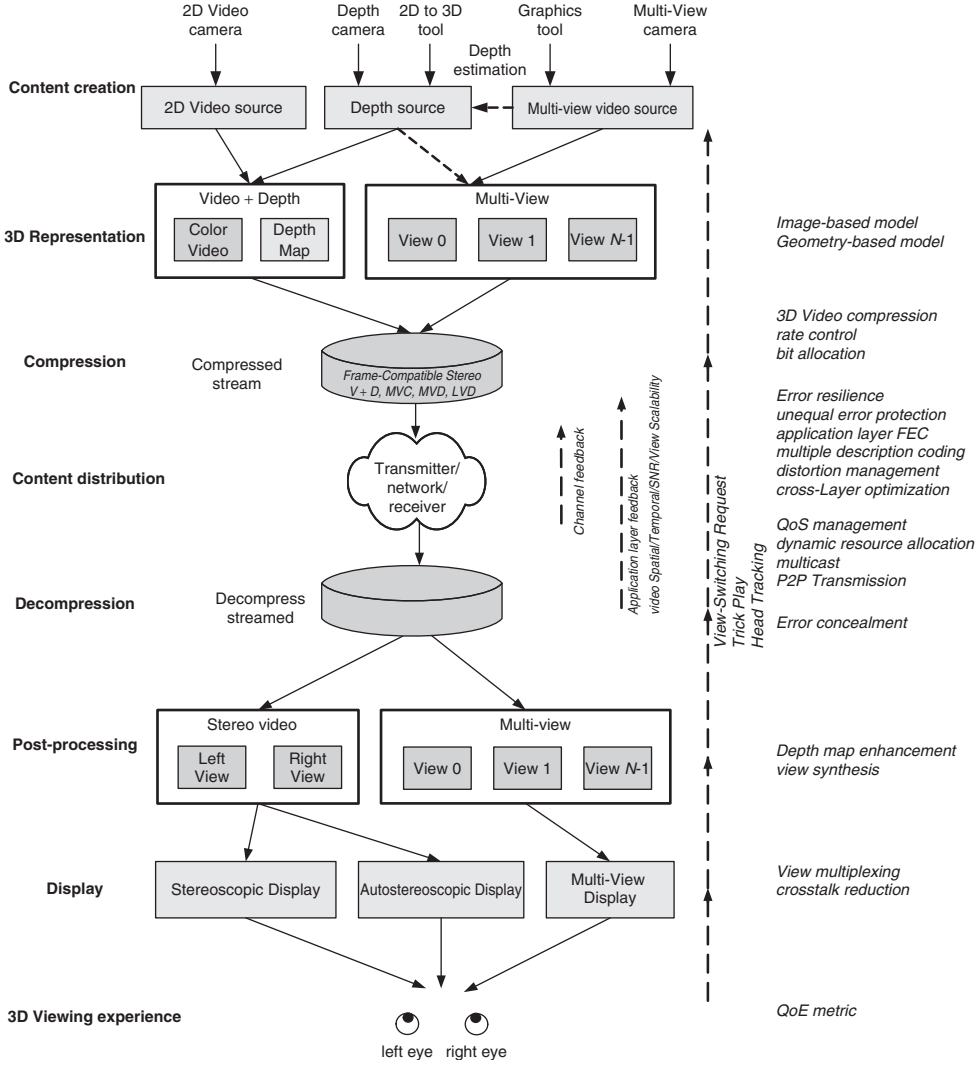

Figure 1.2 End-to-end 3D visual ecosystem.

