Visual Communications and Networking

视频通信与网络

Edited by

Chang Wen Chen

陈长汶 主编

University of Science and
Technology of China Press

中国科学技术大学出版社

当代科学技术基础理论与前沿问题研究丛书

中国科学技术大学

校友文库

Visual Communications and Networking

视频通信与网络

Edited by

Chang Wen Chen

陈长汶 主编

University of Science and Technology of China Press

中国科学技术大学出版社

内容简介

本书收录了一系列有关视频通信与网络各方面最新进展的论文.每篇论文自成一章,分别由视频通信与网络的顶尖专家撰写,共14章.每章的作者至少有一位是中国科学技术大学的校友.本书主要涉及以下几个方面的研究内容:视觉信号的获取、处理,视觉传感器网络、视频信号编码与多尺度视频编码标准及其系统实现,视频通信系统中抗误差技术及媒体安全传输技术,新颖的分布式视频编码技术的发展与挑战,以及各种各样的现代移动视频技术的最新进展.

本书可作为信息科学技术领域高年级本科生和研究生的视频通信与网络参考教材,也可供从事研究和技术开发的专业人员自学和进修.

Visual Communications and Networking
Chang Wen Chen

Published by University of Science and Technology of China Press
96 Jinzhai Road, Hefei 230026, P. R. China

图书在版编目(CIP)数据

视频通信与网络=Visual Communications and Networking:英文/陈长汶主编.—合肥:中国科学技术大学出版社,2012.3
(当代科学技术基础理论与前沿问题研究丛书:中国科学技术大学校友文库)
“十二五”国家重点图书出版规划项目
ISBN 978-7-312-02923-3

Ⅰ.视… Ⅱ.陈… Ⅲ.图像通信—通信网—英文 Ⅳ.TN 919.85

中国版本图书馆CIP数据核字(2011)第259970号

出版发行 中国科学技术大学出版社
安徽省合肥市金寨路96号,230026
http://press.ustc.edu.cn
印　　刷 合肥晓星印刷有限责任公司
经　　销 全国新华书店
开　　本 710 mm × 1000 mm 1/16
印　　张 32
字　　数 639千
版　　次 2012年3月第1版
印　　次 2012年3月第1次印刷
定　　价 98.00元

编　委　会

总　序

大学最重要的功能是向社会输送人才，培养高质量人才是高等教育发展的核心任务．大学对于一个国家、民族乃至世界的重要性和贡献度，很大程度上是通过毕业生在社会各领域所取得的成就来体现的．

中国科学技术大学建校只有短短的五十余年，之所以迅速成为享有较高国际声誉的著名大学，主要就是因为她培养出了一大批德才兼备的优秀毕业生．他们志向高远、基础扎实、综合素质高、创新能力强，在国内外科技、经济、教育等领域做出了杰出的贡献，为中国科大赢得了“科技英才的摇篮”的美誉．

2008年9月，胡锦涛总书记为中国科大建校五十周年发来贺信，对我校办学成绩赞誉有加，明确指出：半个世纪以来，中国科学技术大学依托中国科学院，按照全院办校、所系结合的方针，弘扬红专并进、理实交融的校风，努力推进教学和科研工作的改革创新，为党和国家培养了一大批科技人才，取得了一系列具有世界先进水平的原创性科技成果，为推动我国科教事业发展和社会主义现代化建设做出了重要贡献．

为反映中国科大五十年来的人才培养成果，展示我校毕业生在科技前沿的研究中所取得的最新进展，学校在建校五十周年之际，决定编辑出版《中国科学技术大学校友文库》50种．选题及书稿经过多轮严格的评审和论证，入选书稿学术水平高，被列入“十一五”国家重点图书出版规划．

入选作者中，有北京初创时期的第一代学生，也有意气风发的少年班毕业生；有“两院”院士，也有中组部“千人计划”引进人才；有海内外科研院所、大专院校的教授，也有金融、IT行业的英才；有默默奉献、矢志报国的科技将军，也有在国际前沿奋力拼搏的科研将才；有“文革”后留美学者中第一位担任美国大学系主任的青年教授，也有首批获得新中国博士学位的中年学者……在母校五十周年华诞之际，他们通过著书立说的独特方式，向母校献礼，其深情厚谊，令人感佩！

《文库》于2008年9月纪念建校五十周年之际陆续出版，现已出书53部，在学术界产生了很好的反响．其中，《北京谱仪Ⅱ：正负电子物理》获得中国出版政府奖；中国物理学会每年面向海内外遴选10部“值得推荐的物理学新书”，2009年和2010年，《文库》先后有3部专著入选；新闻出版总署总结“‘十一五’

国家重点图书出版规划”科技类出版成果时，重点表彰了《文库》的2部著作；新华书店总店《新华书目报》也以一本书一个整版的篇幅，多期访谈《文库》作者.此外，尚有十数种图书分别获得中国大学出版社协会、安徽省人民政府、华东地区大学出版社研究会等政府和行业协会的奖励.

