

# CHAPTER 1



# Introduction

Over the past decade, countless multimedia functionalities have been added to mobile devices. For example, front and back video cameras are common features in today's cellular phones. Further, there has been a race to capture, process, and display ever-higher resolution video, making this an area that vendors emphasize and where they actively seek market differentiation. These multimedia applications need fast processing capabilities, but those capabilities come at the expense of increased power consumption. The battery life of mobile devices has become a crucial factor, whereas any advances in battery capacity only partly address this problem. Therefore, the future's winning designs must include ways to reduce the energy dissipation of the system as a whole. Many factors must be weighed and some tradeoffs must be made.

Granted, high-quality digital imagery and video are significant components of the multimedia offered in today's mobile devices. At the same time, there is high demand for efficient, performance- and power-optimized systems in this resource-constrained environment. Over the past couple of decades, numerous tools and techniques have been developed to address these aspects of digital video while also attempting to achieve the best visual quality possible. To date, though, the intricate interactions among these aspects had not been explored.

In this book, we study the concepts, methods, and metrics of digital video. In addition, we investigate the options for tuning different parameters, with the goal of achieving a wise tradeoff among visual quality, performance, and power consumption. We begin with an introduction to some key concepts of digital video, including visual data compression, noise, quality, performance, and power consumption. We then discuss some video compression considerations and present a few video coding usages and requirements. We also investigate the tradeoff analysis—the metrics for its good use, its challenges and opportunities, and its expected outcomes. Finally, there is an introductory look at some emerging applications. Subsequent chapters in this book will build upon these fundamental topics.

# The Key Concepts

This section deals with some of the key concepts discussed in this book, as applicable to perceived visual quality in compressed digital video, especially as presented on contemporary mobile platforms.

## Digital Video

The term *video* refers to the visual information captured by a camera, and it usually is applied to a time-varying sequence of pictures. Originating in the early television industry of the 1930s, video cameras were electromechanical for a decade, until all-electronic versions based on cathode ray tubes (CRT) were introduced. The analog tube technologies were then replaced in the 1980s by solid-state sensors, particularly CMOS active pixel sensors, which enabled the use of digital video.

Early video cameras captured analog video signals as a one-dimensional, time-varying signal according to a pre-defined scanning convention. These signals would be transmitted using analog amplitude modulation, and they were stored on analog video tapes using video cassette recorders or on analog laser discs using optical technology. The analog signals were not amenable to compression; they were regularly converted to digital formats for compression and processing in the digital domain.

Recently, use of all-digital workflow encompassing digital video signals from capture to consumption has become widespread, particularly because of the following characteristics:

- It is easy to record, store, recover, transmit, and receive, or to process and manipulate, video that's in digital format; it's virtually without error, so digital video can be considered just another data type for today's computing systems.
- Unlike analog video signals, digital video signals can be compressed and subsequently decompressed. Storage and transmission are much easier in compressed format compared to uncompressed format.
- With the availability of inexpensive integrated circuits, high-speed communication networks, rapid-access dense storage media, advanced architecture of computing devices, and high-efficiency video compression techniques, it is now possible to handle digital video at desired data rates for a variety of applications on numerous platforms that range from mobile handsets to networked servers and workstations.

Owing to a high interest in digital video, especially on mobile computing platforms, it has had a significant impact on human activities; this will almost certainly continue to be felt in the future, extending to the entire area of information technology.

## Video Data Compression

It takes a massive quantity of data to represent digital video signals. Some sort of data compression is necessary for practical storage and transmission of the data for a plethora of applications. Data compression can be *lossless*, so that the same data is retrieved upon decompression. It can also be *lossy*, whereby only an approximation of the original signal is recovered after decompression. Fortunately, the characteristic of video data is such that a certain amount of loss can be tolerated, with the resulting video signal perceived without objection by the human visual system. Nevertheless, all video signal-processing methods and techniques make every effort to achieve the best visual quality possible, given their system constraints.

Note that video data compression typically involves coding of the video data; the coded representation is generally transmitted or stored, and it is decoded when a decompressed version is presented to the viewer. Thus, it is common to use the terms *compression/decompression* and *encoding/decoding* interchangeably. Some professional video applications may use uncompressed video in coded form, but this is relatively rare.

A *codec* is composed of an encoder and a decoder. Video encoders are much more complex than video decoders are. They typically require a great many more signal-processing operations; therefore, designing efficient video encoders is of primary importance. Although the video coding standards specify the bitstream syntax and semantics for the decoders, the encoder design is mostly open.

