1

Introduction

Networked Audio-Visual Technologies form the basis for the multimedia communication systems that we currently use. The communication systems that must be supported are diverse, ranging from fixed wired to mobile wireless systems. In order to enable an efficient and cost-effective Networked Audio-Visual System, two major technological areas need to be investigated: first, how to process the content for transmission purposes, which involves various media compression processes; and second, how to transport it over the diverse network technologies that are currently in use or will be deployed in the near future. In this book, therefore, visual data compression schemes are presented first, followed by a description of various media transmission aspects, including various channel models, and content and link adaptation techniques.

Raw digital video signals are very large in size, making it very difficult to transmit or store them. Video compression techniques are therefore essential enabling technologies for digital multimedia applications. Since 1984, a wide range of digital video codecs have been standardized, each of which represents a step forward either in terms of compression efficiency or in functionality. The MPEG-x and H.26x video coding standards adopt a hybrid coding approach, employing block-matching motion estimation/compensation, in addition to the discrete cosine transform (DCT) and quantization. The reasons are: first, a significant proportion of the motion trajectories found in natural video can be approximately described with a rigid translational motion model; second, fewer bits are required to describe simple translational motion; and finally, the implementation is relatively straightforward and amenable to hardware solutions. These hybrid video systems have provided interoperability in heterogeneous network systems. Considering that transmission bandwidth is still a valuable commodity, ongoing developments in video coding seek scalability solutions to achieve a one-coding–multiple-decoding feature. To this end, the Joint Video Team of the ITU-T Video Coding Expert Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) have standardized a scalability extension to the existing H.264/AVC codec. The H.264-based Scalable Video Coding (SVC) allows partial transmission and decoding to the bit stream, resulting in various options in terms of picture quality and spatial-temporal resolutions.

In this book, several advanced features/techniques relating to scalable video coding are further described, mostly to do with 3D scalable video coding applications. Applications and scenarios for the scalable coding systems, advances in scalable video coding for 3D video applications, a non-standardized scalable 2D model-based video coding scheme applied on the

Visual Media Coding and Transmission Edited by Ahmet Kondoz

texture, and depth coding of 3D video are all discussed. A scalable, multiple description coding (MDC) application for stereoscopic 3D video is detailed. Multi-view coding and Distributed Video Coding concepts representing the latest advancements in video coding are also covered in significant depth.

The definition of video coding standards is of the utmost importance because it guarantees that video coding equipment from different manufacturers will be able to interoperate. However, the definition of a standard also represents a significant constraint for manufacturers because it limits what they can do. Therefore, in order to minimize the restrictions imposed on manufacturers, only those tools that are essential for interoperability are typically specified in the standard: *the normative tools*. The remaining tools, which are not standardized but are also important in video coding systems, are referred to as *non-normative tools* and this is where competition and evolution of the technology have been taking place. In fact, this strategy of specifying only the bare minimum that can guarantee interoperability ensures that the latest developments in the area of non-normative tools can be easily incorporated in video codecs without compromising their standard compatibility, even after the standard has been finalized. In addition, this strategy makes it possible for manufacturers to compete against each other and to distinguish between their products in the market. A significant amount of research effort is being devoted to the development of non-normative video coding tools, with the target of improving the performance of standard video codecs. In particular, due to their importance, rate control and error resilience non-normative tools are being researched. In this book, therefore, the development of efficient tools for the modules that are non-normative in video coding standards, such as rate control and error concealment, is discussed. For example, multiple video sequence (MVS) joint rate control addresses the development of rate control solutions for encoding video scenes formed from a composition of video objects (VOs), such as in the MPEG-4–standard, and can also be applied to the joint encoding and transcoding of multiple video sequences (VSs) to be transmitted over bandwidth-limited channels using the H.264/AVC standard.

