Introduction to Multimedia 1

1.1 What is Multimedia?

People who use the term “multimedia” may have quite different, even opposing, viewpoints. A consumer entertainment vendor, say a phone company, may think of multimedia as interactive TV with hundreds of digital channels, or a cable-TV-like service delivered over a high-speed Internet connection. A hardware vendor might, on the other hand, like us to think of multimedia as a laptop that has good sound capability and perhaps the superiority of multimedia-enabled microprocessors that understand additional multimedia instructions.

A computer science or engineering student reading this book likely has a more application-oriented view of what multimedia consists of: applications that use multiple modalities to their advantage, including text, images, drawings, graphics, animation, video, sound (including speech), and, most likely, interactivity of some kind. This contrasts with media that use only rudimentary computer displays such as text-only or traditional forms of printed or hand-produced material.

The popular notion of “convergence” is one that inhabits the college campus as it does the culture at large. In this scenario, computers, smartphones, games, digital TV, multimedia-based search, and so on are converging in technology, presumably to arrive in the near future at a final and fully functional all-round, multimedia-enabled product. While hardware may indeed strive for such all-round devices, the present is already exciting—multimedia is part of some of the most interesting projects underway in computer science, with the keynote being *interactivity*. The convergence going on in this field is in fact a convergence of areas that have in the past been separated but are now finding much to share in this new application area. Graphics, visualization, HCI, computer vision, data compression, graph theory, networking, database systems—all have important contributions to make in multimedia at the present time.

Z.-N. Li et al., *Fundamentals of Multimedia*,
Texts in Computer Science, DOI: 10.1007/978-3-319-05290-8_1,
© Springer International Publishing Switzerland 2014

1.1.1 Components of Multimedia

The multiple modalities of text, audio, images, drawings, animation, video, and interactivity in multimedia are put to use in ways as diverse as

- Geographically based, real-time augmented-reality, massively multiplayer online video games, making use of any portable device such as smartphones, laptops, or tablets, which function as GPS-aware mobile game consoles. For example, a game in which players reinforce and link friendly "portals," and attack enemy ones that are played on GPS-enabled devices where the players must physically move to the portals (which are overlaid on real sites such as public art, interesting buildings, or parks) in order to interact with them.
- Shapeshifting TV, where viewers vote on the plot path by phone text-messages, which are parsed to direct plot changes in real-time.
- A camera that suggests what would be the best type of next shot so as to adhere to good technique guidelines for developing storyboards.
- A Web-based video editor that lets anyone create a new video by editing, annotating, and remixing professional videos on the cloud.
- Cooperative education environments that allow schoolchildren to share a single educational game using two mice at once that pass control back and forth.
- Searching (very) large video and image databases for target visual objects, using semantics of objects.
- Compositing of artificial and natural video into hybrid scenes, placing real-appearing computer graphics and video objects into scenes so as to take the physics of objects and lights (e.g., shadows) into account.
- Visual cues of video-conference participants, taking into account gaze direction and attention of participants.
- Making multimedia components *editable*—allowing the user side to decide what components, video, graphics, and so on are actually viewed and allowing the client to move components around or delete them—making components distributed.
- Building "inverse-Hollywood" applications that can recreate the process by which a video was made, allowing storyboard pruning and concise video summarization.

From a computer science student's point of view, what makes multimedia interesting is that so much of the material covered in traditional computer science areas bears on the multimedia enterprise. In today's digital world, multimedia content is recorded and played, displayed, or accessed by digital information content processing devices, ranging from smartphones, tablets, laptops, personal computers, smart TVs, and game consoles, to servers and datacenters, over such distribution media as tapes, harddrives, and disks, or more popularly nowadays, wired and wireless networks. This leads to a wide variety of research topics:

- **Multimedia processing and coding.** This includes audio/image/video processing, compression algorithms, multimedia content analysis, content-based multimedia retrieval, multimedia security, and so on.
- **Multimedia system support and networking.** People look at such topics as network protocols, Internet and wireless networks, operating systems, servers and clients, and databases.