Chapter 2 has a detailed discussion of video data compression, while the important data compression algorithms and standards can be found in Chapter 3.

## Noise Reduction

Although compression and processing are necessary for digital video, such processing may introduce undesired effects, which are commonly termed *distortions* or *noise*. They are also known as *visual artifacts*. As noise affects the fidelity of the user's received signal, or equivalently the visual quality perceived by the end user, the video signal processing seeks to minimize the noise. This applies to both analog and digital processing, including the process of video compression.

In digital video, typically we encounter many different types of noise. These include noise from the sensors and the video capture devices, from the compression process, from transmission over lossy channels, and so on. There is a detailed discussion of various types of noise in Chapter 4.

## Visual Quality

Visual quality is a measure of perceived visual deterioration in the output video compared to the original signal, which has resulted from lossy video compression techniques. This is basically a measure of the *quality of experience* (QoE) of the viewer. Ideally, there should be minimal loss to achieve the highest visual quality possible within the coding system.

Determining the visual quality is important for analysis and decision-making purposes. The results are used in the specification of system requirements, comparison and ranking of competing video services and applications, tradeoffs with other video measures, and so on.

Note that because of compression, the artifacts found in digital video are fundamentally different from those in analog systems. The amount and visibility of the distortions in video depend on the contents of that video. Consequently, the measurement and evaluation of artifacts, and the resulting visual quality, differ greatly from the traditional analog quality assessment and control mechanisms. (The latter, ironically, used signal parameters that could be closely correlated with perceived visual quality.)

Given the nature of digital video artifacts, the best method of visual quality assessment and reliable ranking is subjective viewing experiments. However, subjective methods are complex, cumbersome, time-consuming, and expensive. In addition, they are not suitable for automated environments.

An alternative, then, is to use simple error measures such as the *mean squared error* (MSE) or the *peak signal to noise ratio* (PSNR). Strictly speaking, PSNR is only a measure of the signal fidelity, not the visual quality, as it compares the output signal to the input signal and so does not necessarily represent perceived visual quality. However, it is the most popular metric for visual quality used in the industry and in academia. Details on this use are provided in Chapter 4.

## Performance

Video coding *performance* generally refers to the speed of the video coding process: the higher the speed, the better the performance. In this context, *performance optimization* refers to achieving a fast video encoding speed.

In general, the performance of a computing task depends on the capabilities of the processor, particularly the *central processing unit* (CPU) and the *graphics processing unit* (GPU) frequencies up to a limit. In addition, the capacity and speed of the main memory, auxiliary cache memory, and the disk input and output (I/O), as well as the cache hit ratio, scheduling of the tasks, and so on, are among various system considerations for performance optimization.

Video data and video coding tasks are especially amenable to parallel processing, which is a good way to improve processing speed. It is also an optimal way to keep the available processing units busy for as long as necessary to complete the tasks, thereby maximizing resource utilization. In addition, there are many other performance-optimization techniques for video coding, including tuning of encoding parameters. All these techniques are discussed in detail in Chapter 5.

## Power Consumption

A mobile device is expected to serve as the platform for computing, communication, productivity, navigation, entertainment, and education. Further, devices that are implantable to human body, that capture intrabody images or videos, render to the brain, or securely transmit to external monitors using biometric keys may become available in the future. The interesting question for such new and future uses would be how these devices can be supplied with power. In short, leaps of innovation are necessary in this area. However, even while we await such breakthroughs in power supply, know that some externally wearable devices are already complementing today's mobile devices.

Power management and optimization are the primary concerns for all these existing and new devices and platforms, where the goal is to prolong battery life. However, many applications are particularly power-hungry, either by their very nature or because of special needs, such as on-the-fly binary translation.

Power—or equivalently, energy—consumption thus is a major concern. Power optimization aims to reduce energy consumption and thereby extend battery life. High-speed video coding and processing present further challenges to power optimization. Therefore, we need to understand the power management and optimization considerations, methods, and tools; this is covered in Chapters 6 and 7.

## Video Compression Considerations

A major drawback in the processing, storage, and transmission of digital video is the huge amount of data needed to represent the video signal. Simple scanning and binary coding of the camera voltage variations would produce billions of bits per second, which without compression would result in prohibitively expensive storage or transmission devices. A typical high-definition video (three color planes per picture, a resolution of  $1920 \times 1080$  pixels per plane, 8 bits per pixel, at a 30 pictures per second rate) necessitates a data rate of approximately 1.5 billion bits per second. A typical transmission channel capable of handling about 5 Mbps would require a 300:1 compression ratio. Obviously, lossy techniques can accommodate such high compression, but the resulting reconstructed video will suffer some loss in visual quality.