这套发端于五十周年校庆之际的文库，能在两年的时间内形成现在的规模，并取得这样的成绩，凝聚了广大校友的智慧和对母校的感情.学校决定，将《中国科学技术大学校友文库》作为广大校友集中发表创新成果的平台，长期出版.此外，国家新闻出版总署已将该选题继续列为“十二五”国家重点图书出版规划，希望出版社认真做好编辑出版工作，打造我国高水平科技著作的品牌.

成绩属于过去，辉煌仍待新创.中国科大的创办与发展，首要目标就是围绕国家战略需求，培养造就世界一流科学家和科技领军人才.五十年来，我们一直遵循这一目标定位，积极探索科教紧密结合、培养创新拔尖人才的成功之路，取得了令人瞩目的成就，也受到社会各界的肯定.在未来的发展中，我们依然要牢牢把握“育人是大学第一要务”的宗旨，在坚守优良传统的基础上，不断改革创新，进一步提高教育教学质量，努力践行严济慈老校长提出的“创寰宇学府，育天下英才”的使命.

是为序.

侯建国

中国科学技术大学校长
中国科学院院士
第三世界科学院院士
2010年12月

Preface

With unprecedented advance in communication and networking technologies, enormous amount of multimedia data, especially images and videos, have been delivered to various media consumers accessed through heterogeneous networks and terminals. It has been predicted that by 2013 video streaming will be 90% of all Internet consumer traffic and 64% of mobile consumer traffic. Therefore, developing visual information processing technologies for efficient, robust, and secure transmission of visual information, in particular image and video data, to Internet and mobile consumers shall have a profound impact on quality of service for massive media consumers.

The research in visual communications and networking has witnessed significant flourishing since the widespread penetration of Internet towards the end of 20th century. In this relatively young yet very vigorous field of research, we can find numerous researchers who are graduated from University of Science and Technology of China (USTC). We can easily identify several USTC alumni who are among the pioneers in visual communication and networking. Two such pioneers are Dr. Ya-Qin Zhang who is now Vice President for Microsoft Corporation and Professor Weiping Li who has recently returned to Alma Mater and has been appointed Dean of the School of Information Science and Technology of USTC. Both of them have served as the Editor-in-Chief for the top Journal in this field—"IEEE Trans. Circuits and Systems for Video Technology", true recognition of their seminal contributions to video technology.

On the occasion of USTC's 50th Anniversary in 2008, it is therefore quite appropriate for me, a member of USTC class 1978 and a veteran researcher in visual communications and networking, to gather a group of USTC alumni to collectively contribute a book to be dedicated to the 50th Anniversary of our Alma Mater—USTC. I did try to recruit both Dr. Zhang and Professor Li to contribute to this edited book. Unfortunately, Dr. Ya-Qin Zhang was unable to contribute due to his administrative responsibilities. However, Professor Weiping Li, along with his two former students who are both USTC alumni, has actually contributed a separate book of their own to the Book Series for the USTC's 50th Anniversary.

After contacting several dozen of USTC alumni who are working in this emerging area of visual communications and networking, I was able to select 14 chapters for this edited book, covering a wide variety of topics ranging from visual information acquisition to video coding standards, and from video streaming to secure visual information transmission. This book is part of a book series in electrical electronic engineering and computer science to commemorate USTC's 50th anniversary. This series is intended to promote the national and international reputations of USTC and report the cutting-edge research from USTC alumni. All contributors are USTC alumni and they share the common pride that they have received their education from USTC, their Alma Mater.

We begin this book with two chapters in the important area of image acquisition. Chapter 1 is contributed by USTC alumnus Professor Xin Li from West Virginia University in USA. In this chapter, Prof. Xin Li provides a systematic survey of over seventy published works in the emerging technique of image demosaicing—the problem of interpolating full-resolution color images from so-called color-filter-array (CFA) samples. In this chapter, after extensive review, fundamental differences among competing approaches have been identified.

Chapter 2 of this book is contributed by USTC alumnus Dr. Min Wu, who is now a Senior Research Engineer in Mako Surgical Corp. in USA. This chapter addresses important technical issues in wireless imaging sensor networks with a wide range of applications, such as surveillance, traffic control, and environmental monitoring etc. The main results in this chapter reveal that there exists a clear need to develop more efficient image/video source coding and communication protocols to boost wireless imaging sensor network applications.