The goal of wireless communication is to allow a user to access required services at any time with no regard to location or mobility. Recent developments in wireless communications, multimedia technologies, and microelectronics technologies have created a new paradigm in mobile communications. Third/fourth-generation (3G/4G) wireless communication technologies provide significantly higher transmission rates and service flexibility over a wide coverage area, as compared with second-generation (2G) wireless communication systems. High-compression, error-robust multimedia codecs have been designed to enable the support of multimedia application over error-prone bandwidth-limited channels. The advances of VLSI and DSP technologies are enabling lightweight, low-cost, portable devices capable of transmitting and viewing multimedia streams. The above technological developments have shifted the service requirements of mobile communication from conventional voice telephony to business- and entertainment-oriented multimedia services in wireless communication systems. In order to successfully meet the challenges set by the latest current and future audiovisual communication requirements, the International Telecommunication Union-Radio communications (ITU-R) sector has elaborated on a framework for global 3G standards by recognizing a limited number of radio access technologies. These are: Universal Mobile Telecommunications System (UMTS), Enhanced Data rates for GSM Evolution (EDGE), and CDMA2000. UMTS is based on Wideband CDMA technology and is employed in Europe and Asia using the frequency band around 2 GHz. EDGE is based on TDMA technology and uses the same air interface as the successful 2G mobile system GSM. General Packet Radio Service (GPRS) and High-Speed Circuit Switched Data (HSCSD) are introduced by Phase 2 + of the GSM standardization process. They support enhanced services with data rates up to 144 kbps in the packet-switched and circuit-switched domains, respectively. EDGE, which is the evolution of GPRS and HSCSD, provides 3G services up to 500 kbps within GSM carrier spacing of 200 kHz. CDMA2000 is based on multi-carrier CDMA technology and provides the upgraded solution for existing IS-95 operators, mainly in North America. EDGE and UMTS are the most widely accepted 3G radio access technologies. They are standardised by the 3rd Generation Partnership Project (3GPP). Even though EDGE and UMTS are based on two different multiple-access technologies, both systems share the same core network. The evolved GSM core network serves for a common GSM/UMTS core network that supports GSM/GPRS/EDGE and UMTS access. In addition, Wireless Local Area Networks (WLAN) are becoming more and more popular for communication at homes, offices and indoor public areas such as campus environments, airports, hotels, shopping centres and so on. IEEE 802.11 has a number of physical layer specifications with a common MAC operation. IEEE 802.11 includes two physical layers – a frequency-hopping spread-spectrum (FHSS) physical layer and a direct-sequence spread-spectrum (DSSS) physical layer – and operates at 2 Mbps. The currently deployed IEEE 802.11b standard provides an additional physical layer based on a high-rate direct-sequence spread-spectrum (HR/DSSS). It operates in the 2.4 GHz unlicensed band and provides bit rates up to 11 Mbps. IEEE 802.11a standard for 5 GHz band provides high bit rates up to 54 Mbps and uses a physical layer based on orthogonal frequency division multiplexing (OFDM). Recently, IEEE 802.11g standard has also been issued to achieve such high bit rates in the 2.4 GHz band.

The Worldwide Interoperability for Microwave Access (WiMAX) is a telecommunications technology aimed at providing wireless data over long distances in different ways, from point-to-point links to full mobile cellular access. It is based on the IEEE 802.16 standard, which is also called WirelessMAN. The name WiMAX was created by the WiMAX Forum, which was formed in June 2001 to promote conformance and interoperability of the standard. The forum describes WiMAX as "a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL". Mobile WiMAX IEEE 802.16e provides fixed, nomadic and mobile broadband wireless access systems with superior throughput performance. It enables non-line-of-sight reception, and can also cope with high mobility of the receiving station. The IEEE 802.16e enables nomadic capabilities for laptops and other mobile devices, allowing users to benefit from metro area portability of an xDSL-like service.

Multimedia services by definition require the transmission of multiple media streams, such as video, still picture, music, voice, and text data. A combination of these media types provides a number of value-added services, including video telephony, E-commerce services, multi-party video conferencing, virtual office, and 3D video. 3D video, for example, provides more natural and immersive visual information to end users than standard 2D video. In the near future, certain 2D video application scenarios are likely be replaced by 3D video in order to achieve a more involving and immersive representation of visual information and to provide more natural methods of communication. 3D video transmission, however, requires more resources than the conventional video communication applications.