- **Multimedia tools, end systems, and applications.** These include hypermedia systems, user interfaces, authoring systems, multimodal interaction, and integration: "ubiquity"—Web-everywhere devices, multimedia education, including computer supported collaborative learning and design, and applications of virtual environments.

Multimedia research touches almost every branch of computer science. For example, data mining is an important current research area, and a large database of multimedia data objects is a good example of just what big data we may be interested in mining; telemedicine applications, such as "telemedical patient consultative encounters," are multimedia applications that place a heavy burden on network architectures. Multimedia research is also highly interdisciplinary, involving such other research fields as electric engineering, physics, and psychology; signal processing for audio/video signals is an essential topic in electric engineering; color in image and video has a long-history and solid foundation in physics; more importantly, all multimedia data are to be perceived by human beings, which is, certainly, related to medical and psychological research.

1.2 Multimedia: Past and Present

To place multimedia in its proper context, in this section we briefly scan the history of multimedia, a relatively recent part of which is the connection between multimedia and hypermedia. We also show the rapid evolution and revolution of multimedia in the new millennium with the new generation of computing and communication platforms.

1.2.1 Early History of Multimedia

A brief history of the use of multimedia to communicate ideas might begin with newspapers, which were perhaps the *first* mass communication medium, using text, graphics, and images. Before still-image camera was invented, these graphics and images were generally hand-drawn.

Joseph Nicéphore Niépce captured the first natural image from his window in 1826 using a sliding wooden box camera [1,2]. It was made using an 8-h exposure on pewter coated with bitumen. Later, Alphonse Giroux built the first commercial camera with a double-box design. It had an outer box fitted with a landscape lens, and an inner box holding a ground glass focusing screen and image plate. Sliding the inner box makes the objects of different distances be focused. Similar cameras were used for exposing wet silver-surfaced copper plates, commercially introduced in 1839. In the 1870s, wet plates were replaced by the more convenient dry plates. Figure 1.1 (image from author's own collection) shows an example of a nineteenth century dry-plate camera, with bellows for focusing. By the end of the nineteenth

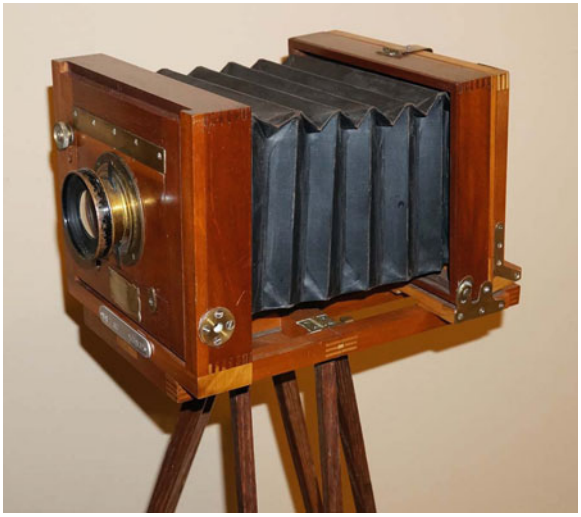

Fig. 1.1 A vintage dry-plate camera. E&H T Anthony model Champion, circa 1890

century, film-based cameras were introduced, which soon became dominant until replaced by digital cameras.

Thomas Alva Edison's phonograph, invented in 1877, was the first device that was able to record and reproduce sound. It originally recorded sound onto a tinfoil sheet phonograph cylinder [3]. Figure 1.2 shows an example of an Edison's phonograph (Edison GEM, 1905; image from author's own collection).

The phonographs were later improved by Alexander Graham Bell. Most notable improvements include wax-coated cardboard cylinders, and a cutting stylus that moved from side to side in a "zig zag" pattern across the record. Emile Berliner further transformed the phonograph cylinders to gramophone records. Each side of such a flat disk has a spiral groove running from the periphery to near the center, which can be conveniently played by a turntable with a tonearm and a stylus. These components were improved over time in the twentieth century, which eventually enabled quality sound reproducing that is very close the origin. The gramophone record was one of the dominant audio recording formats throughout much of the twentieth century. From the mid-1980s, phonograph use declined sharply because of the rise of audio tapes, and later the *Compact Disc* (CD) and other digital recording formats [4]. Figure 1.3 shows the evolution of audio storage media, starting from the Edison cylinder record, to the flat vinyl record, to magnetic tapes (reel-to-reel and cassette), and modern digital CD.