Chapter 3 is contributed by USTC alumnus Professor Lei Cao of the University of Mississippi in USA. In this chapter, unequal error protection for progressive image transmission with fixed codeword length is studied. First, it has been proved that for source codes with convex rate-distortion (R-D) function and practically used channel codes, the channel code rates for sequential packets in the optimal rate allocation are indeed non-decreasing. Analysis and simulation results are presented to demonstrate the performance of image transmission over error prone channels using rate compatible punctured turbo codes.

The next five chapters introduce various video coding technologies and standards. Chapter 4 is contributed by USTC alumnus Dr. Haoping Yu of

Huawei Technologies Co. in USA. This chapter presents an overview of newly completed high profile video coding standards by the Joint Video Team (JVT) of MPEG and ITU-T for high quality video applications. These new profiles have been developed for applications in the professional and semi-professional domains with enhanced compression capabilities. This chapter discusses the design, the coding structure and tools, the operation modes, the profile definitions, and the target application environments of these new profiles.

Chapter 5 provides comprehensive overview of scalable video coding algorithms and applications and is contributed by USTC Guest Professor Dr. Feng Wu and USTC alumni Dr. Shipeng Li. Both of them have been considered as leading experts in video coding, especially scalable video coding. Scalable video coding offers a good compression solution for applications that require the compressed video stream to be adaptive to transmission bandwidth fluctuations. This chapter will focus on a particular technique of fine granularity scalable (FGS) video coding in which enhancement layer streams can be arbitrarily truncated and provide flexible and precise adaptation on channel bandwidth variations.

Chapter 6 introduces an overview of the development about Chinese national standard named AVS (the Audio Video Coding Standard of China) in recent years. This chapter is contributed by USTC Professor Houqiang Li and his student Hui Liu. This AVS standard is mainly driven by the fast paced booming of China's information technology industry with an aim at establishing general technical standards for the compression, decoding, processing, and the representation of digital audio-video based on homegrown innovations and publicly available technologies. This AVS standard has attracted many attentions from researchers and engineers internationally and has been used on a trial basis for IPTV, mobile TV, and terrestrial broadcasting in China.

Chapter 7 presents the development of emerging real-time video coding technology and the implementation of these technologies on digital signal processor (DSP) architectures. This chapter is contributed by USTC alumnus Dr. Lulin Chen. Dr. Chen is now a Senior Research Engineer in Huawei Technologies Co. in USA. In this chapter, the architectures of real-time HD video coding: ASIC, FPGA, DSP, GPP, and GPU etc are reviewed in detail. Among all these architectures, the architecture based on DSP has the properties of high programmability and low development time, capable of keeping pace with the arrival of new standards and new application requirements within a short time period. This chapter's focus is on the implementation of real-time video coding based on DSP.

Chapter 8 presents a novel technique for the decoding of Huffman variable length codes (VLC). This chapter is contributed again by Prof. Lei Cao of University of Mississippi in USA. In general, VLCs are key components in most image and video coding standards and applications to achieve entropy approaching compression performance. However, these codes are sensitive to transmission error and may cause potential error propagation in the process of decoding. This chapter details an approach in designing the decoding scheme for VLCs aiming at reducing both Levenshtein distance and the decoding complexity by extracting the self-synchronization string (SSS) information from a given Huffman equivalent code. Such SSS information will then be incorporated into maximum a posteriori probability (MAP) decoding to achieve improved performance with reduced decoding complexity.

Chapters 9, 10, and 11 are all concerned with video communications and networking in different aspects. Chapter 9 is contributed by USTC alumnus Dr. Ye-Kui Wang and his colleague Miska Hannuksela when he worked at Nokia Research Center. Dr. Wang is now working with Qualcomm Inc., USA. This chapter discusses the topic of error robust video communications. This chapter provides an overview of typical video communication systems in terms of system framework and transmission protocol stack, focusing on packet based video transmission over the Internet Protocol. Channel coding and joint source channel error control schemes are also described, including a variety of retransmission and forward error correction strategies.

Chapter 10 presents a video coding scheme based on wavelet transform and is contributed by USTC alumnus Professor Jianyu Dong of California State University in Los Angeles in USA. Wavelet-based video codec has unique advantages in streaming applications. Its great scalability and high precision in rate control are ideal for the design of a real time video streaming system over heterogeneous networks to meet different QoS requirements of various users. This chapter provides an overview of current streaming technologies for 3D wavelet-based video. Specifically, the characteristics of 3D wavelet video structure are be analyzed, the design issues of real-time video streaming system over lossy networks are discussed, and the concept of rate-distortion optimized streaming is introduced.

