Video Coding: Part II of Fundamentals of Source and Video Coding

Heiko Schwarz  
Fraunhofer Institute for Telecommunications — Heinrich Hertz Institute  
Germany  
heiko.schwarz@hhi.fraunhofer.de

Thomas Wiegand  
Berlin Institute of Technology and Fraunhofer Institute for Telecommunications — Heinrich Hertz Institute  
Germany  
thomas.wiegand@tu-berlin.de
Editorial Scope

Topics

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multirate signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing

Information for Librarians

Foundations and Trends® in Signal Processing, 2016, Volume 10, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Full text is available at 10.1561/2000000078
Video Coding: Part II of Fundamentals of Source and Video Coding

Heiko Schwarz
Fraunhofer Institute for Telecommunications — Heinrich Hertz Institute
Germany
heiko.schwarz@hhi.fraunhofer.de

Thomas Wiegand
Berlin Institute of Technology and Fraunhofer Institute for Telecommunications — Heinrich Hertz Institute
Germany
thomas.wiegand@tu-berlin.de
## Contents

1 Introduction  
1.1 The Video Communication Problem  
1.2 Scope and Overview of the Text  

2 Acquisition, Representation, Display, and Perception  
2.1 Fundamentals of Image Formation  
2.2 Visual Perception  
2.3 Representation of Digital Images and Video  
2.4 Image Acquisition  
2.5 Display of Images and Video  
2.6 Chapter Summary  

3 Video Coding Overview  
3.1 Properties of Digital Video Signals  
3.2 Intra-Picture Coding  
3.3 Hybrid Video Coding  
3.4 Coding Efficiency  
3.5 Chapter Summary  

4 Video Encoder Control  
4.1 Encoder Control using Lagrange Multipliers  
4.2 Lagrangian Optimization in Hybrid Video Encoders  

Full text is available at 10.1561/2000000078
Abstract

Digital video coding technologies have become an integral part of the way we create, communicate, and consume visual information. In the first part of this two-part text, we introduced the fundamental source coding techniques entropy coding, quantization, prediction, and transform coding. The present second part describes the application of these techniques to video coding. We introduce the basic design of hybrid video encoders and decoders, explain the basic concepts of intra-picture coding, motion-compensated prediction, and prediction error coding, and discuss encoder optimization techniques. Special emphasis is put on a fair analysis of various design aspects and coding tools in terms of coding efficiency. We highlight the application of the discussed concepts in modern video coding standards and compare important standards with respect to the achievable coding efficiency.


DOI: 10.1561/2000000078.
The application areas of digital video today range from multimedia messaging, video telephony, and video conferencing over mobile television, wireless and wired Internet video streaming, standard- and high-definition television broadcasting, subscription and pay-per-view services to personal video recorders, digital camcorders, and optical storage media such as the digital versatile disc (DVD) or the Blu-ray disc. Ultra-high definition (UHD) television sets, with a resolution of $3840 \times 2160$ image points, four times as many as high-definition screens, have recently become available for end consumers. Due to its horizontal resolution, this UHD format is often also called 4K format. Internet streaming providers have started to produce and deliver content in 4K. At the time of writing this text, the first satellite broadcasters are testing a new UHD infrastructure; the first UHD demo channels are already on air [230] and the first sport events and concerts have been successfully broadcasted live in 4K [231].

One of the key techniques that enabled the variety of digital video applications is video coding, also called video compression. Even though the main task of video coding is to compress visual information so that it requires as little bit rate as possible, the availability of advanced
video coding techniques also enables a number of new applications. For example, the availability of the improved video coding standard H.264 | MPEG-4 AVC [121] was a driving factor for the broad introduction of high-definition television (HDTV) and several video streaming services. Future video services and applications will benefit from the even more efficient standard H.265 | MPEG-H HEVC [123]. In fact, the percentage of the Internet traffic that is caused by the transmission of compressed video data increases continuously. According to a study by Cisco [48], 66% of the bits transmitted in the consumer Internet in 2013 were video data. For 2018, an increase to 79% is predicted.

Even though the various applications of digital video differ in the spatial resolution, the required compression ratio, and the acceptable video quality, the same basic video compression principles are employed. In the following, we give an overview of the main elements of a video communication chain, from the capturing of pictures to display and human perception. The main focus of the present text lies however on the description of the fundamental principles of video compression. In that context, we will also discuss the improvements in video coding technology that led to a continuous increase of coding efficiency from one generation of video coding standards to the next.

### 1.1 The Video Communication Problem

The most important processing steps of a typical video transmission system are illustrated in the block diagram of Figure 1.1. At the beginning of the signal processing chain, a digital video signal is captured by a camera. The lens of the camera projects an image of a 3D scene onto the surface of an image sensor, which samples the optical signal. The resulting raw data samples are further processed inside the camera and transformed into a representation format. The video signal eventually delivered by the camera consists of arrays of discrete-amplitude samples. An optional preprocessing step may be applied, for example, for improving the contrast and color representation or for reducing the noise; the latter has typically a beneficial effect on the following coding step. The video encoder maps the sample arrays of the representation...
format into a so-called *bitstream*, which usually has a much lower bit rate than the raw data samples.

In the simplest case, the video bitstream generated by the encoder is *stored*, typically inside a container format such as the Audio Video Interleave (AVI), Quicktime, or Matroska format. In most applications, however, the compressed video is transmitted. The transmission chain usually consists of a channel encoder, a modulator, the actual physical transmission channel, a demodulator, and a channel decoder. The *channel encoder* extends the bitstream by adding structured redundancy suitable for detecting or correcting potential transmission errors at the receiver side. The *modulator* maps the resulting data stream into an analog signal, which is transmitted over the physical *channel*. At the receiver side, the *demodulator* extracts a digital data stream from the received analog signal. The *channel decoder* produces the received bitstream by detecting and correcting transmission errors and extracting the video data packets from the data stream. Note that if not all transmission errors can be corrected, the received bitstream is not identical to the bitstream generated by the video encoder.
1.1. The Video Communication Problem

The video decoder reconstructs the sample arrays from the received bitstream or, alternatively, the bitstream read from a storage device. Optionally, the decoded signal may be postprocessed in order to reduce the impact of transmission errors and coding artifacts on the subjective video quality. At the end of the communication chain, the video signal is typically displayed and perceived by human beings.

