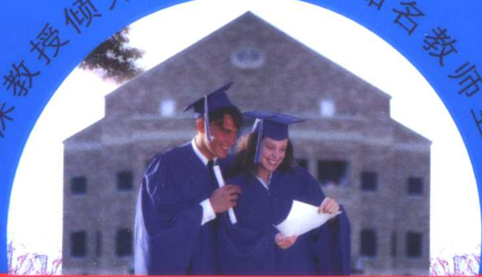
国外资深教授倾力之作　国内知名教师全力推荐

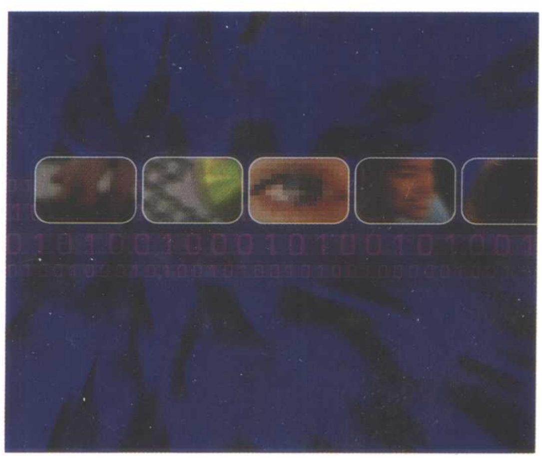


国外高校电子信息类优秀教材

压缩视频通信

Compressed Video Communications

(英文影印版)



Abdul H. Sadka 著

科学出版社
www.sciencep.com

国外高校电子信息类优秀教材(英文影印版)

压缩视频通信

Compressed Video Communications

Abdul H. Sadka 著

科学出版社
北京

图字 01-2003-7120

内 容 简 介

本书为国外高校电子信息类优秀教材(英文影印版)之一。

本书深入浅出地介绍了自1989年以来几乎全部ISO/ITU-T颁布的数字视频压缩算法标准与多媒体应用建议,加入了IP与ATM的一般技术内容,并体现了最新的研究成果。内容包括数字视频压缩算法、压缩视频通信的流控制、压缩视频通信的纠错、无线IP网络的视频通信、内部网络视频通信的代码转换等。

本书可作为"视频通信"课程的专业教材,也可作为相关工程技术人员的参考书。

图书在版编目(CIP)数据

压缩视频通信/(美)萨迪卡(Sadka, A. H.)著. 影印本. —北京:科学出版社,2004
(国外高校电子信息类优秀教材)
ISBN 7-03-013242-4

Ⅰ.压… Ⅱ.萨… Ⅲ.图像通信-高等学校-教材-英文 Ⅳ.TN919.8

中国版本图书馆CIP数据核字(2004)第029136号

责任编辑:李 宇 匡 敏
责任印制:安春生/封面设计:陈 敬

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com
双青印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2004年5月第 一 版 开本:787×1092 1/16
2004年5月第一次印刷 印张:18 1/4
印数:1—3 000 字数:338 000

定价: 28.00元 (含光盘)

(如有印装质量问题,我社负责调换〈环伟〉)

国外高校电子信息类优秀教材(英文影印版)

丛书编委会

(按姓氏笔画排序)

To both my parents,
my lovely wife and
our little daughter.

Preface

This book examines the technologies underlying the compression and transmission of digital video sequences over networking platforms. The incorporated study covers a large spectrum of topics related to compressed video communications. It presents to readers a comprehensive and structured analysis of the issues encountered in the transmission of compressed video streams over networking environments. This analysis identifies the problems impeding the progress of compressed video communication technologies and impairing the quality of networked video services; it also presents a wide range of solutions that would help in the quality optimisation of video communication services. The book is a unique reference in which the author has combined the discussions of several topics related to digital video technology such as compression, error resilience, rate control, video transmission over mobile networks and transcoding. All the techniques, algorithms, tools and mechanisms connected to the provision of video communication services are explained and analysed from the application layer down to the transport and network layers. The description of these technologies is accompanied by a large number of video subjective and objective results in order to help graduate students, engineers and digital video researchers make the best understanding of the presented material and build a comprehensive view on the progress of this rather fertile and exciting field of technology. A number of video clips have also been prepared and put at the disposal of readers on the supplementary CD in order to back up the results depicted and conclusions reached within the book chapters.

