Preface

Recent advances in network infrastructures and multimedia systems have facilitated the introduction of new multimedia services such as Internet Protocol Television (IPTV), Mobile Television, Video on Demand (VoD), Videoconferencing, Tele-medicine, and several other applications. Multimedia-content users are demanding to be always connected to the infrastructure and able to seamlessly access these services from any location and using any device. These services are already consuming around 50% of the Internet traffic and are expected to surpass peer-to-peer file sharing traffic by the end of this year (Cisco, 2011). Moreover, both industries and research communities are pushing towards the introduction of innovative services such as 3D Television (3DTV) and Free View Television (FVTV) which demand a significant increase in bandwidth requirements. Broadcast stereoscopic transmission over DVB-S2 (Digital Video Broadcasting - Satellite) is already with us, with channels like Eurosport 3D and Sky 3D providing such services. These offer a 3D experience by transmitting the left and right eye video side by side using MPEG-4 encoding techniques and MPEG-2 transport stream over satellite links. The received video is then presented to the user on a 3D display and has to wear 3D glasses to view the content. Furthermore, High-Definition Television (HDTV) is becoming common and countries like Japan are experimenting also with Super HDTV services that demand more data to be transmitted. All this increase in content and services has lead Cisco in predicting that the video traffic will reach 62% of the global Internet traffic by the end of 2015 (Cisco, 2011). These predictions indicate that audiovisual services are expected to provide a significant economic value for the coming years (Foster & Broughton, 2011).

The massive bandwidths occupied by video traffic and the huge demand provide several challenges to the service providers who cannot guarantee high quality content over bandwidth-limited channels (Wenger, 2003). Furthermore, the best effort nature of the IP core and unreliability of wireless networks can significantly reduce the quality of the transmitted content (Wenger, 2003), (Stockhammer, Hannuksela, & Wiegand, 2003). Moreover, these services are required on different devices having different resolution, frame rate and quality requirements. All these challenges have inspired both academia and industry to concentrate their research efforts to implement more advanced video compression schemes for both single and multiview video.

In addition, the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) have joined their efforts to try to develop video coding systems which deliver the best video quality at the lowest rates possible. These standardization bodies gave birth to H.264/AVC (ITU-T, 2007), which is the state of the art video coding system that is widely deployed today. They are also expected to draft the High Efficiency Video Coding (HEVC) in 2013, where it is expected to achieve twice the compression efficiency of H.264/AVC at high and ultra high definition television (Sullivan & Ohm, 2010). These developments are also accompanied by the standardization of Multiview Video Coding (MVC) and Scalable Video Coding (SVC), where the first one targets 3DTV and FVTV
applications while the latter is targeted for heterogeneous delivery of video content. These are included as extensions to the H.264 standard, H.264/MVC and H.264/SVC. Current standardization work is also being done for Multiview plus Depth Coding (MVD) which is expected to give better performance than MVC in both 3DTV and FVTV applications.

The multimedia signal processing and communication research community has also concentrated its research effort in the design of novel error resilient strategies, quality of experience modeling, cross layer optimization, peer-to-peer video streaming, and specialized multimedia network protocols. In addition to this research, the community has paved the way to the development of novel applications such as telemedicine, real-time intelligent surveillance systems, distributed multimedia retrieval services and several others.

The continuous and fast evolution of this research area has inspired the editors to develop this book, which covers a broad spectrum within the field of multimedia signal processing and communications. The editors felt the need to put the latest research and the future directions in this area together in one book, with the prime purpose to stimulate more research and innovation in multimedia systems. Another major objective of this book was to introduce this research topic to the non-expert readers in order to target this book to a wider audience including engineering and computer science students. The editors have therefore contacted several international experts in the concerned fields who cover different aspects of Multimedia signal processing and communications and were asked to write self-contained chapters about their latest research. The chapters had to be targeted for both expert and non-expert readers, and include future research trends within each field tackled in the chapter.

The book was purposely divided into three sections. The first section focuses on the state of the art video compression schemes and the basic principles deployed by these standards. This section explains the processes adopted by the state of the art codecs such as H.264/AVC, H.264/SVC, H.264/MVC, and those that will be probably deployed by the future H.265/HEVC. These standards will be covered in some depth highlighting the advantages and disadvantages of the architectures and mention future research trends in this area. The theory and future advances in this area are provided in this section. The second section presents the challenges involved when dealing with the transmission of video content over bandwidth-limited and noisy channels. The video artifacts which are produced because of these impairments are discussed in some detail. This section covers the error resilient and error concealment tools adopted by both single and multiview video coding schemes and will present novel solutions which are intended to inspire the readers. This section also treats Peer-to-Peer Video streaming, that is a transmission paradigm which is gaining interest by several companies, especially in China and United States. Peer-to-peer solutions can offer different data paths to a destination and can overcome the asymmetry between uplink and downlink in legacy telecommunication networks such as 3G networks. The third and last section is dedicated towards applications where it presents three promising solutions which may emerge in future technology. These include subject identification using low resolution cameras, the deployment of 4G technology for e-health systems, and image retrieval using Peer-to-Peer networks. Several other applications such as FVTV, 3DTV, layered coding and Peer-to-Peer video streaming were mentioned in other chapters which are incorporated in the other two sections. This list of applications is far from comprehensive and several other applications exist but is representative of the wide possibilities these new technologies are offering.