Different media types have different quality-of-service (QoS) requirements and enforce conflicting constraints on the communication networks. Still picture and text data are categorized as background services and require high data rates but have no constraints on the transmission delay. Voice services, on the other hand, are characterized by low delay. However, they can be coded using fixed low-rate algorithms operating in the 5–24 kbps range. In contrast to voice and data services, low-bit-rate video coding involves rates at tens to hundreds of kbps. Moreover, video applications are delay sensitive and impose tight constraints on system resources. Mobile multimedia applications, consisting of multiple signal types, play an important role in the rapid penetration of future communication services and the success of these communication systems. Even though the high transmission rates and service flexibility have made wireless multimedia communication possible over 3G/4G wireless communication systems, many challenges remain to be addressed in order to support efficient communications in multi-user, multi-service environments. In addition to the high initial cost associated with the deployment of 3G systems, the move from telephony and low-bit-rate data services to bandwidth-consuming 3G services implies high system costs, as these consume a large portion of the available resources. However, for rapid market evolvement, these wideband services should not be substantially more expensive than the services offered today. Therefore, efficient system resource (mainly the bandwidth-limited radio resource) utilization and QoS management are critical in 3G/4G systems.

Efficient resource management and the provision of QoS for multimedia applications are in sharp conflict with one another. Of course, it is possible to provide high-quality multimedia services by using a large amount of radio resources and very strong channel protection. However, this is clearly inefficient in terms of system resource allocation. Moreover, the perceptual multimedia quality received by end users depends on many factors, such as source rate, channel protection, channel quality, error resilience techniques, transmission/processing power, system load, and user interference. Therefore, it is difficult to obtain an optimal source and network parameter combination for a given set of source and channel characteristics. The time-varying error characteristics of the radio access channel aggravate the problem. In this book, therefore, various QoS-based resource management systems are detailed. For comparison and validation purposes, a number of wireless channel models are described. The key QoS improvement techniques, including content and link-adaptation techniques, are covered.

Future media Internet will allow new applications with support for ubiquitous media-rich content service technologies to be realized. Virtual collaboration, extended home platforms, augmented, mixed and virtual realities, gaming, telemedicine, e-learning and so on, in which users with possibly diverse geographical locations, terminal types, connectivity, usage environments, and preferences access and exchange pervasive yet protected and trusted content, are just a few examples. These multiple forms of diversity requires content to be transported and rendered in different forms, which necessitates the use of context-aware content adaptation. This avoids the alternative of predicting, generating and storing all the different forms required for every item of content. Therefore, there is a growing need for devising adequate concepts and functionalities of a context-aware content adaptation platform that suits the requirements of such multimedia application scenarios. This platform needs to be able to consume low-level contextual information to infer higher-level contexts, and thus decide the need and type of adaptation operations to be performed upon the content. In this way, usage constraints can be met while restrictions imposed by the Digital Rights Management (DRM) governing the use of protected content are satisfied.

In this book, comprehensive discussions are presented on the use of contextual information in adaptation decision operations, with a view to managing the DRM and the authorization for adaptation, consequently outlining the appropriate adaptation decision techniques and adaptation mechanisms. The main challenges are found by identifying integrated tools and systems that support adaptive, context-aware and distributed applications which react to the characteristics and conditions of the usage environment and provide transparent access and delivery of content, where digital rights are adequately managed. The discussions focus on describing a scalable platform for context-aware and DRM-enabled adaptation of multimedia content. The platform has a modular architecture to ensure scalability, and well-defined interfaces based on open standards for interoperability as well as portability. The modules are classified into four categories, namely: 1. Adaptation Decision Engine (ADE); 2. Adaptation Authoriser (AA); 3. Context Providers (CxPs); and 4. Adaptation Engine Stacks (AESs), which comprise Adaptation Engines (AEs). During the adaptation decision-taking stage the platform uses ontologies to enable semantic description of real-world situations. The decision-taking process is triggered by low-level contextual information and driven by rules provided by the ontologies. It supports a variety of adaptations, which can be dynamically configured. The overall objective of this platform is to enable the efficient gathering and use of context information, ultimately in order to build content adaptation applications that maximize user satisfaction.