Abdul H. Sadka

Acknowledgement

The author would like to thank his research assistants, namely Dr S. Dogan for his valuable contribution to Chapter 6, and both Dr S. Dogan and Dr S. Worrall for their assistance with preparing the video clips on the supplementary CD. Particular thanks also go to our computing staff for their understanding and cooperation in dealing with persistent requests for extra video storage space and processing power.

About the Author

Dr Sadka obtained a PhD in Electrical & Electronic Engineering from the University of Surrey in 1997, an MSc with distinction in Computer Engineering from the Middle East Technical University in 1993, and a BSc in Computer and Communications Engineering from the American University of Beirut in 1990. In October 1997, he became a lecturer in Multimedia Communication Systems in the Centre for Communication Systems Research at the University of Surrey, and has been a member of the academic staff of the university since then.

He and his researchers have pioneered work on various aspects of video coding and transcoding, error resilience techniques in video communications and video transmissions over networks. His work in this area has resulted in numerous novel techniques for error and flow control in digital video communications. He has contributed to several short courses delivered to UK industry on various aspects of media compression and multimedia communication systems. He served on the program, advisory and technical committees of various specialised international conferences and workshops. He acts as consultant on image/video compression and mobile video communications to potential companies in the UK Telecommunications industry. He is well supported by the industry in the form of research projects and consultancy contracts. He runs a private consultancy company VIDCOM.

Dr Sadka is the instigator of numerous funded projects covering a wide spectrum of multimedia communications. He has been serving as a referee to a number of *IEE Transactions* and *IEE Electronics Letters* since 1997. He has 2 patents filed in the area of video compression and mobile video systems and over 40 publications in peer refereed conferences and journals, in addition to several contributions to multi-authored books on mobile multimedia communications in the UK and abroad. He is a member of the IEE and a chartered electrical engineer.

Contents

1

Introduction

1.1 Background

Both the International Standardisation Organisation (ISO) and the International Telecommunications Union (ITU) standardisation bodies have been releasing recommendations for universal image and video coding algorithms since 1985. The first image coding standard, namely JPEG (Joint Picture Experts Group), was released by ISO in 1989 and later by ITU-T as a recommendation for still image compression. In December 1991, ISO released the first draft of a video coding standard, namely MPEG-1, for audiovisual storage on CD-ROM at 1.5–2 Mbit/s. In 1990, CCITT issued its first video coding standard which was then, in 1993, subsumed into an ITU-T published recommendation, namely ITU-T H.261, for low bit rate communications over ISDN networks at $p \times 64$ kbit/s. ITU-T H.262, alternatively known as MPEG-2, was then released in 1994 as a standard coding algorithm for HDTV applications at 4–9 Mbit/s. Then, in 1996, standardisation activities resulted in releasing the first version of a new video coding standard, namely ITU-T H.263, for very low bit rate communications over PSTN networks at less than 64 kbit/s. Further work on improving the standard has ended up with a number of annexes that have produced more recent and comprehensive versions of the standard, namely H.263+ and H.263++ in 1998 and 1999 respectively. In 1998, the ISO MPEG (Motion Picture Experts Group) AVT (Audio Video Transport) group put forward a new coding standard, namely MPEG-4, for mobile audiovisual communications. MPEG-4 was the first coding algorithm that used the object-based strategy in its layering structure as opposed to the block-based frame structure in its predecessors. In March 2000, the standardisation sector of ISO published the most recent version of a standard recommendation, namely JPEG-2000 for still picture compression. Most of the aforementioned video coding algorithms have been adopted as the standard video codecs used in contemporary multimedia communication standards such as ITU-T H.323 and H.324 for the provision of multimedia communications over packet-switched and circuit-switched networks respectively. This remarkable evolution of video coding technology has underlined the development of a multitude of novel signal compression techniques that aimed to optimise the compression efficiency and quality of service of standard video coders. In this book, we put at the disposal of readers a comprehensive but simple explanation of the basic principles of video coding techniques employed in this long series of standards. Emphasis is placed on the major building blocks that constitute the body of the standardised video coding algorithms. A large number of tests are carried out and included in the book to enable the readers to evaluate the performance of the video coding standards and establish comparisons between them where appropriate, in terms of their coding efficiency and error robustness.