This monograph focuses on the video encoder and video decoder, which are often summarized under the term video codec. The encoder maps the samples of the original video pictures to a set of so-called coding parameters and writes these coding parameters to a bitstream. The bitstream represents the input video in compressed form and is transmitted to the video decoder. The format in which the coding parameters are written to the bitstream is referred to as bitstream syntax. It has to be known to both the encoder and the decoder. The video decoder parses the received bitstream according to the given bitstream syntax and thereby decodes the transmitted coding parameters. Finally, the video pictures are reconstructed by following a defined decoding process, which is controlled by the transmitted coding parameters.

For achieving the required transmission bit rates, video codecs apply lossy coding algorithms. Hence, even in error-free transmission scenarios, the digital video signal reconstructed by the decoder is different from the encoder’s input signal. Since, in most applications, the coded video is perceived by human beings, the degradation of the perceived video quality should be as small as possible. The basic video coding problem can be stated as representing a video signal with the highest possible subjective quality without exceeding an available bit rate or, alternatively, as conveying the video signal with the lowest possible bit rate while maintaining a specified subjective quality. In practice, the subjective quality of a video signal is very hard to specify and therefore objective distortion measures calculated based on the differences between the original and decoded sample values are often used instead. The ability of a codec to choose a suitable trade-off between signal distortion and bit rate is referred to as its coding efficiency. Besides the coding efficiency, the applicability of a video codec for a certain communication scenario is also influenced by the implementation com-
plexity of the algorithms used as well as the structural and processing delay of the codec, which determine, among other factors, the crucial end-to-end delay between capturing and displaying a video picture.

In the most common setup of video codecs, the video decoder merely extracts the coding parameters from the received bitstream and follows a defined decoding process for reconstructing the video pictures. Given a particular bitstream syntax and decoding process, all decoder implementations generate the same (or, sometimes, nearly the same) video pictures. The achievable coding efficiency is limited by the set of syntax features and coding tools that are supported in the bitstream syntax and the decoding process. However, the actual coding efficiency of a bitstream is highly dependent on the encoder implementation. For a given bitstream syntax and decoding process, both the bit rate and the reconstruction quality are determined by the encoding algorithm that maps the original pictures into a sequence of coding parameters.

1.2 Scope and Overview of the Text

The present text provides a description of the fundamentals concepts of video coding. It is aimed at aiding students and engineers to investigate the subject. Since the topic of video coding and video communication is too broad and too deep to exhaustively describe all its aspects in the chosen presentation format, we concentrate on the signal processing in video encoders and decoders. This also means that we will leave out a number of areas, including software and hardware implementation aspects, the topics of pre- and postprocessing, and the whole field of video transmission and error-robust coding.

The intention of this text is to provide an in-depth treatment of the basic principles and coding tools found in modern video codecs. Subjects that we consider particularly important will be covered in greater detail. For giving examples and analyzing certain coding tools, we will often refer to video coding standards. These standards do not only represent the dominant technology for real-world applications, but also reflect the state-of-the-art in the field of video coding. In fact, many advanced coding tools have been developed in international standard-
Chapter 1: Scope and Overview of the Text

1.2. Scope and Overview of the Text

Video coding approaches that are less relevant for practical applications, such as 3D subband coding or distributed video coding, will not be explained in this text. Moreover, we will neither discuss scalable nor 3D video coding. These coding schemes typically represent extensions of conventional video coding designs. Even though they include some additional coding tools, the same fundamental concepts as in conventional video coding are employed.

The monograph is divided into two parts. In the first part [301], we introduced the fundamental source coding techniques entropy coding, quantization, prediction, and transform coding and analyzed their coding efficiency based on simple models for 1D random signals. The present second part describes the application of these techniques to video coding. We describe the basic structure of video codecs, discuss the fundamental concepts of video coding, and highlight their application in modern video coding standards. The effectiveness of various coding tools will be demonstrated based on experimental results.

Section 2 gives an overview of the acquisition, representation, and display of video signals. It describes the raw data formats used in video coding applications and highlights the relationship between the acquisition, representation, and display of video signals and the way we perceive visual information. Section 3 introduces the basic principles of hybrid video coding and describes the structure of typical video encoders and decoders. It further introduces the measures that we will use for comparing the coding efficiency of different codecs. Section 4 describes the concept of a Lagrangian encoder control, which we will use in all coding experiments. The usage of a unified and highly effective encoder control allows us to fairly compare different coding tools in terms of coding efficiency. Section 5 discusses the application of transform coding in video codecs and introduces techniques for intra-picture coding. Section 6 describes coding tools for inter-picture coding. It introduces the concept of motion-compensated prediction and analyzes several design aspects in terms of coding efficiency. Section 7 compares important video coding standards with respect to their coding efficiency. A summary of important results is given in Section 8.


References


References


References


[71] A. Fuldseth, G. Bjøntegaard, M. Budagavi, and V. Sze. CE10: Core transform design for HEVC. Joint Collaborative Team on Video Coding, doc. JCTVC-G495, November 2011.


References


References


Full text is available at 10.1561/2000000078
References


References


References


References


References


References


References


References


References


References


References


References


References


References


Full text is available at 10.1561/2000000078


