Cambridge University Press
978-0-521-11432-5 - Compression for Multimedia
Irina Bocharova
Excerpt
More information

1 Introduction

According to wikipedia[1] "*multimedia is the use of several different media (e.g. text, audio, graphics, animation, video, and interactivity) to convey information*." In a more narrow sense, multimedia is a set of software and hardware means used to create, store, and transmit information presented in various digital formats. Although multimedia data is a general term referring to any type of information, we will focus on multimedia data such as speech, images, audio, and video that are originally analog waveforms.

Due to the dramatic progress in microelectronic technologies during the last decades, TV, photography, sound and video recording, communication systems etc., which came into the world and during at least half of the previous century were developed as analog systems, have been almost completely replaced by digital systems. At the same time numerous digital areas and systems such as video conferencing via the Internet, IP-telephony, multi-user games etc. using digitized speech, images, audio, and video appeared as well. Relatively recently multimedia computer technologies started to penetrate into education, medicine, scientific research, entertainment, advertising, and marketing, as well as into many other universally important areas.

Everything mentioned above motivates a deep study of multimedia compression and intensive research in this area. In order to use analog multimedia signals in digital systems it is necessary to solve two main problems. The first problem, related to using these kinds of signal in digital systems, is how to convert them into digital forms. However, it is not enough simply to digitize them. The number of bits required to store images, audio or video signals converted into digital form is so large that this circumstance limits the efficiency of the corresponding digital systems. Thus, the second problem is to compress multimedia data in order to transmit them faster and to store them more efficiently.

Typically, digitizing *multimedia signals* with a high precision results in large files containing the obtained *multimedia data*. Surely, the exact meaning of the words "large file" or "small file" depends on the level of existing microelectronic technologies. About 20 years ago when multimedia compression was not an everyday attribute of our lives, a floppy disk of size 1.44 Mbytes was a typical storage medium. Data files of size exceeding one diskette were considered as "huge" files at that time. It seemed completely impossible to store, on any kind of existing memory, for example, a digitized

[1] http://en.wikipedia.org/wiki/Multimedia

color image of size 1408 × 1152 pixels. Each pixel of such an image is represented by 3 bytes and thus the image requires 4.9 Mbytes of memory for storing. Transmitting of a color image of size 288 × 352 through the Plain Old Telephone Service (POTS) networks also looked extremely impractical. Since POTS networks were originally designed for the transmission of analog data they needed a so-called modem to convert the digital data to be transmitted into analog form. It is easy to compute that transmitting $288 \times 352 \times 24 = 2.4$ Mb through the telephone channel using a standard modem with transmitting rate equal to 33.6 kb/s requires approximately 1 min (72 s).

Nowadays a variety of storage devices of large capacity are offered by different companies. Transmitting images through POTS networks also became a thing of the past. New kinds of wideband channel are introduced. In the late nineties Digital Subscriber Line (DSL) modems were developed. They are used to communicate over the same twisted-pair metallic cable as used by telephone networks. Using not a voice channel band but the actual range of frequencies supportable on a twisted-pair circuit, they provide transmitting rates exceeding 1 Mb/s. So, it might seem that compression will not be needed in the future. But this is not the case. Together with increasing storage and channel capacities our requirements of the quality of digital multimedia also increase. First of all, during the last decade typical resolutions of images and video became significantly higher. For example, 2–4 Mpixel digital photocameras are replaced by 8–10 Mpixel cameras. A color picture taken by a 10 Mpixel camera requires 30 Mbytes of memory for storing, i.e. only 66 uncompressed pictures can be stored on a Compact-Flash (CF) memory of rather large size, 2 Gbytes, say. For this reason each photocamera has an embedded image compression algorithm. Moreover, the majority of photocameras do not have a mode which allows us to store the uncompressed image.

Let us consider another example. One second of video film with resolution 480 × 720 pixels recorded with 30 frames/s requires approximately 31 Mbytes of memory. It means that only 21 s of this film can be recorded on a 650 Mbytes Compact Disc (CD) without compression. However, neither does using 15.9 Gbyte Digital Versatile Disc (DVD) solve the problem of storing video data since only 8.5 minutes of video film with such resolution and frame rate can be recorded on the DVD of such a capacity. As for transmitting high-resolution video through communication channels, it is still an even more complicated problem than storing it. For example, it takes 4 s to transmit a color image of size 480 × 720 pixels by using the High Data rate DSL (HDSL) modem with rate 2 Mb/s. It means that a film with such a frame size can be transmitted with frame rate equal to 0.25 frame/s only.

