Preface

In the past decades, digital video experiences have been continuously evolving towards a higher degree of immersion and realism. This development has been possible thanks to a number of technological advances. New acquisition and display devices enable capturing and rendering of video with unprecedented realism, through technologies such as Ultra High Definition (UHD), High Frame Rate (HFR), High Dynamic Range (HDR), and Wide Color Gamut (WCG). Wireless networks such as 5G and future 6G promise to deliver broadband communications and ultra low-latency (1 ms) to support telepresence in future applications such as *tactile Internet*. On the processing side, the advent of powerful graphic processing units (GPU) with massive parallelization have facilitated the surge of sophisticated image processing, computer vision and machine learning methods. Thanks to these new tools, it is possible to capture the content of a scene beyond the capabilities of the acquisition sensor, leading to computational photography and new rendering techniques. Visualization and interaction devices such as head-mounted displays (HMD), glasses-free multiview (or light field) displays, and haptic interfaces start to hit the market. At the same time, video formats such as volumetric video that provide six-degrees-of-freedom interaction are being standardized. All together, these technologies will contribute to the development of the *metaverse*, a set of perpetual and concurrent virtual spaces implementing various forms of *extended reality* (XR).

In this context, immersive video technologies enable new ways of visual communication and storytelling. The applications of immersive video are numerous and go beyond entertainment: they include interactions with objects and robots, tele-surgery, collaborative working, virtual driving, serious games, remote assistance and therapy, and many other examples. Immersive video technologies represent a major step forward in video technology, and are attracting a growing interest from manufacturers and service providers, with the goal to enhance the quality of experience of users.

This book aims to provide an overview and introduction to these new immersive video technologies. Specifically, we address different stages in the content production and delivery chain from acquisition and representation, to coding, streaming, visualization and quality monitoring. The book focuses on three main immersive modalities: omnidirectional video, light fields, and volumetric video. While not necessarily covering all the possible immersive video formats, these modalities are representative of the main commercially available (or viable) solutions nowadays. Furthermore, they pave the way to future immersive technologies where the boundaries between modalities fade away and vanish. We have intentionally left out other traditional immersive video formats, such as stereo and multiview video which have been thoroughly studied before and for which we provide references to previous books and surveys whenever needed. We also do not examine holography. While holography is perhaps the ultimate form of immersive video, its deployment is still far from the consumer market, mainly due to optical and physical limitations. The interested reader can easily find dedicated books covering holography both from the optical and signal processing perspectives.

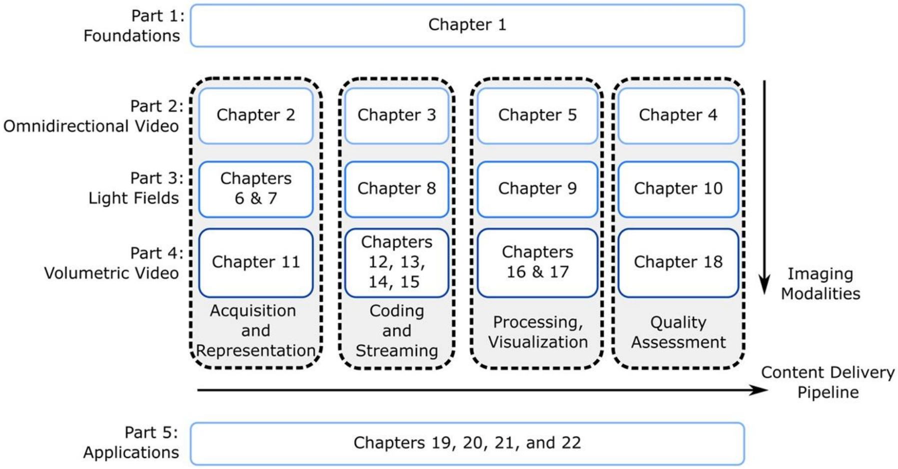

Figure 0.1 The organization of the book and the two main axes.

The organization of the book follows the two axes introduced above, as illustrated in Fig. 0.1: immersive video modalities on one hand, and stages of the content delivery pipeline on the other hand. There are three main immersive imaging modalities, which are reflected in the main book parts: Part 2 discusses the omnidirectional video (ODV) technologies; Part 3 focuses on light fields (LF); and Part 4 scrutinizes techniques for volumetric video (VV). The different stages of the content delivery pipeline are reflected in the different chapters for each part, and cover image acquisition and representation, coding and streaming, processing, visualization, and quality assessment. Part 1 and Part 5 are transversal, as they provide a general introduction and discuss different applications of immersive video, respectively. We describe the structure of the book in further detail in the following.