Motion pictures were originally conceived of in the 1830s to observe motion too rapid for perception by the human eye. Edison again commissioned the invention of a motion picture camera in 1887 [5]. Silent feature films appeared from 1910 to 1927; the silent era effectively ended with the release of *The Jazz Singer* in 1927.

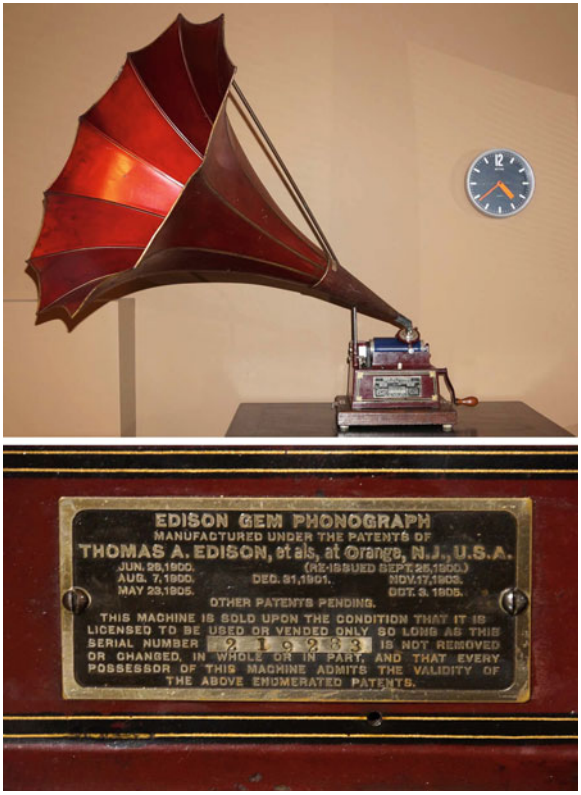
EDISON GEM PHONOGRAPH
MANUFACTURED UNDER THE PATENTS OF
THOMAS A. EDISON, et als, at Orange, N.J., U.S.A.
JUN. 26,1900.
(RE-ISSUED SEPT. 25,1900.)
AUG. 7, 1900.
DEC. 31,1901.
NOV. 17,1903.
MAY 23,1905.
OCT. 3. 1905.
OTHER PATENTS PENDING.
THIS MACHINE IS SOLD UPON THE CONDITION THAT IT IS LICENSED TO BE USED OR VENDED ONLY SO LONG AS THIS SERIAL NUMBER 219283 IS NOT REMOVED OR CHANGED, IN WHOLE OR IN PART, AND THAT EVERY POSSESSOR OF THIS MACHINE ADMITS THE VALIDITY OF THE ABOVE ENUMERATED PATENTS.

Fig. 1.2 An Edison phonograph, model GEM. Note the patent plate in the *bottom* picture, which suggests that the importance of patents had long been realized and also how serious Edison was in protecting his inventions. Despite the warnings in the plate, this particular phonograph was modified by the original owner, a good DIYer 100 years ago, to include a more powerful spring motor from an Edison Standard model and a large flower horn from the Tea Tray Company

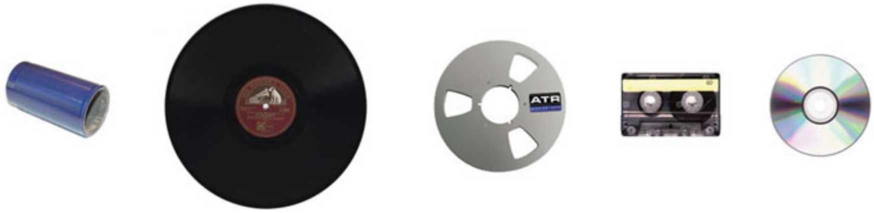

Fig. 1.3 Evolution of audio storage media. *Left* to *right* an Edison cylinder record, a flat vinyl record, a reel-to-reel magnetic tape, a cassette tape, and a CD