The first two sections contain introductory material suitable for understanding video compression and transmission systems. These sections are intended to provide an introduction to the basic concepts and methodologies in multimedia signal processing and communications and to develop a foundation
that can be used as the basis for further study and research in the respective fields. Therefore, non-expert readers are advised to start reading from the first chapter to the last. On the other hand, expert readers can read each chapter individually as they are self-contained. The knowledge and mathematical complexity presented within this book remains at a level well within the grasp of undergraduate and graduate students who have followed introductory courses in signal processing and communications.

The material in this book is organized as follows.

Chapter 1 describes in detail the basic components which make up any video coding standard available today. These principles are essential for non-experts, since the knowledge delivered in this chapter is essential for the proper understanding of the latter chapters. This book chapter also presents chronologically the complex evolution of video coding standards found today and also briefly introduces the High Efficiency Video Coding (HEVC) standard which is expected to be drafted by the end of 2013. Throughout this chronological representation of the standards the authors highlight the major contribution by each standard. Koumaras and Kourtis present existing subjective evaluation methodologies and explain Reduced reference objective methodology which can model the quality experienced by the end user.

Chapter 2 explains the visual distortions present in modern video communication including but not limited to blocking artifacts, blurring artifacts, ringing artifacts, basis pattern artifacts and others. In this chapter, the author, Unterweger, investigates the artifacts caused by H.264/MVC and Scalable Video Coding and analyzes the effect of distorted depth maps which are used for view rendering for both 3DTV and FVT applications. The author shows that higher quality can be achieved if the encoder adopts objective metrics which calculate the distortions included by the encoder rather than using the mean absolute error (MAE) adopted by all known video coding standards.

Chapter 3 provides an overview of motion estimation for video coding, which is the major computational complexity bottleneck in existing video encoders. Several motion estimation techniques are analyzed and the drawbacks of the different motion estimation schemes found in literature are pointed out in this chapter. Golam Sorwar and Manzur Murshed demonstrate that the fast motion estimation algorithms are directional and based on the unimodal error surface assumption, which does not always hold true in real world video sequences. The author presents the Distance dependent Thresholding Search (DTS) as a promising alternative solution.

Chapter 4 looks at the possible requirements of Regions-of-Interest (ROI) within video and images. Such a need occurs when a user is only interested in a particular part of the presented media. This is more pronounced in mobile applications where the resources and bandwidth are more limited, suggesting that higher quality video can only be afforded in smaller regions of the video or image. Furthermore, the region-of-interest should be tracked to guarantee a good quality of experience and adapted to the type of display available at the receiver. Grois and Hadar use their experience to show how the scalable video coding extension of H.264 can be used to provide this service. The complexity of the system is also considered in the study to find a good performance-complexity compromise. The authors show through experimental results the advantages of using such a technique and look into future directions in this field.

Chapter 5 delves into the current state of 3D video coding. 3D video is expected to be the next major advancement in multimedia systems that will provide user with a 3D scene immersive experience. This is provided through the depth impression of the observed scenery. 3D video has a long history dating back to concepts proposed in the 19th century but it is only now that technology is reaching the state to provide mass distribution of such content. In this chapter, Lee and Ho describe the overall technologies used in 3D video systems starting from the content capturing to their display. The authors focus on the
recent standardization activities that were and are undertaken by the MPEG group that are associated with the area of 3D video coding.

Chapter 6 focuses on motion estimation in 3D television. As image sensors become smaller and cheaper and computational power increases, more image and video data can be acquired and transmitted. The improvements in computational power allow also better processing tools for the analysis and reconstruction of 3D scenes. Compression plays an important role in transmitting all the generated data. Estrela and Franz tackle the issue of motion compensation/motion estimation in 3D TV. Motion vectors (MVs) can be exploited in a number of sub-systems within the end-to-end 3D video transmission chain. The authors show how motion compensation/motion estimation can be used to improve dynamic scene acquisition, content creation, 2D to 3D conversion, compression, decoding, scene rendering, error concealment, virtual/augmented reality handling, intelligent content retrieval and displaying. Furthermore, the authors argue that since most of the techniques currently applied to 3D are imported from 2D solutions, there is still a lot of room for improvement.