Chapter 11 addresses a very important problem in secure media transmission and is contributed by another USTC alumnus Dr. Qibin Sun who is now working at Cisco Systems Shanghai as a Senior Director. In this chapter, after brief introduction of a list of signature based robust authentication schemes for multimedia applications, a more detailed description on robust multimedia authentication is highlighted. This chapter argues that a good semi-fragile authentication system should not only be secure enough to the malicious attacks but also be robust enough to the acceptable manipulations. This chapter shows that it is both necessary and possible to develop a scalable, robust and secure media (image and video) authentication framework.

The next two chapters present some unconventional video coding and communication techniques that are based on Slepian-Wolf and Wyner-Ziv distributed source coding theory and principles. Chapter 12 is contributed by Dr. Guogang Hua who is now a Research Engineer in Qualcomm Inc. in USA. This chapter tries to present the fundamentals of distributed video coding (DVC), current research themes, and encouraging results of DVC, as well as its applications. The new DVC architectures call for low power and low complexity at the mobile/sensor video encoding unit. Besides the lower complexity encoder, DVC also has the benefits over standard video coding in terms of improved error robustness as well as great potential in layered video coding and multi-terminal or multi-view video coding.

Chapter 13 is contributed by USTC alumna Dr. Qian Xu and her colleague and PhD supervisor in Texas A&M University. Dr. Xu is now a Media Architect at Intel Corp. in USA. This chapter focuses on the emerging layered Wyner-Ziv video compression for robust video delivery. The layered video coding system presented in this chapter achieves scalability as the layered Wyner-Ziv bitstream enhances the standard base layer bitstream in such a way that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. A practical distributed joint source and channel coding (JSCC) scheme is also presented that exploits a single digital fountain Raptor code for both compression and protection for transmission over erasure channels. The joint design based on novel distributed JSCC paradigm is superior to designs where compression and protection coding are treated separately.

Chapter 14 is contributed by Professor Chang Wen Chen, the Editor of this book. This chapter begins with discussions on some recent advances in contemporary technologies in multimedia communications over mobile wireless links, including some modern communication and networking technologies for reliable links, mobile devices technologies for handheld receivers, and related signal processing technologies for the enhancement of multimedia communication performances. Through the analysis of these advances, several emerging technical challenges and associated research opportunities in the development of future multimedia communication systems have been identified. This chapter concludes with a summary and illustrates the potential impact of these emerging technologies on the development of next generation mobile multimedia communication systems.

In summary, this book illustrates a very diverse range of contemporary image and video communication technologies that are shaping modern society in various aspects. These technologies are indeed at the forefront of the visual communication research and development. I am very pleased that many top researchers in this emerging field are actually USTC alumni. The authors of these chapters in this book represent only a small fraction of USTC alumni who were able to contribute with tight schedule.

I believe all authors of this edited book must have plenty pleasant memory of their undergraduate and graduate educational years in USTC. The rigorous training and unselfish teaching of USTC faculty must have imprinted on each of its graduates the spirit of pursuing for the truth in their scientific and engineering career. The unprecedented success of USTC graduates in various scientific and engineering disciplines has proved such spirit of USTC.

I would like to thank USTC Press for initiating the publication of such a wonderful series dedicated to the 50th Anniversary of USTC. Without the coordinated efforts of the authors, editors, and publication staff members, this book would not have been published smoothly.

Chang Wen Chen

Hefei, Anhui, China

Contents

Chapter 1

Image Demosaicing: A Systematic Survey and a New Contribution

Xin Li
West Virginia University, Morgantown

Abstract

Image demosaicing is a problem of interpolating full-resolution color images from so-called color-filter-array (CFA) samples. In this chapter, we provide a systematic survey of over seventy published works in this field since 1999. Our review attempts to address important issues to demosaicing and identify fundamental differences among competing approaches. Our findings suggest most existing works belong to the class of sequential demosaicing, i.e., luminance channel is interpolated first and then chrominance channels are reconstructed based on recovered luminance information. Despite its popularity, we emphasize it still represents an ad-hoc understanding of inter-channel dependency within color images.

We report our comparative study results with a collection of eleven competing algorithms whose source codes or executables are provided by the authors. Our comparison is performed on two data sets: Kodak PhotoCD (popular choice) and McMaster high-quality images (more challenging). While most existing demosaicing algorithms achieve good performance on the Kodak data set, their performance on the McMaster one (images with varying-hue and high-saturation edges) degrades significantly. Such observation suggests the importance of properly addressing the issue of mismatch between assumed model and observation data in demosaicing. An empirical Bayesian model-averaging (EBMA) framework is then presented to combine complementary demosaicing techniques for better performance. We conclude this chapter by highlighting several issues deserving further investigation such as model validation, benchmark selection and performance evaluation.