The considered examples show that actually it does not matter how much we can increase telecommunication bandwidth or disk storage capacity, there will always remain a need to compress multimedia data in order to transmit them faster and to store them more efficiently.

Multimedia technologies continuously find new applications that create new problems in the multimedia compression field. Recently, new tendencies in multimedia compression have arisen. One of many newly intensively developed areas is Digital Multimedia Broadcasting (DMB), often called "mobile TV." This technology is used in order to send multimedia data to mobile phones and laptops. The world's first DMB service appeared in South Korea in 2005. This line of development requires a significant revision of existing compression algorithms in order to tailor them to the specific needs of broadcasting systems.

The Depth Image Based Rendering (DIBR) technique for three-dimensional television (3D-TV) systems is another quickly developed area that is closely related to multimedia compression. Each 3D image is represented as the corresponding 2D image and an associated per-pixel "depth" information. As a result, the number of bits to represent a 3D image drastically increases. It requires modifications of known 2D compression techniques in order to efficiently compress 3D images and video.

It is, needless to say, about a variety of portable devices such as Personal Digital Assistant (PDA), smartphone, Portable Media Player (iPOD), and many others intended for loading multimedia contents. Each such device requires a new compression algorithm taking into account its specific features.

Multimedia compression systems can be split into two large classes. The first class is *lossless compression systems*. With lossless compression techniques the original file can be recovered exactly, bit by bit after compression and decompression. To do this we usually use well-known methods of discrete source coding such as Huffman coding, arithmetic coding, or coding based on the Ziv–Lempel algorithms.

Lossy compression implies that we remove or reduce the redundancy of the multimedia data at the cost of changing or distorting the original file. These methods exploit the tradeoff of compression versus distortion. Among lossy techniques are compression techniques based on transform coding (coding of the discrete Fourier transform or discrete cosine transform coefficients, wavelet filtering), predictive coding etc.

In this book lossy compression techniques and their applications to image, video, speech, and audio compression are considered. The book provides rather deep knowledge of lossy compression systems. Modern multimedia compression techniques are analyzed and compared in terms of achieving known theoretical limits. Some implementation issues important for the efficient implementation of existing multimedia compression standards are discussed also.

The book is intended for undergraduate students. The required prerequisite is an elementary knowledge of linear systems, Fourier transforms, and signal processing. Some prior knowledge of information theory and random processes would be useful. The book can be also recommended for graduate students with an interest in compression techniques for multimedia.

The book consists of 10 chapters and an Appendix. In Chapters 2 and 3 basic theoretical aspects of source coding with fidelity criteria are given. In particular, in Chapter 3 the notion of the rate-distortion function is introduced. This function is a theoretical limit for achievable performances of multimedia systems. This chapter can be recommended for readers who are doing research in the multimedia compression area. In order to compare the performances of different quantizers some results of the high-resolution quantization theory are given in Section 3.3. This section can be omitted for readers who have no prior knowledge of information theory.

In Chapters 4, 5, and 6, commonly used coding techniques are described. Chapter 4 is devoted to linear predictive coding. It begins in Section 4.1 with descriptions of discrete-time filters by different means which are presented for the sake of completeness and can be omitted by readers familiar with this subject.

Chapters 7, 8, 9, and 10 are devoted to modern standards for speech, image, video, and audio compression, respectively. For readers who are not familiar with lossless coding techniques, the basics of lossless coding are briefly overviewed in the Appendix.

Cambridge University Press
978-0-521-11432-5 - Compression for Multimedia
Irina Bocharova
Excerpt
More information

2 Analog to digital conversion

Analog to digital transformation is the first necessary step to load multimedia signals into digital devices. It contains two operations called sampling and quantization. The theoretical background of sampling is given by the famous sampling theorem. The first attempts to formulate and prove the sampling theorem date back to the beginning of the twentieth century. In this chapter we present Shannon's elegant proof of the sampling theorem. Consequences of sampling "too slowly" in the time and frequency domains are discussed. Quantization is the main operation which determines the quality–compression ratio tradeoff in all lossy compression systems. We consider different types of quantizer commonly used in modern multimedia compression systems.

2.1 Analog and digital signals

First, we introduce some definitions.

- A function $f(x)$ is *continuous* at a point $x = a$ if $\lim_{x \to a} f(x) = f(a)$. We say a function is continuous if it is continuous at every point in its domain (the set of its input values).
- We call a set of elements a *discrete* set if it contains a finite or countable number of elements (elements of a countable set can be enumerated).