However, video compression techniques aim at providing the best possible visual quality at a specified data rate. Depending on the requirements of the applications, available channel bandwidth or storage capacity, and the video characteristics, a variety of data rates are used, ranging from 33.6 kbps video calls in an old-style public switched telephone network to ~20 Mbps in a typical HDTV rebroadcast system.

## Varying Uses

In some video applications, video signals are captured, processed, transmitted, and displayed in an on-line manner. Real-time constraints for video signal processing and communication are necessary for these applications. The applications use an end-to-end real-time workflow and include, for example, video chat and video conferencing, streaming, live broadcast, remote wireless display, distant medical diagnosis and surgical procedures, and so on.

A second category of applications involve recorded video in an off-line manner. In these, video signals are recorded to a storage device for archiving, analysis, or further processing. After being used for many years, the main storage medium for the recorded video is shifted from analog video tapes to digital *DV* or *Betacam* tapes, optical discs, hard disks, or flash memory. Apart from archiving, stored video is used for off-line processing and analysis purposes in television and film production, in surveillance and monitoring, and in security and investigation areas. These uses may benefit from video signal processing as fast as possible; thus, there is a need to speed up video compression and decompression processes.

## Conflicting Requirements

The conflicting requirements of video compression on modern mobile platforms pose challenges for a range of people, from system architects to end users of video applications. Compressed data is easy to handle, but visual quality loss typically occurs with compression. A good video coding solution must produce videos without too much loss of quality.

Furthermore, some video applications benefit from high-speed video coding. This generally implies a high computation requirement, resulting in high energy consumption. However, mobile devices are typically resource constrained and battery life is usually the biggest concern. Some video applications may sacrifice visual quality in favor of saving energy.

These conflicting needs and purposes have to be balanced. As we shall see in the coming chapters, video coding parameters can be tuned and balanced to obtain such results.

## Hardware vs. Software Implementations

Video compression systems can be implemented using dedicated application-specific integrated circuits (ASICs), field-programmable gate arrays (FPGAs), GPU-based hardware acceleration, or purely CPU-based software.

The ASICs are customized for a particular use and are usually optimized to perform specific tasks; they cannot be used for purposes other than what they are designed for. Although they are fast, robust against error, yield consistent, predictable, and offer stable performance, they are inflexible, implement a single algorithm, are not programmable or easily modifiable, and can quickly become obsolete. Modern ASICs often include entire microprocessors, memory blocks including read-only memory (ROM), random-access memory (RAM), flash memory, and other large building blocks. Such an ASIC is often termed a system-on-chip (SoC).

FPGAs consist of programmable logic blocks and programmable interconnects. They are much more flexible than ASICs; the same FPGA can be used in many different applications. Typical uses include building prototypes from standard parts. For smaller designs or lower production volumes, FPGAs may be more cost-effective than an ASIC design. However, FPGAs are usually not optimized for performance, and the performance usually does not scale with the growing problem size.

Purely CPU-based software implementations are the most flexible, as they run on general-purpose processors. They are usually portable to various platforms. Although several performance-enhancement approaches exist for the software-based implementations, they often fail to achieve a desired performance level, as hand-tuning of various parameters and maintenance of low-level codes become formidable tasks. However, it is easy to tune various encoding parameters in software implementations, often in multiple passes. Therefore, by tuning the various parameters and number of passes, software implementations can provide the best possible visual quality for a given amount of compression.

GPU-based hardware acceleration typically provides a middle ground. In these solutions, there are a set of programmable execution units and a few performance- and power-optimized fixed-function hardware units. While some complex algorithms may take advantage of parallel processing using the execution units, the fixed-function units provide fast processing. It is also possible to reuse some fixed-function units with updated parameters based on certain feedback information, thereby achieving multiple passes for those specific units. Therefore, these solutions exhibit flexibility and scalability while also being optimized for performance and power consumption. The tuning of available parameters can ensure high visual quality at a given bit rate.

## Tradeoff Analysis

Tradeoff analysis is the study of the cost-effectiveness of different alternatives to determine where benefits outweigh costs. In video coding, a tradeoff analysis looks into the effect of tuning various encoding parameters on the achievable compression, performance, power savings, and visual quality in consideration of the application requirements, platform constraints, and video complexity.