In 1895, Guglielmo Marconi conducted the first wireless radio transmission at Pontecchio, Italy, and a few years later (1901), he detected radio waves beamed across the Atlantic [6]. Initially invented for telegraph, radio is now a major medium for audio broadcasting. In 1909, Marconi shared the Nobel Prize for Physics.[1]

Television, or TV for short, was the new medium for the twentieth century [7]. In 1884, Paul Gottlieb Nipkow, a 23-year-old university student in Germany, patented the first electromechanical television system which employed a spinning disk with a series of holes spiraling toward the center. The holes were spaced at equal angular intervals such that, in a single rotation, the disk would allow light to pass through each hole and onto a light-sensitive selenium sensor which produced the electrical pulses. As an image was focused on the rotating disk, each hole captured a horizontal "slice" of the whole image. Nipkow's design would not be practical until advances in amplifier tube technology, in particular, the cathode ray tube (CRT), became available in 1907. Commercially available since the late 1920s, CRT-based TV established video as a commonly available medium and has since changed the world of mass communication.

All these media mentioned above are in the *analog* format, for which the time-varying feature (variable) of the signal is a continuous representation of the input, i.e., analogous to the input audio, image, or video signal. The connection between *computers* and *digital media*, i.e., media data represented using the discrete binary format, emerged actually only over a short period:

1967 Nicholas Negroponte formed the Architecture Machine Group at MIT.

1969 Nelson and van Dam at Brown University created an early hypertext editor called FRESS [8]. The present-day Intermedia project by the Institute for Research in Information and Scholarship (IRIS) at Brown is the descendant of that early system.

1976 The MIT Architecture Machine Group proposed a project entitled "Multiple Media." This resulted in the *Aspen Movie Map*, the first videodisk, in 1978.

1982 The *Compact Disc* (CD) was made commercially available by Philips and Sony, which was soon becoming the standard and popular medium for digital audio data, replacing the analog magnetic tape.

1985 Negroponte and Wiesner co-founded the MIT Media Lab, a leading research institution investigating digital video and multimedia.

1990 Kristina Hooper Woolsey headed the Apple Multimedia Lab, with a staff of 100. Education was a chief goal.

1991 MPEG-1 was approved as an international standard for digital video. Its further development led to newer standards, MPEG-2, MPEG-4, and further MPEGs, in the 1990s.

[1] Reginald A. Fessenden, of Quebec, beat Marconi to human voice transmission by several years, but not all inventors receive due credit. Nevertheless, Fessenden was paid $2.5 million in 1928 for his purloined patents.

1991 The introduction of PDAs in 1991 began a new period in the use of computers in general and multimedia in particular. This development continued in 1996 with the marketing of the first PDA with no keyboard.

1992 JPEG was accepted as the international standard for digital image compression, which remains widely used today (say, by virtually every digital camera).

1992 The first audio multicast on the multicast backbone (MBone) was made.

1995 The JAVA language was created for platform-independent application development, which was widely used for developing multimedia applications.

1996 DVD video was introduced; high-quality, full-length movies were distributed on a single disk. The DVD format promised to transform the music, gaming, and computer industries.

1998 Handheld MP3 audio players were introduced to the consumer market, initially with 32 MB of flash memory.

1.2.2 Hypermedia, WWW, and Internet

The early studies laid a solid foundation for the capturing, representation, compression, and storage of each type of media. Multimedia however is not simply about putting different media together; rather, it focuses more on the integration of them so as to enable rich interaction amongst them, and also between media and human beings.

In 1945, as part of MIT's postwar deliberations on what to do with all those scientists employed on the war effort, Vannevar Bush wrote a landmark article [9] describing what amounts to a hypermedia system, called "Memex." Memex was meant to be a universally useful and personalized memory device that even included the concept of associative links—it really is the forerunner of the World Wide Web. After World War II, 6,000 scientists who had been hard at work on the war effort suddenly found themselves with time to consider other issues, and the Memex idea was one fruit of that new freedom.

In the 1960s, Ted Nelson started the Xanadu project and coined the term *hypertext*. Xanadu was the first attempt at a hypertext system—Nelson called it a "magic place of literary memory."