In the real world *analog signals* are continuous functions of continuous arguments such as time, space, or any other continuous physical variables, although we often use mathematical models with not continuous analog signals such as the saw-tooth signal. We consider mainly time signals which can take on a continuum of values over a defined interval of time. For example, each value can be a real number.

Discrete signals can be discrete over a set of function values and (or) over a set of argument values. In other words, if the analog time signals are sampled, we call this set of numbers which can take on an infinity of values within a certain defined range, a *discrete-time* or *sampled* system. If the sample values are constrained to belong to a discrete set, the system becomes *digital*.

Such signals as images, speech, audio, and video are originally analog signals. In order to convert them into digital signals, we should perform the following two operations:

- First, the signal has to be sampled (the time axis must be discretized or quantized).
- The second operation is to transform the sample values (the obtained list of numbers) in such a manner that each resulting number belongs to a discrete alphabet. We call this operation *quantization*.

We start with a discussion of *sampling* which is a technique of converting an analog signal with a continuous time axis into real values in discrete-time.

Let $x(t)$ be a continuous time function. *Sampling* is taking samples of this function at time instants $t = nT_s$ for all integer values n, where T_s is called *sampling period*. The value $f_s = 1/T_s$ is called *sampling frequency*. Thus, instead of the function $x(t)$ we study the sequence of samples $x(nT_s)$, $n = 0, 1, 2, \dots$ The first question is: does sampling introduce distortion of the original continuous time function $x(t)$? The second question is: how does the distortion, if any, depend on the value of T_s? The answers are given by the so-called *sampling theorem* (Whittaker 1915; Nyquist 1928; Kotelnikov 1933; Oliver *et al.* 1948), sometimes known as the Nyquist–Shannon–Kotelnikov theorem and also referred to as the Whittaker–Shannon–Kotelnikov sampling theorem.

Before considering the sampling theorem we briefly review the notions which will be used to prove this theorem.

The Fourier transform of the analog signal $x(t)$ is given by the formula

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt$$

where $\omega = 2\pi f$ is the *radian frequency* (f is the frequency in Hz). This function is in general complex, with $X(f) = A(f)e^{j\phi(f)}$, where $A(f) = |X(f)|$ is called the *spectrum* of $x(t)$ and $\phi(f)$ is the *phase*. We can also represent $x(t)$ in terms of its Fourier transform via the *inversion formula*

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j\omega t}\, df.$$

The Fourier transform is closely related to the Laplace transform. For continuous time functions existing only for $t \geq 0$, the Laplace transform is defined as a function of the complex variable s by the following formula

$$L(s) = \int_{0}^{\infty} x(t)e^{-st}\, dt.$$

For $s = j\omega$ the Laplace transform for $x(t)$, $t \geq 0$ coincides with the Fourier transform for this function, if $L(s)$ has no poles on the imaginary axis.

If $x(t)$ is a periodic time function with period p it can be represented as the Fourier series expansion

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos(kt\omega_p) + b_k \sin(kt\omega_p))$$

where

$$\omega_p = \frac{2\pi}{p},$$
$$a_0 = \frac{2}{p}\int_{-p/2}^{p/2} x(t)\,dt,$$
$$a_k = \frac{2}{p}\int_{-p/2}^{p/2} x(t)\cos(kt\omega_p)\,dt, \quad k = 1, 2, \ldots,$$
$$b_k = \frac{2}{p}\int_{-p/2}^{p/2} x(t)\sin(kt\omega_p)\,dt, \quad k = 1, 2, \ldots$$

To replace two of the integrals ($\cos(kt\omega_p)$ and $\sin(kt\omega_p)$) by one for each index k (after simple derivations) we obtain

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k \cos(kt\omega_p + \varphi_k)$$

where φ_k is the initial phase.

Using Euler's formula $e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$ we obtain the complex form of the Fourier series expansion

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \frac{A_k}{2}\left(\exp(jkt\omega_p + j\varphi_k) + \exp(-jkt\omega_p - j\varphi_k)\right)$$
$$= \sum_{k=-\infty}^{\infty} c_k e^{jkt\omega_p}$$

where

$$c_k = \frac{1}{2}A_k e^{j\varphi_k} = \frac{1}{2}(a_k - jb_k) = \frac{1}{p}\int_{-p/2}^{p/2} x(t)\exp(-jkt\omega_p)\,dt,$$
$$c_0 = 1/2a_0\,,\ b_0 = 0,$$
$$a_k = a_{-k}\,,\ b_k = -b_{-k}.$$

It is said that the Fourier series expansion for a periodical function with period p decomposes this function into a sum of harmonical functions with frequencies $k\omega_p$, $k = 1, 2, \ldots$ The Fourier transform for a nonperiodical function represents this function as a sum of an infinite number of harmonical functions with frequencies which differ in infinitesimal quantities. Notice that a nonperiodical function of finite length T also can be decomposed into the Fourier series expansion. To do this we have to construct its periodical continuation with period T.