The foundational definitions regarding immersion and presence, which define immersive video, are discussed in Part 1. Chapter 1 is the only chapter in this part and provides the theoretical background for the modalities discussed in the book and definitions for the relevant terms that are necessary to understand the needs and the concepts of different immersive video technologies. The main stages of video content delivery chain are briefly explained, also laying out the limitations and challenges for immersive video.

Omnidirectional video and how it is captured, processed, transmitted, and used are discussed in Part 2. In Chapter 2, Maugey discusses existing omnidirectional video acquisition techniques including how ODVs are represented and rendered after acquisition. A multi-view ODV dataset capture is also discussed in the same chapter. Rossi, Guedes and Toni focus in Chapter 3 on the user behavior in consuming ODVs and the implications on ODV streaming systems. They also provide a comprehensive overview of the research efforts in this field. Croci, Singla, Fremeret, Raake, and Smolic deal with subjective and objective quality assessment for ODV in Chapter 4, discussing different methodologies. Finally, in Chapter 5, Chao, Battisti, Lebreton, and Raake discuss how to estimate visual saliency for ODVs, and how to collect visual attention data from observers.

Light fields are discussed in Part 3. This part starts with two chapters by Herfet, Chelli, and Le Pendu: Chapter 6 explains LF image and video acquisition; Chapter 7 presents different LF representations. Stepanov, Valenzise, and Dufaux discuss compression and transmission for LFs and describe the advances in JPEG Pleno standardization in Chapter 8. In Chapter 9, Keinert et al. discuss how LF can be captured, processed, and rendered for media applications, and how they can be integrated into a VR environment, as well as their quality enhancement through neural networks. Lastly, Ak and Le Callet present an overview of various subjective and objective factors affecting the Quality of Experience for LFs in Chapter 10.

Volumetric video (VV) is among the most compelling immersive technologies as it enables a full interaction between users and the scene. Part 4 regroups several volumetric video representations. This part starts with Chapter 11, where Eisert, Schreer, Feldmann, Hellge, and Hilsmann focus on various stages of VV acquisition, interaction, streaming and rendering, including interesting discussions on how to animate and make transformations on the volumetric content. In Chapter 12, Garus, Milovanović, Jung, and Cagnazzo present the MPEG Immersive Video (MPEG-I) coding standard, which is the successor of 3D-HEVC and implements a 2D-based texture-plus-depth scene representation for free viewpoint video. A native 3D representation is point clouds. The state-of-the-art techniques for point cloud compression, including available standards and recent learning-based approaches, are presented in Chapter 13 by Valenzise, Quach, Tian, Pang, and Dufaux. In Chapter 14, Marvie, Krivokuća, and Graziosi discuss another representation for VV that is more common in computer graphics, i.e., 3D dynamic meshes, and they provide a review of state-of-the-art dynamic mesh compression. In Chapter 15, Viola and Cesar offer an overview of the current approaches and implementations for VV streaming, focusing on both theoretical approaches and their adaptation to networks and users. In Chapter 16, Chen and Zheng discuss various processing methods that can be used to analyze, modify, and synthesize VVs for different tasks. Liu, Zhong, Mantel, Forchhammer, and Mantiuk describe the concept of computational 3D displays, underlying key performance factors from a visual perception point of view. Lastly, in Chapter 18, Alexiou and others present an overview of subjective and objective quality methods used for VVs in point cloud and 3D mesh formats.

Finally, Part 5 focuses on various applications that can benefit from immersive video technologies. In Chapter 19, Palomar, Kumar, Wang, Pelanis, and Cheikh describe how 3D organs models are used in mixed reality, after their digital twins are obtained by magnetic resonance and computed tomography scans. Helzle describes the use of light fields and immersive LED walls for visual effects in immersive media productions in Chapter 20. Young, O'Dwyer and Smolic discuss how volumetric video can be used as a novel medium for creative storytelling by describing some of the experiments conducted within the V-SENSE project in Chapter 21. In the last Chapter 22, Li and Cesar describe various social VR applications using immersive video technologies.

By covering the essential immersive video technologies in use today, this book has the ambition to become a reference for those interested in the fundamentals of immersive video, the various stages of the content lifespan, as well as the potential practical applications. Assuming basic knowledge on image and video processing, the book should be of interest to a broad audience with different backgrounds and expectations, including engineers, practitioners, researchers, as well as professors, graduate and undergraduate students, and managers making technological decisions about immersive video.

The Editors
Giuseppe Valenzise, Martin Alain, Emin Zerman, and Cagri Ozcinar