We may think of a book as a *linear* medium, basically meant to be read from beginning to end. In contrast, a hypertext system is meant to be read nonlinearly, by following links that point to other parts of the document, or indeed to other documents. Figure 1.4 illustrates this familiar idea.

Douglas Engelbart, greatly influenced by Vannevar Bush's "As We May Think," demonstrated the *On-Line System* (NLS), another early hypertext program in 1968. Engelbart's group at Stanford Research Institute aimed at "augmentation, not automation," to enhance human abilities through computer technology. NLS consisted of such critical ideas as an outline editor for idea development, hypertext links, teleconferencing, word processing, and email, and made use of the mouse pointing device, windowing software, and help systems [10].

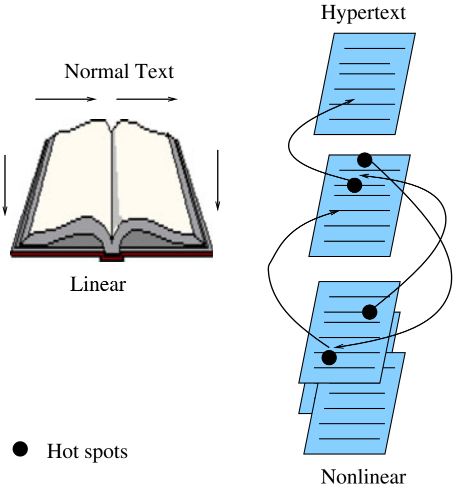

Fig. 1.4 Hypertext is nonlinear

Hypermedia, again first introduced by Ted Nelson, went beyond text-only. It includes a wide array of media, such as graphics, images, and especially the continuous media—sound and video, and links them together. The *World Wide Web* (WWW or simply Web) is the best example of a hypermedia application, which is also the largest.

Amazingly, this most predominant networked multimedia applications has its roots in nuclear physics! In 1990, Tim Berners-Lee proposed the World Wide Web to CERN (European Center for Nuclear Research) as a means for organizing and sharing their work and experimental results. With approval from CERN, he started developing a hypertext server, browser, and editor on a NeXTStep workstation. His team invented the *Hypertext Markup Language* (HTML) and the *Hypertext Transfer Protocol* (HTTP) for this purpose, too.

HyperText Markup Language (HTML)

It is recognized that documents need to have formats that are human-readable and that identify structure and elements. Charles Goldfarb, Edward Mosher, and Raymond Lorie developed the Generalized Markup Language (GML) for IBM. In 1986, the ISO released a final version of the Standard Generalized Markup Language (SGML), mostly based on the earlier GML.

HTML is a language for publishing hypermedia on the Web [11]. It is defined using SGML and derives elements that describe generic document structure and formatting. Since it uses ASCII, it is portable to all different (even binary-incompatible) computer hardware, which allows for global exchange of information. The current version of HTML is 4.01, and a newer version, HTML5, is still under development.

HTML uses tags to describe document elements. The tags are in the format `<token params>` to define the start point of a document element and `</token>` to define the end of the element. Some elements have only inline parameters and do not require ending tags. HTML divides the document into a `HEAD` and a `BODY` part as follows:

```
<HTML>
<HEAD>
...
</HEAD>
<BODY>
...
</BODY>
</HTML>
```

The `HEAD` describes document definitions, which are parsed before any document rendering is done. These include page title, resource links, and meta-information the author decides to specify. The `BODY` part describes the document structure and content. Common structure elements are paragraphs, tables, forms, links, item lists, and buttons.

A very simple HTML page is as follows:

```
<HTML>
<HEAD>
 <TITLE>
 A sample webpage.
 </TITLE>
 <META NAME = "Author" CONTENT = "Cranky Professor">
</HEAD> <BODY>
 <P>
 We can put any text we like here, since this is
 a paragraph element.
 </P>
</BODY>
</HTML>
```

Naturally, HTML has more complex structures and can be mixed with other standards. The standard has evolved to allow integration with script languages, dynamic manipulation of almost all elements and properties after display on the client side (*dynamic HTML*), and modular customization of all rendering parameters using a markup language called *Cascading Style Sheets* (CSS). Nonetheless, HTML has rigid, nondescriptive structure elements, and modularity is hard to achieve.