The first stage of the whole pipeline is the content creation. The goal of the content creation stage is to produce 3D content based on various data sources or data generation devices. There are three typical ways of data acquisition which result in different types of data formats. The first is to use a traditional 2D video camera, which captures 2D images; the image can be derived for 3D data representation in the later stage of the pipeline. The second type is to use a depth video camera to measure the depth of each pixel corresponding to its counterpart color image. The registration of depth and 2D color image may be needed if sensors are not aligned. Note that in some depth cameras, the spatial resolution is lower than that of a 2D color camera. The depth image can also be derived from a 2D image with 2D-to-3D conversion tools; often the obtained depth does not have a satisfactory precision and thus causes QoE issues. The third type is to use an N-view video camera, which consists of an array of 2D video cameras located at different positions around one scene and all cameras are synchronized to capture video simultaneously, to generate N-view video. Using graphical tools to model and create 3D scene is another approach which could be time consuming, but it is popular nowadays to combine both graphical and video capturing and processing methods in the 3D content creation.

In the next stage, the collected video/depth data will be processed and transformed into 3D representation formats for different targeted applications. For example, the depth image source can be used in image plus depth rendering or processed for N-view application. Since the amount of acquired/processed/transformed 3D scene is rather large compared to single-view video data, there is a strong need to compress the 3D scene data. On the other hand, applying traditional 2D video coding schemes separately to each view or each different data type is inefficient as there exist certain representation/coding redundancies among neighboring views and different data types. Therefore, a dedicated compression format is needed at the compression stage to achieve better coding efficiency. In the content distribution stage, the packet loss during data delivery plays an important role in the final QoE, especially for streaming services. Although certain error concealment algorithms adopted in the existing 2D decoding and post-processing stages may alleviate this problem, directly applying the solution developed for 2D video system may not be sufficient. This is because the 3D video coding introduces more coding dependencies, and thus error concealment is much more complex compared to that in 2D systems. Besides, the inter-view alignment requirement in 3D video systems also adds plenty of difficulties which do not exist in 2D scenarios. The occlusion issue is often handled at the post-processing stage, and the packet loss will make the occlusion post-processing even more difficult. There are also some other application layer approaches to relieve the negative impact of packet loss, such as resilient coding and unequal error protection (UEP), and those technologies can be incorporated into the design of the 3D visual communication system to enrich the final QoE. At the final stage of this 3D visual ecosystem, the decoded and processed 3D visual data will be displayed on its targeted 3D display. Depending on the type of 3D display, each display has its unique characteristics of artifacts and encounters different QoE issues.

1.2.1 3D Modeling and Representation

3D scene modeling and representation is the bridging technology between the content creation, transmission, and display stages of a 3D visual system. The 3D scene modeling

and representation approaches can be classified into three main categories: geometry-based modeling, image based modeling, and hybrid modeling. Geometry-based representation typically uses polygon meshes (called surface-based modeling), 2D/3D points (called point-based modeling), or voxels (called volume-based modeling) to construct a 3D scene. The main advantage is that, once geometry information is available, the 3D scene can be rendered from any viewpoint and view direction without any limitation, which meets the requirement for a free-viewpoint 3D video system. The main disadvantage is in the computational cost of rendering and storing, which depends on the scene complexity, that is the total number of triangles used to describe the 3D world. In addition, geometry-based representation is generally an approximation to the 3D world. Although there are offline photorealistic rendering algorithms to generate views matching our perception of the real world, the existing algorithms using graphics pipeline still cannot produce realistic views on the fly.

The image based modeling goes to the other extreme, not using any 3D geometry, but using a set of images captured by a number of cameras with predesigned positions and settings. This approach tends to generate high quality virtual view synthesis without the effort of 3D scene reconstruction. The computation complexity via image based representation is proportional to the number of pixels in the reference and output images, but in general not to the geometric complexity such as triangle counts. However, the synthesis ability of image based representation has limitations on the range of view change and the quality depends on the scene depth variation, the resolution of each view, and the number of views. The challenge for this approach is that a tremendous amount of image data needs to be stored, transferred, and processed in order to achieve a good quality synthesized view, otherwise interpolation and occlusion artifacts will appear in the synthesized image due to lack of source data.