1 Introduction

Consumer-level digital cameras were introduced in the middle of 1990s; in the past decade, the digital camera market has grown rapidly to exceed that of film cameras. Today, there are consumer-grade point-and-shoot cameras with over 8 million pixels; professional-grade single-lens refection (SLR) cameras with more than 12 million pixels are also available. Resolution, light sensitivity, and dynamic range of the sensors have significantly improved such that image quality of digital cameras has become comparable to that of film cameras. Among the various operations during the image capture process, color demosaicing, also called color filter array (CFA) interpolation, is a crucial step affecting the image quality.

Why do we need CFA in color imaging? To produce a color image, there should be at least three color samples at each pixel location. One approach is to use beam-splitters along the optical path to project the image onto three separate sensors as illustrated in Fig. 1. Using a color filter in front of each sensor, three full-channel color images are obtained. This is a costly approach as it

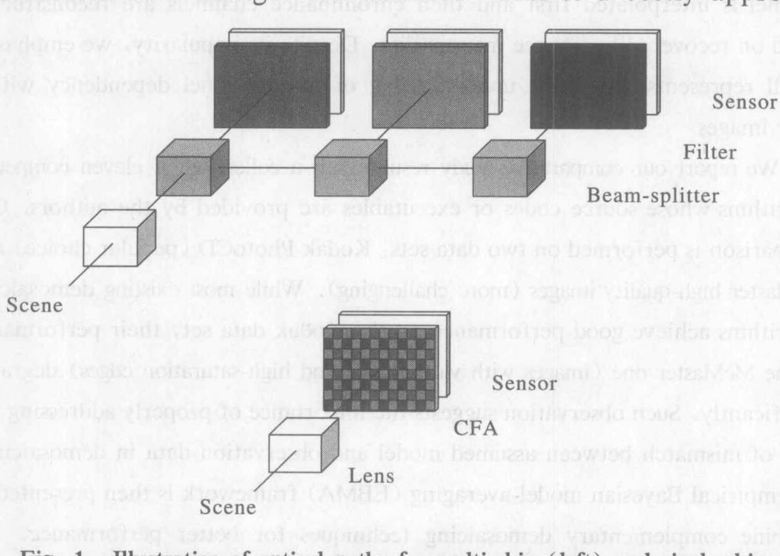

Fig. 1 Illustration of optical paths for multi-chip (left) and single-chip (right) digital cameras

requires three charge-coupled device (CCD) sensors and moreover these sensors have to be aligned precisely (a nontrivial challenge to mechanical design). A more cost-effective solution is to put a color filter array (CFA) in front of the sensor to capture one color component at a pixel and then interpolate the missing two color components[1]. Because of the mosaic pattern of the CFA, this interpolation process of CFA patterns has been widely known as "image demosaicing".

Among many CFA patterns, the most commonly used is so-called Bayer pattern[2] (refer to Fig. 2). The Bayer pattern measures the green image on a quincunx grid (half of the image resolution) and the red and blue images on rectangular grids (quarter of the image resolution). The green channel is measured at a higher sampling rate than the other two because the peak sensitivity of the human visual system (HVS) lies in the medium wavelengths, corresponding to the green portion of the spectrum[3]. Although we limit our discussion to the demosaicing problem with reference to the Bayer pattern here, the demosaicing algorithms developed for Bayer pattern can also be extended to other patterns. Systematic analysis and comparison of different CFA patterns can be found in recent works[4-5].

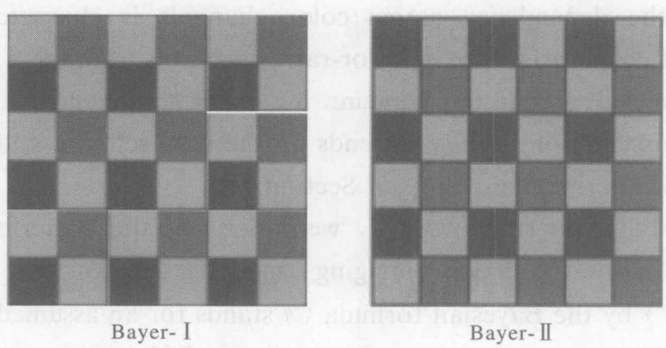

Fig. 2 Bayer pattern used in single-chip digital cameras (US Patent 3971065)

2 Problem Formulations

Intuitively image demosaicing can be best understood as an extension of image interpolation for grayscale images. If we denote full-resolution color image by S