IAIN E. RICHARDSON

CODING VIDEO

A PRACTICAL GUIDE TO HEVC AND BEYOND

WILEY

Coding Video

Coding Video

A Practical Guide to HEVC and Beyond

Iain E. Richardson
Director, Vcodex Ltd
Delft
The Netherlands

WILEY

This edition first published 2024

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data applied for

Hardback ISBN 9781118711781

Cover Design: Wiley
Cover Image: © Aleksandra Konoplia/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Pondicherry, India

To Naomi

Contents

About the Author

Iain E. Richardson is an internationally recognised expert on video compression and digital video communications. He is the author of four other books about video coding, which include two widely cited books on the H.264 Advanced Video Coding standard. For over thirty years, he has carried out research in the field of video compression and video communications, as a Professor at the Robert Gordon University in Aberdeen, Scotland, and as an independent consultant with his own company, Vcodex. He advises companies and delivers lectures on video compression technology and is sought after as an expert witness in litigation cases involving video coding.

Acknowledgements

This book would never have left my desk without the dedication, support and encouragement of my colleague Naomi Reid, who read many drafts, acted as a sounding board and kept the manuscript moving forwards until it was complete. Thank you, Naomi! I would also like to thank the ever-amazing Pat Ballantyne for reading and commenting on the final draft and for her constant support. Thanks to the editorial team at John Wiley & Sons, to Elecard, Parabola Research and Solveig Multimedia for providing access to their visualisation tools and to everyone who has asked me to explain video coding to them over the years.

About the Companion Website

This book is accompanied by a companion website

www.wiley.com/go/richardson/codingvideo1

1

Introduction

The scene: my video coding research lab at the Robert Gordon University, Aberdeen, Scotland, 1994. 'You're working on video compression?' asked my visitor. 'Isn't that going to be unnecessary in a few years? After all, network bandwidth is increasing every year. We've already got 10 Mbps Ethernet and we'll soon have 100 Mbps. Video compression will be redundant by the year 2000'.

My visitor predicted that research into video compression would fade away as ever-increasing bandwidth and storage capacity made it possible to send uncompressed, full-quality video with ease. He offered me a wager, that my research topic, video coding, would be redundant by the year 2000. If only I had taken that bet! In fact, at the present time, the need for efficient and effective video coding is perhaps greater than ever. Generation and consumption of video content have grown more quickly than available bandwidth has in recent years. This is because of the emergence of user-generated content that comes most notably from mobile devices. Consumption has also increased through changing viewing patterns – often with multiple simultaneous streams being viewed in a single household – and the move to higher-resolution content, such as High Definition and beyond.

This book is about video coding, also known as video compression. A video encoder converts digital video into a compressed form, that is, a form that takes up less storage or transmission capacity, known as coded video. A video decoder does the opposite, converting coded video into uncompressed digital video. The combination of a video encoder and a video decoder is known as a video codec, and the processes of compression and decompression are often described as video coding.

1.1 Why Write This Book?

I have written four books about video coding. Why write another one?

It is now 2024, and video coding is embedded in modern life as never before. Compressed video made a major impact in the 1990s with the advent of digital television and DVD video. Other applications followed throughout the early 2000s, and now compressed video

Coding Video: A Practical Guide to HEVC and Beyond, First Edition. Iain E. Richardson.

Companion website: www.wiley.com/go/richardson/codingvideo1

is a fundamental part of an ever-expanding list of consumer and business applications. These include, but are not limited to:

- Recording, playing and sharing video on mobile devices.
- Video calling and videoconferencing, both of which received a massive boost following the COVID lockdowns of 2020.
- Video streaming, which is now overtaking digital TV broadcasting as the dominant way in which we watch video.
- Video in social media, with platforms such as Facebook, Instagram and TikTok increasingly relying on video as well as still images.
- Security and surveillance, from commercial security cameras to police bodycams and video-enabled doorbells.
- Automotive video, from dashcams to in-car entertainment.

Since I wrote my last book in 2010, we have seen a significant increase in the use of video compression and a shift in the way coded video is used, as well as the release of a series of new industry standards and formats for video compression. These formats include H.265/High Efficiency Video Coding (HEVC), H.266/Versatile Video Coding, VP9, AV1 and more.

These new standards and formats are publicly available documents. However, the published standards are not intended for the casual reader and can be challenging to understand. This is due in part to their focus on highly detailed technical descriptions of how standard-based video codecs work. Even a reader with a solid background in computer science or electronic engineering may need a more approachable way into the video coding standards.

I have spent much of my career explaining video coding to students, researchers, engineers and professionals. I have always tried to find approachable and intuitive explanations of how video coding works. I wrote this book primarily to gather this material in one place, and I hope that it can provide a more user-friendly guide to video coding that will enable readers to engage with the video coding standards and understand how to get the most out of video compression.