The hybrid approach can leverage these two representation methods to find a compromise between the two extremes according to given constraints. By adding geometric information into image based representation, the disocclusion and resolution problem can be relieved. Similarly, adding image information captured from the real world into geometry-based representation can reduce the rendering cost and storage. As an example, using multiple images and corresponding depth maps to represent 3D scene is a popular method (called depth image based representation), in which the depth maps are the geometric modeling component, but this hybrid representation can reduce the storage and processing of many extra images to achieve the same high-quality synthesized view as the image based approach. All these methods are demonstrated in detail in Chapters 2 and 4.

1.2.2 3D Content Creation

Other than graphical modeling approaches, the 3D content can be captured by various processes with different types of cameras. The stereo camera or depth camera simultaneously captures video and associated per-pixel depth or disparity information; the multi-view camera captures multiple images simultaneously from various angles, then multi-view matching (or correspondence) process is required to generate the disparity map for each pair of cameras, and then the 3D structure can be estimated from these disparity maps. The most challenging scenario is to capture 3D content from a normal 2D (or monoscopic) camera, which lacks of disparity or depth information, and where a

2D-to-3D conversion algorithm has to be triggered to generate an estimated depth map and thus the left and right views. The depth map can be derived from various types of depth cues, such as the linear perspective property of a 3D scene, the relationship between object surface structure and the rendered image brightness according to specific shading models, occlusion of objections, and so on. For complicated scenes, the interactive 2D-to-3D conversion, or offline conversion, tends to be adopted, that is, human interaction is required at certain stages of the processing flow, which could be in object segmentation, object selection, object shape or depth adjustment, object occlusion order specification, and so on. In Chapter 4, a few 2D-to-3D conversation systems are showcased to give details of the whole process flow.

1.2.3 3D Video Compression

Owing to the huge amount of 3D video data, there is a strong need to develop efficient 3D video compression methods. The 3D video compression technology has been developed for more than a decade and there have been many formats proposed. Most 3D video compression formats are built on state-of-the-art video codecs, such as H.264. The compression technology is often a tradeoff between the acceptable level of computation complexity and affordable budget in the communication bandwidth. In order to reuse the existing broadcast infrastructure originally designed for 2D video coding and transmission, almost all current 3D broadcasting solutions are based on a frame-compatible format via spatial subsampling approach, that is, the original left and right views are subsampled into half resolution and then embedded into a single video frame for compression and transmission over the infrastructure as with 2D video, and at the decoder side the demultiplexing and interpolation are conducted to reconstruct the dual views. The subsampling and merging can be done by either (a) side-by-side format, proposed by Sensio, RealD, and adopted by Samsung, Panasonic, Sony, Toshiba, JVC, and DirectTV (b) over/under format, proposed by Comcast, or (c) checkerboard format. A mixed-resolution approach is proposed, which is based on the binocular suppression theory showing that the same subjective perception quality can be achieved when one view has a reduced resolution. The mixed-resolution method first subsamples each view to a different resolution and then compresses each view independently.

Undoubtedly, the frame-compatible format is very simple to implement without changing the existing video codec system and underlying communication infrastructure. However, the correlation between left and right views has not been fully exploited, and the approach is mainly oriented to the two-view scenario but not to the multi-view 3D scenario. During the past decade, researchers have also investigated 3D compression from the coding perspective and 3D video can be represented in the following formats: two-view stereo video, video-plus-depth (V+D), multi-view video coding (MVC), multi-view video-plus-depth (MVD), and layered depth video (LDV). The depth map is often encoded via existing a 2D color video codec, which is designed to optimize the coding efficiency of the natural images. It is noted that depth map shows different characteristics from natural color image. Researchers have proposed several methods to improve the depth-based 3D video compression. In nowadays, free-viewpoint 3D attracts a lot of attention, in which the system allows end users to change the view position and angle to enrich their immersive experience. Hybrid approaches combining