Chapter 7 introduces the different video streaming procedures available. The author describes the network protocols adopted in different video streaming services, and how these protocols can be used to achieve an acceptable level of quality including methodologies that achieve congestion control for video streaming services. This chapter explains in depth the various error resilient mechanisms available in the H.264/AVC video coding system and explains different error concealment strategies that can be used by the decoder to alleviate the effect of transmission errors or packet loss.

Chapter 8 explains how distortions in a frame can propagate in subsequent frames because of the motion compensation prediction process. This chapter explains the rate distortion optimization process adopted by most video compression encoders, and explains how error resilient rate distortion optimization can be used to increase error resilience. The authors propose two novel schemes that can be used which detect the regions which might contain distortions and increases the probability for these regions to be intra encoded, thus stopping the temporal propagation of that particular distortion.

Chapter 9 introduces Free-ViewPoint (FVP) 3DTV and tackles view interpolation, coding, and streaming. Interpolation is a technique used to create a new view in between existing reference views. This gives the user the possibility to interactively select a viewing angle and point of the scene. Moreover, it improves the coding efficiency of multi-view video sequences as the interpolation can be done at the receiver avoiding the transmission of additional streams from the server. Therefore, inter-view coding can be done by exploiting the redundancies within the views. S. Zinger, Do, de With, Pertovic, and Y. Morvan give an overview of the process needed in 3D free-view video communication. This consists of the interpolation techniques, free-view coding, and the introduction of a scalable free-view video streaming architecture. The authors consider a 3D video which is represented by texture and depth images and apply Depth Image Based Rendering (DIBR) for reconstruction of the required viewpoint. The authors analyze the impact of using interpolation to generate intermediate views and propose solutions that can be applied to improve the image quality of the generated view. They give particular attention to the current state-of-the-art interpolation methods for FVP and look at the challenges ahead.

Chapter 10 discusses the application of peer-to-peer solutions to video streaming. Internet has seen high growth in video streaming over the last decade. Most of this delivery is done through unicast links between a server and a number of clients. This implies that the same content is being sent to very distinct receiver terminals, placing a problem to both scalability and heterogeneity in content development and to network providers as the users demanding the service increase. These problems can be mitigated through the use of IP Multicast protocols and Quality of Service (QoS) solutions but not all the networks are
equipped with such facilities. Thus the main technology available relies on the best-effort and unicast nature of the Internet. This creates a demand for the development of new protocols and architectures that are capable of supporting the distribution of such large content in an efficient and transparent way. Several Peer-to-Peer protocols have been developed in the last years but these are mainly targeted for file sharing and are not necessarily good for multimedia and need adaptation and improvements. Monteiro, Cruz, Nunes, Patrikakis, Papaoulakis, and Calafate present the current state of research and development in Peer-to-Peer video streaming over the Internet. They focus on available peer-to-peer protocols, the associated architectures and current video coding techniques. Furthermore, they highlight currently funded European project in this area.

Chapter 11 presents a video surveillance application where the system is able to detect particular individuals when using low resolution video sequences. The presented system is able to detect the faces in low resolution video sequences. Object tracking and super-resolution techniques were combined to facilitate person identification. The object detection strategy was used to detect the people of interest, greatly reducing the amount of data for super resolution. A number of instances of these regions were used to obtain a super resolved faces, where improvement of around 7% were achieved for faces having an eye to eye distance of 14 pixels.

Chapter 12 gives details on the possible application of medical video transmission over wireless networks. 4G-based mobile networks are expected to provide higher bandwidth capabilities and thus the transmission of multimedia becomes more accessible and has lower latency than with current solutions. Medical video requires higher resolutions than other video applications since medical professionals need to make accurate diagnostics using this data. Thus 4G networks will be capable of presenting the high bandwidth, high data rates, and better quality of service demanded by such a service. The success of such implementations relies on the compatibility of emerging broadband wireless networks, such as mobile WiMAX and LTE networks, with upcoming m-health and e-health systems. Istepanian, Alinejad, and Philip highlight some of the challenges faced when developing m-health solutions over broadband wireless systems. They show and explain the importance of providing medical Quality of Service (m-QoS) and medical Quality of Experience (m-QoE) concepts when developing 4G-based medical video transmission systems. The authors also present a validation scenario to illustrate the concepts behind a typical medical video streaming application that can be used in a 4G-health system scenario. Furthermore, the authors comment on the need of more work and regulations to make such technology mainstream within the health industry.

Chapter 13 presents the application of peer to peer networks to facilitate content based image retrieval services. This chapter introduces the image features adopted for the application and the MPEG-7 standard descriptors. The author presents a number of similarity measures that can be used and then provide an in depth description of the peer to peer content based image retrieval system where multiple cues are used for the queries. This chapter demonstrates that reconfigurable P2P CIBR achieves the highest retrieval efficiency for different time to lives and thus can be a promising solution for content based image retrieval applications.

Reuben A. Farrugia
University of Malta, Malta

Carl J. Debono
University of Malta, Malta
REFERENCES