From a network perspective, coded video streams are to be transmitted over a variety of networking platforms. In certain cases, these streams are required to travel across a number of asymmetric networks until they get to their final destination. For this reason, the coded video bit streams have to be transmitted in the form of packets whose structure and size depend on the underlying transport protocols. During transmission, these packets and the enclosed video payload are exposed to channel errors and excessive delays, hence to information loss. Lost packets impair the reconstructed picture quality if the video decoder does not take any action to remedy the resulting information loss. This book covers a whole range of error handling mechanisms employed in video communications and provides readers with a comprehensive analysis of error resilience techniques proposed for contemporary video coding algorithms. Moreover, this book provides the readers with a complete coverage of the quality of service issues associated with video transmissions over mobile networks. The book addresses the techniques employed to optimise the quality of service for the provision of real-time MPEG-4 transmissions over GPRS radio links with various network conditions and different error patterns.

On the other hand, to allow different video coding algorithms to interoperate, a heterogeneous video transcoder must be employed to modify the bit stream generated by the source video coder in accordance with the syntax of the destination coder. Some heterogeneous video transcoders are enabled to operate in two or more directions allowing incompatible streams to flow across 2 or more networks for inter-network video communications. If both sender and receiver are utilising the same video coding algorithm but are yet located on dissimilar networks of different bandwidth characteristics, then a homogeneous video transcoder is required to adapt the information rate of flowing coded streams to the available bandwidth of the destination network. Hybrid video transcoding algorithms have both heterogeneous and homogeneous transcoding capabilities to adapt the transmitted coded video streams to the destination network in terms of both the end-user video decoding syntax and the destination network capacity respectively. This book addresses the technologies underpinning both the homogeneous and heterogeneous transcoding algorithms and presents the solutions proposed for the improvement of quality of service for both transcoding scenarios in error-free and error-prone environments.

1.2 Source Material

ITU has specified a number of test video sequences for use in the performance evaluation process of its proposed video coding paradigms. In this book, we focus most of the conducted tests and experiments onto six different conventional ITU test head-and-shoulder sequences to verify the study made on the performance of the presented video coding schemes and the efficiency of the corresponding error control algorithms. These test sequences have been selected to reflect a wide range of video sequences with different properties and behaviour. Foreman, Miss America, Carphone, Grandma, Suzie and Claire are the six sequences used throughout the book to conduct a large number of experiments and produce subjective and objective results. Other video sequences, such as Stefan and Harry for instance, are used only sporadically throughout the book with minor emphasis placed on their use and corresponding test results. All of the six chosen sequences represent a head-and-shoulder type of scene with different contrast and activity. Foreman is the most active scene of all since it includes a shaky background, high noise and a fair amount of bi-directional motion of the foreground object. Claire and Grandma are both typical head-and-shoulder sequences with uniform and stationary background and minimal amount of activity confined to moving lips and flickering eyelids. Both Claire and Grandma are low motion video sequences with moderate contrast and noise and a uniform background. Miss America is rather more active than Claire and Grandma with the subject once moving her shoulders before a static camera. Suzie is another head-and-shoulder video sequence with high contrast and moderate noise. It contains a fast head motion with the subject, being the foreground, holding a telephone handset with a stationary and plain-textured background. The carphone sequence shows a moving background with fair details. Though not a typical head-and-shoulder sequence, Carphone shows a talking head in a moving vehicle with more motion in the foreground object and a non-uniform changing background. All these sequences are in discrete video YUV format with a decomposition ratio 4:2:0, a QCIF (Quadrature Common Intermediate Format) resolution of 176 pixels by 144 lines and a temporal resolution (frame rate) of 25 frames per second. Figure 1.1 depicts some original frames extracted from each one of the six sequences.