1.2 What Is in the Book?

In this book, I give an overview of digital video and video compression and a short history of some of the key concepts and standards. I then discuss each of the main core concepts of video coding. I explain each concept in general terms and discuss how each is put into practice in certain video coding standards. I have chosen to focus on the widely used H.26x series of standards, with a particular focus on H.265/HEVC. Each chapter is liberally illustrated with diagrams and examples.

In Chapter 2, I provide an overview of a video codec and introduce the concepts of performance and visual quality.

Chapter 3 is a short history of video coding. This chapter begins with a description of the basic concepts that developed between the 1950s and 1990s and goes on to describe the major video coding standards from H.261 through to H.266 and AV1.

Chapter 4 explains the structural elements used in video coding. A video codec processes a video by breaking it down into structures such as Groups of Pictures, pictures, slices, tiles, basic coding units, prediction blocks and transform blocks.

Chapters 5 and 6 deal with prediction. The information in a block of video pixels can be significantly reduced if we can predict the pixels from previously sent information. Intra prediction (see Chapter 5) involves predicting blocks using information that is available in the same frame of video. This is known as intra-frame or spatial prediction. Inter prediction (see Chapter 6) creates predictions using information that is available in other, previously coded frames. This is known as inter-frame or temporal prediction.

Chapter 7 examines transforms and quantisation. A transform converts a block of pixels, or difference samples, into another domain in which the important visual information is concentrated into a smaller number of values. Quantisation removes some of the resulting information, deliberately reducing the precision of the data in order to compress it, hopefully without sacrificing too much visual quality.

Chapter 8 looks at the final stage in a video encoder, entropy coding, in which data are converted into a compressed bitstream. I discuss the main types of entropy coding, with a particular emphasis on arithmetic coding, which has become the method of choice for entropy coding in recent video coding standards.

Chapter 9 deals with filtering, in particular the types of video filtering that are carried out as part of a video encoding or decoding process. In this context, video pixels are filtered to try to improve the performance of all the other compression processes.

In Chapter 10, I discuss how coded video can be stored and transmitted. Video is often stored together with audio in container files such as an MP4 file. Transport or transmission of video is an area that continues to develop, with the emergence of adaptive bitrate streaming as one of the most important ways of sending video across the internet.

Chapter 11 considers how video codecs are implemented in software or hardware and how the performance of video codecs can be measured and compared. I provide some suggestions for trying this yourself, for example, by experimenting with the various public-domain software video codecs and by analysing coded video to see how it is actually put together.

1.3 How Should You Use This Book?

I wrote this book with several audiences in mind. Each chapter starts by setting out basic concepts and continues into more technical detail. I have deliberately not assumed specialised background knowledge on the part of the reader. I hope that the book will be useful to:

- Students who are studying multimedia processing and communications.
- Academics and researchers, as the book presents a way to understand and approach the extensive research literature on video coding and provides a platform to develop ideas and research topics.
- Engineers and implementers, as a bridge between the basic concepts of video coding and the often challenging density of the video coding standards.
- Other professionals, as a more accessible way into this important technical topic.

I have written the book so that it can be read from start to finish, and I hope that at least some readers will do just that. Of course, you may want to dip in and out and concentrate on particular topics of interest. I would recommend reading Chapter 2, which will give an overview of how each topic fits into the overall video codec system or model. From Chapter 4 onwards, each chapter begins with an explanation of the basic concepts of the video compression process such as prediction and transform. The chapter then explains how the process is put into practice according to specific standards with a particular focus on H.265/HEVC.

At some point, you may want to go deeper into the topics presented here. For example, you might want to find out exactly how motion vectors are communicated in H.265 or H.266. There are many excellent articles in journals that can explain topics further, such as the *IEEE Transactions on Circuits and Systems for Video Technology*, which has published several special issues and special sections on H.265/HEVC [1] and a special section on H.266/Versatile Video Coding [2]. If there is any doubt or ambiguity about a particular process that has been specified in one of the standards, the standard document itself should be consulted.

Experimenting is an excellent way to deepen your understanding of a topic. Chapter 11 gives practical guidance on how to start experimenting with standard-based codecs. Examples of experiments can include changing encoding parameters to understand how they can affect the compressed video or using syntax analysers to visualise the way the video is encoded. Please visit the companion website, https://www.vcodex.com/coding-video-book, for further examples and resources.

I hope you are now ready to walk through the processes of coding video, starting with an overview of the video codec and the visual information it is designed to compress.

References

1 Gharavi, H. (2012). Combined issue on high efficiency video coding (HEVC) standard and research. *IEEE Transactions on Circuits and Systems for Video Technology* 22 (12): 1646–1646. doi: https://doi.org/10.1109/TCSVT.2012.2226073.