2.1.1 Sampling theorem

If $x(t)$ *is a signal whose Fourier transform* $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt$ *is identically zero* $X(f) = 0$ *for* $|f| > f_{\rm H}$ *, then* $x(t)$ *is completely determined by its samples taken every* $1/(2f_{\rm H})$ s. *The frequency* $f_{\rm s} = 1/T_{\rm s} = 2f_{\rm H}$ Hz *is called the Nyquist sampling rate.*

Proof. Since $X(f) = 0$ for $|f| > f_{\rm H}$ we can continue $X(f)$ periodically. Then we obtain the periodical function $\hat{X}(f)$ with period equal to $2f_{\rm H}$. The function $\hat{X}(f)$ can be decomposed into the Fourier series expansion

$$\hat{X}(f) = \sum_{k=-\infty}^{\infty} a_k e^{j2\pi fk/(2f_{\rm H})}$$

where

$$a_k = \frac{1}{2f_{\rm H}} \int_{-f_{\rm H}}^{f_{\rm H}} \hat{X}(f)e^{-j2\pi fk/(2f_{\rm H})}\,df. \tag{2.1}$$

Since $X(f)$ is the Fourier transform of $x(t)$ then $x(t)$ can be represented in terms of its Fourier transform via the inversion formula

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df = \int_{-f_{\rm H}}^{f_{\rm H}} \hat{X}(f)e^{j2\pi ft}\,df.$$

Consider the values of the time function $x(t)$ in the discrete points $t = k/(2f_{\rm H})$ for all integers k. They can be expressed as follows

$$x\left(\frac{k}{2f_{\rm H}}\right) = \int_{-f_{\rm H}}^{f_{\rm H}} \hat{X}(f)e^{j2\pi fk/(2f_{\rm H})}\,df. \tag{2.2}$$

Comparing (2.1) and (2.2), we obtain that

$$a_k = \frac{1}{2f_{\rm H}} x\left(\frac{-k}{2f_{\rm H}}\right).$$

Thus, if the time function $x(t)$ is known at points $\ldots, -2/(2f_{\rm H}), -1/(2f_{\rm H}), 0, 1/(2f_{\rm H}), 2/(2f_{\rm H}), \ldots$ then the coefficients a_k are determined. These coefficients in turn determine $\hat{X}(f)$ and thereby they determine $X(f)$. On the other hand, $X(f)$ determines $x(t)$ for all values of t. It means that there exists a unique time function which does not contain frequencies higher than $f_{\rm H}$ and passes through the given sampling points spaced $1/(2f_{\rm H})$ s.

In order to reconstruct the time function $x(t)$ using its sampling points $x(k/(2f_{\rm H}))$ we notice that

$$X(f) = \begin{cases} \sum_{k=-\infty}^{\infty} a_k e^{j2\pi fk/(2f_{\rm H})}, & \text{if } |f| \le f_{\rm H} \\ 0, & \text{if } |f| > f_{\rm H}. \end{cases}$$

To simplify our notations we introduce the *sinc-function* which is defined as

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

Using the inverse transform we obtain

$$\begin{aligned}
x(t) &= \int_{-f_H}^{f_H} X(f) e^{j2\pi f t}\, df \\
&= \int_{-f_H}^{f_H} \sum_{k=-\infty}^{\infty} a_k e^{j2\pi f (\frac{k}{2f_H}+t)}\, df \\
&= \sum_k a_k \int_{-f_H}^{f_H} e^{j2\pi f (\frac{k}{2f_H}+t)}\, df \\
&= \sum_{k=-\infty}^{\infty} a_k \Bigg(\int_{-f_H}^{f_H} \underbrace{\cos\left(2\pi f \left(\frac{k}{2f_H}+t\right)\right)}_{\text{even}} df \\
&\quad + j \int_{-f_H}^{f_H} \underbrace{\sin\left(2\pi f \left(\frac{k}{2f_H}\right)\right)}_{\text{odd}} df \Bigg) \\
&= 2 \sum_{k=-\infty}^{\infty} a_k \int_0^{f_H} \cos\left(2\pi f \left(\frac{k}{2f_H}+t\right)\right) df \\
&= 2f_H \sum_{k=-\infty}^{\infty} a_k \operatorname{sinc}\left(2f_H \left(\frac{k}{2f_H}+t\right)\right).
\end{aligned}$$