1.3 Video Quality Assessment and Performance Evaluation

Since compression at low bit rates results in inevitable quality degradation, the performance of video coding algorithms must be assessed with regard to the quality of the reconstructed video sequence. Both subjective and objective methods are usually adopted to evaluate the performance of video coding algorithms. The decoded video quality can be measured by simply comparing the

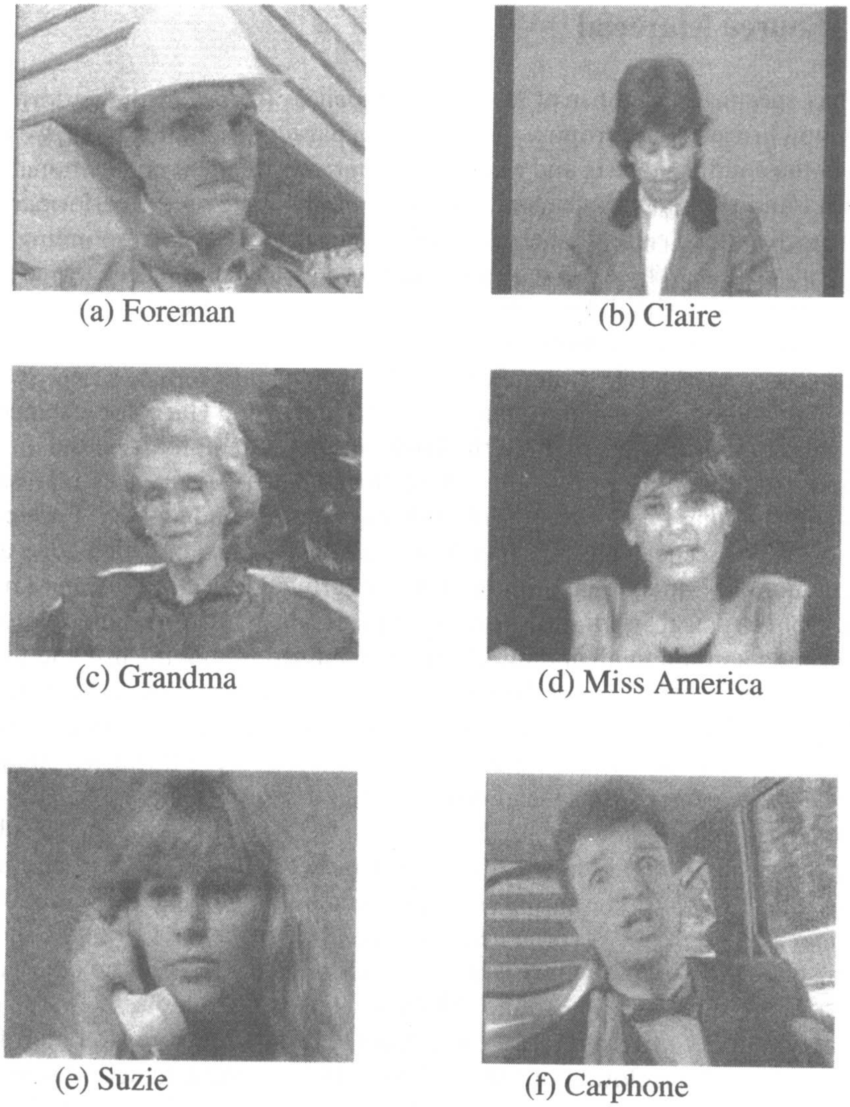


Figure 1.1 Original frames of used ITU test sequences

original and reconstructed video sequences. Although the subjective evaluation of decoded video quality is quite cumbersome compared to the calculation of numerical values for the objective quality evaluation, it is still preferable especially for low and very low bit rate compression because of the inconsistency between the existing numerical quality measurements and the Human Visual System (HVS). On the other hand, in error-prone environments, errors might corrupt the coded video stream in a way that causes a merge or split in the transmitted video frames. In this case, using the objective numerical methods to compare the original and reconstructed video sequences would incorporate some errors in associating the peer frames (corresponding frames between the two sequences) in both sequences with each other. This leads to an inaccurate evaluation of the coder performance. A subjective measurement in this case would certainly yield a fairer and more precise evaluation of the decoded video quality.