2 Boyce, J.M., Chen, J., Liu, S. et al. (2021). Guest editorial introduction to the special section on the VVC standard. *IEEE Transactions on Circuits and Systems for Video Technology* 31 (10): 3731–3735. doi: https://doi.org/10.1109/TCSVT.2021.3111712.

2

Video Coding and Video Quality

2.1 Introduction

Video coding or video compression bridges the gap between displayable or uncompressed video and practical methods of storing or transmitting data. Digital video is made up of a series of frames or pictures, each of which contains hundreds of thousands or millions of picture elements or pixels. Storing or transmitting video in its original, uncompressed form can require an impractically large amount of storage capacity or bandwidth. A video encoder compresses video into a smaller and more compact form known as coded video, suitable for storage or transmission. A video decoder reverses this process, extracting displayable video from coded video (see Figure 2.1). The complementary pair of an encoder and a decoder is a video codec.

In this chapter, I will introduce each of the following aspects of a video codec, before covering them in further detail throughout the book:

Inputs and outputs: The data inputs and outputs for a video codec. The input to an encoder and the output of a decoder are uncompressed or raw video data, i.e. video that can be displayed. The output of an encoder and the input to a decoder are an encoded, compressed bitstream that is suitable for storage or transmission.

Data structures: A video codec processes data in a hierarchy from a complete video sequence, e.g. a video scene or programme, through groups of coded frames and individual coded frames, down to basic coding units (Coding Tree Units [CTUs], Macroblocks [MBs] or Superblocks) and sub-blocks.

Prediction: Some or all of the data to be coded are predicted from previously coded data. We can predict a block of video data from neighbouring data in the same video frame. This is known as intra prediction. Inter prediction uses video data from previously coded frames to predict a block of video data. Once the prediction is created, it is subtracted from the original block to form a residual block, which typically contains less information than the original block.

Transform and quantisation: Blocks of video data or residual data are transformed into another form, such as a frequency domain representation, and then quantised to reduce the data precision. This has the effect of removing or discarding much of the information in a block, at the expense of a loss of visual quality.

Coding Video: A Practical Guide to HEVC and Beyond, First Edition. Iain E. Richardson.

Companion website: www.wiley.com/go/richardson/codingvideo1

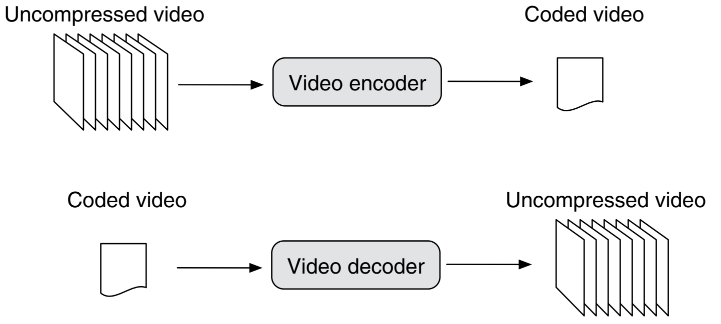

Figure 2.1 Video encoder and decoder

Bitstream coding: Quantised blocks and information needed for decoding are converted into a compressed digital bitstream, using a form of entropy coding such as variable-length coding or arithmetic coding.

Transport: A coded video bitstream is arranged in a suitable form for storage and/or transmission. This might include splitting the data into packets of a convenient size for network transmission or streaming, or placing the video data in a container file, along with other information such as an audio track.

Decoding: A video decoder reverses the encoding processes to create a video sequence that can be displayed. Depending on the encoding process, the decoded version may not be identical to the original video sequence.

Performance measurement: We can measure the performance of a video codec by considering how much it compresses video, i.e. how small the compressed bitrate is; by considering how good the decoded video looks, i.e. how much it is distorted by, compared to the original video; and by considering how much computation is required at the encoder and the decoder. These three dimensions of compression, quality and computation can be used to compare the effectiveness of different video codecs.

2.2 An Overview of Video Coding

2.2.1 Just How Much Data Are We Talking About?

Figure 2.2 shows the same frame of a video clip at four different resolutions:

1) **Standard Definition (**SD**)**: Contains 720 × 576 pixels
2) **720p Definition**[1]: Contains 1280 × 720 pixels
3) **High Definition (1080p)**: Contains 1920 × 1080 pixels
4) **Ultra-High Definition (**UHD**/4K)**: Contains 3840 × 2160 pixels.

One second of original, uncompressed SD video, captured at 25 frames per second, takes up around 15.5 Mbytes of storage space. This means that this video clip would require

1 720p is an intermediate resolution between Standard Definition and 1080p (Full HD).