Note that the tuning of video coding parameters affects performance as well as visual quality, so a good video coding solution balances performance optimization with achievable visual quality. In Chapter 8, a case study illustrates this tradeoff between performance and quality.

It is worthwhile to note that, while achieving high encoding speed is desirable, it may not always be possible on platforms with different restrictions. In particular, achieving power savings is often the priority on modern computing platforms. Therefore, a typical tradeoff between performance and power optimization is considered in a case study examined in Chapter 8.

## Benchmarks and Standards

The benchmarks typically used today for ranking video coding solutions do not consider all aspects of video. Additionally, industry-standard benchmarks for methodology and metrics specific to tradeoff analysis do not exist. This standards gap leaves the user guessing about which video coding parameters will yield satisfactory outputs for particular video applications. By explaining the concepts, methods, and metrics involved, this book helps readers understand the effects of video coding parameters on the video measures.

## Challenges and Opportunities

Several challenges and opportunities in the area of digital video techniques have served as the motivating factors for tradeoff analysis.

- The demand for compressed digital video is increasing. With the desire to achieve ever-higher resolution, greater bit depth, higher dynamic range, and better quality video, the associated computational complexity is snowballing. These developments present a challenge for the algorithms and architectures of video coding systems, which need to be optimized and tuned for higher compression but better quality than standard algorithms and architectures.

- Several international video coding standards are now available to address a variety of video applications. Some of these standards evolved from previous standards, were tweaked with new coding features and tools, and are targeted toward achieving better compression efficiency.
- Low-power computing devices, particularly in the mobile environment, are increasingly the chosen platforms for video applications. However, they remain restrictive in terms of system capabilities, a situation that presents optimization challenges. Nonetheless, tradeoffs are possible to accommodate goals such as preserving battery life.
- Some video applications benefit from increased processing speed. Efficient utilization of resources, resource specialization, and tuning of video parameters can help achieve faster processing speed, often without compromising visual quality.
- The desire to obtain the best possible visual quality on any given platform requires careful control of coding parameters and wise choice among many alternatives. Yet there exists a void where such tools and measures should exist.
- Tuning of video coding parameters can influence various video measures, and desired tradeoffs can be made by such tuning. To be able to balance the gain in one video measure with the loss in another requires knowledge of coding parameters and how they influence each other and the various video measures. However, there is no unified approach to the considerations and analyses of the available tradeoff opportunities. A systematic and in-depth study of this subject is necessary.
- A tradeoff analysis can expose the strengths and weaknesses of a video coding solution and can rank different solutions.

## The Outcomes of Tradeoff Analysis

Tradeoff analysis is useful in many real-life video coding scenarios and applications. Such analysis can show the value of a certain encoding feature so that it is easy to make a decision whether to add or remove that feature under the specific application requirements and within the system restrictions. Tradeoff analysis is useful in assessing the strengths and weaknesses of a video encoder, tuning the parameters to achieve optimized encoders, comparing two encoding solutions based on the tradeoffs they involve, or ranking multiple encoding solutions based on a set of criteria.

It also helps a user make decisions about whether to enable some optional encoding features under various constraints and application requirements. Furthermore, a user can make informed product choices by considering the results of the tradeoff analysis.



# Emerging Video Applications

Compute performance has increased to a level where computers are no longer used solely for scientific and business purposes. We have a colossal amount of compute capabilities at our disposal, enabling unprecedented uses and applications. We are revolutionizing human interfaces, using vision, voice, touch, gesture, and context. Many new applications are either already available or are emerging for our mobile devices, including perceptual computing, such as 3-D image and video capture and depth-based processing; voice, gesture, and face recognition; and virtual-reality-based education and entertainment.

These applications are appearing in a range of devices and may include synthetic and/or natural video. Because of the fast pace of change in platform capabilities, and the innovative nature of these emerging applications, it is quite difficult to set a strategy on handling the video components of such applications, especially from an optimization point of view. However, by understanding the basic concepts, methods, and metrics of various video measures, we'll be able to apply them to future applications.

## Summary

This chapter discussed some key concepts related to digital video, compression, noise, quality, performance, and power consumption. It presented various video coding considerations, including usages, requirements, and different aspects of hardware and software implementations. There was also a discussion of tradeoff analysis and the motivations, challenges, and opportunities that the field of video is facing in the future. This chapter has set the stage for the discussions that follow in subsequent chapters.