There are two broad types of subjective quality evaluation, namely rating scale methods and comparison methods (Netravali and Limb, 1980). In the first method, an overall quality rating is assigned to the image (usually the last frame of a video sequence) by using one of several given categories. In the second method, a quality impairment of a standard type is introduced to the original image until the viewer decides the impaired and reference images are of equal quality. However, throughout this book, pair comparison is used where the original sequence and decoded sequence frames are displayed side by side for subjective quality evaluation. Original sequence frames are used as reference to demonstrate the performance of a video coding algorithm in error-free environments. However, when the aim is to evaluate the performance of an error resilience technique, the original frames are then replaced by error-free decoded ones since the improvement is then intended to be shown on the error performance of the coder (decoded video quality in error-prone environments) and not on its error-free compression efficiency.

The quality of the video sequence can also be measured by using some mathematical criteria such as signal-to-noise ratio (SNR), peak-to-peak signal-to-noise ratio (PSNR) or mean-squared-error (MSE). These measurement criteria are considered to be objective due to the fact that they rely on the pixel luminance and chrominance values of the input and output video frames and do not include any subjective human intervention in the quality assessment process. For image and video, PSNR is preferred for objective measurements and is frequently used by the video coding research community, although the other two criteria are still occasionally used. PSNR and MSE are defined in Equations 1.1 and 1.2, respectively:

$$PSNR = 10\log_{10}\frac{255^2}{\frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[x(i,j) - \hat{x}(i,j)]^2} \tag{1.1}$$

$$MSE = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[x(i,j) - \hat{x}(i,j)]^2 \tag{1.2}$$

where M and N are the dimensions of the video frame in width and height respectively, and $x(i,j)$ and $\hat{x}(i,j)$ are the original and reconstructed pixel luminance or chrominance values at position (i,j).

Additionally, for a fair performance evaluation of a video coding algorithm, the bit rate must also be included. The output bit rate of a video coder is expressed in bits per second (bit/s). Since the bit rate is directly proportional to the number of pixels per frame and the number of frames coded per second, both the picture resolution and frame rate have to be indicated in the evaluation process as well. QCIF picture resolution and a frame rate of 25 frames per second have been adopted throughout the book unless otherwise specified.

1.4 Outline of the Book

The book is divided into six chapters covering the core aspects of video communication technologies. Chapter 1 presents a general historical background of the area and introduces to the reader the conventional ITU video sequences used for low bit rate video compression experiments. This chapter also discusses the conventional methods used for assessing the video quality and evaluating the performance of a video compression algorithm both subjectively and objectively.

Chapter 2 presents an overview of the core techniques employed in digital video compression algorithms with emphasis on standard techniques. The author highlights the major motivations for video compression and addresses the main issues of contemporary video coding techniques, such as model-based, segmentation-based and vector-based coders. The standardised block-transform video coders are then analysed and their performance is evaluated in terms of their quality/bit rate optimisation. A comprehensive comparison of ITU-T H.261 and H.263 is carried out in terms of their compression efficiency and robustness to errors. Emphasis is placed on the improvements brought by the latter by highlighting its performance in both the baseline and full-option modes. Then, the object-based video coding techniques are addressed in full details and particular attention is given to the ISO MPEG-4 video coding standard. The main techniques used in MPEG-4 for shape, motion and texture coding are covered, and the coder performance is evaluated in comparison to the predecessor H.263 standard. Finally, the concept of layered video coding is described and the performance of a layered video coder is analysed objectively with reference to a single layer coder for both quality and bit rates achieved.

Chapter 3 analyses the flow control mechanisms used in video communications. The factors that lead to bit rate variability in video coding algorithms are first described and alternatives to variable rate, fixed quality video coding are examined. Fixed rate video coding is then discussed by explaining several techniques used to achieve a regulated output bit rate. A variety of bit rate control algorithms are presented and their performance is evaluated using PSNR and bit rate values. Furthermore, particular attention is given to the feed-forward MB-based bit rate control algorithm which outperforms the standard-compliant rate control algorithm used in H.263 video coder. The performance of the feed-forward rate control