Simple derivations complete the proof:

$$\begin{aligned}
x(t) &= 2f_H \sum_{k=-\infty}^{\infty} a_k \operatorname{sinc}(2f_H t + k) \\
&= \sum_{k=-\infty}^{\infty} x\left(\frac{-k}{2f_H}\right) \operatorname{sinc}(2f_H t + k) \\
&= \sum_{i=-\infty}^{\infty} x\left(\frac{i}{2f_H}\right) \operatorname{sinc}(2f_H t - i),\ i = -k. \qquad (2.3)
\end{aligned}$$

□

In other words, the time function $x(t)$ can be represented as a sum of elementary functions in the form $\operatorname{sinc}(\alpha)$, $\alpha = 2f_H t - i$, centered in the sampling points. The sinc-function $\operatorname{sinc}(\alpha)$ is shown in Fig. 2.1. It is equal to 1 in the point $\alpha = 0$, that is, $t = i/(2f_H)$ and is equal to zero in other sampling points.

It follows from (2.3) that at time instants $t = kT_s = k/(2f_H)$ the values of $x(t)$ coincide with the sample values $x(k/(2f_H))$. For the other time instants, it is necessary to sum up an infinite number of series terms in order to reconstruct the exact value of $x(t)$. Therefore, we conclude that in order to reconstruct the function $x(t)$, it is necessary to generate an infinite train of impulses which have form $\operatorname{sinc}(\alpha)$ and are proportional to samples, and to summarize them.

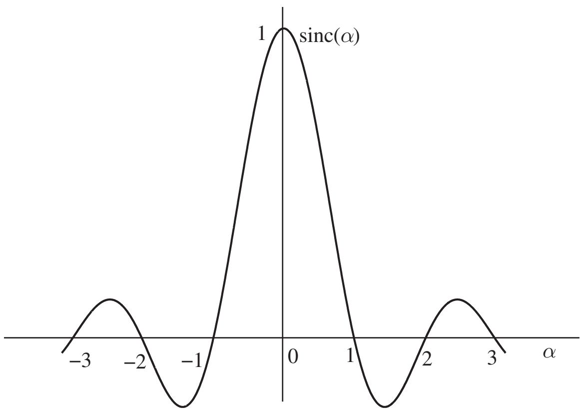

Figure 2.1 Function sinc(α)

The representation of $x(t)$ in the form (2.3) is a particular case of the so-called *orthogonal decomposition* of the function over a system of basis functions. In our case the role of basis functions is played by the sinc-functions $\mathrm{sinc}(2f_{\mathrm{H}}t - i)$ which we call *sampling functions*. They are orthogonal since

$$\int_{-\infty}^{\infty} \mathrm{sinc}(2f_{\mathrm{H}}t - j)\mathrm{sinc}(2f_{\mathrm{H}}t - i)\,dt = \begin{cases} 1/(2f_{\mathrm{H}}), & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

2.1.2 Historical background

The sampling theorem has a rather long history. It started in 1897 when the theorem was partly proved by the French mathematician Emil Borel. He showed that any continuous-time function $x(t)$ whose spectrum is limited by the maximal frequency f_{H} is uniquely defined by its samples with frequency $2f_{\mathrm{H}}$. However, he wrote nothing about how to reconstruct $x(t)$ from these samples. Then in 1915 the English mathematician Edmund Whittaker almost completed the proof by finding the so-called "cardinal function" which had the form

$$\sum_k x(kT_{\mathrm{S}})\mathrm{sinc}\left(\frac{t}{T_{\mathrm{S}}} - k\right)$$

but he never stated that this reconstructed function coincides with the original function $x(t)$. Kinnosuke Ogura actually was the first who in 1920 proved the sampling theorem (ignoring some theoretical nuances). In 1928 Harry Nyquist improved the proof by Ogura and in 1933 (independently of Nyquist) Vladimir Kotelnikov published the theorem in its contemporary form. In 1948 Claude Shannon, who was not aware of Kotelnikov's results relying only on Nyquist's proof, formulated and proved the theorem once more.

It follows from the sampling theorem that, if the Fourier transform of a time function is nonzero over a finite frequency band, then taking samples of this time function in the