Extensible Markup Language (XML)

There was also a need for a markup language for the Web that has modularity of data, structure, and view. That is, we would like a user or an application to be able to *define* the tags (structure) allowed in a document and their relationship to each other, in one place, then define data using these tags in another place (the XML file), and finally, define in yet another document how to render the tags.

Suppose we wanted to have stock information retrieved from a database according to a user query. Using XML, we would use a global *Document Type Definition* (DTD) we have already defined for stock data. Your server-side script will abide by the DTD rules to generate an XML document according to the query, using data from your database. Finally, we will send users your *XML Style Sheet* (XSL), depending on the type of device they use to display the information, so that our document looks best both on a computer with a 27-in. LED display and on a small-screen cellphone.

The original XML version was XML 1.0, approved by the W3C in February 1998, and is currently in its fifth edition as of 2008. The original version is still recommended. The second version XML 1.1 was introduced in 2004 and is currently in its second edition as of 2006. XML syntax looks like HTML syntax, although it is much stricter. All tags are lowercase, and a tag that has only inline data has to terminate itself, for example, `<token params />`. XML also uses namespaces, so that multiple DTDs declaring different elements but with similar tag names can have their elements distinguished. DTDs can be imported from URIs as well. As an example of an XML document structure, here is the definition for a small XHTML document:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0"
 "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transition.dtd">
 <html xmlns="http://www.w3.org/1999/xhtml">
 ... [html that follows
   the above-mentioned
   XML rules]
 </html>
```

All XML documents start with `<?xml version="ver"?>`. `<!DOCTYPE ...>` is a special tag used for importing DTDs. Since it is a DTD definition, it does not adhere to XML rules. `xmlns` defines a unique XML namespace for the document elements. In this case, the namespace is the XHTML specifications website.

In addition to XML specifications, the following XML-related specifications are standardized:

- **XML Protocol**. Used to exchange XML information between processes. It is meant to supersede HTTP and extend it as well as to allow interprocess communications across networks.
- **XML Schema**. A more structured and powerful language for defining XML data types (tags). Unlike a DTD, XML Schema uses XML tags for type definitions.

- **XSL**. This is basically CSS for XML. On the other hand, XSL is much more complex, having three parts: *XSL Transformations* (XSLT), *XML Path Language* (XPath), and *XSL Formatting Objects*.

The WWW quickly gained popularity, due to the amount of information available from web servers, the capacity to post such information, and the ease of navigating such information with a web browser, particularly after Marc Andreessen's introduction of Mosaic browser in 1993 (later became Netscape).

Today, the Web technology is maintained and developed by the World Wide Web Consortium (W3C), together with the Internet Engineering Task Force (IETF) to standardize the technologies. The W3C has listed the following three goals for the WWW: universal access of web resources (by everyone everywhere), effectiveness of navigating available information, and responsible use of posted material.

It is worth mentioning that the Internet serves as the underlying vehicle for the WWW and the multimedia content shared over it. Starting from the Advanced Research Projects Agency Network (ARPANET) with only two nodes in 1969, the Internet gradually became the dominating global network that interconnects numerous computer networks and their billions of users with the standard Internet protocol suite (TCP/IP). It evolved together with digital multimedia. On one hand, the Internet carries much of the multimedia content. It has largely swept out optical disks as the storage and distribution media in the movie industry. It is currently reshaping the TV broadcast industry with an ever-accelerating speed. On the other hand, the Internet was not initially designed for multimedia data and was not quite friendly to multimedia traffic. Multimedia data, now occupying almost 90 % of the Internet bandwidth, is the key driving force toward enhancing the existing Internet and toward developing the next generation of the Internet, as we will see in Chaps. 15 and 16.

1.2.3 Multimedia in the New Millennium

Entering the new millennium, we have witnessed the fast evolution toward a new generation of social, mobile, and cloud computing for multimedia processing and sharing. Today, the role of the Internet itself has evolved from the original use as a communication tool to provide easier and faster sharing of an infinite supply of information, and the multimedia content itself has also been greatly enriched. High-definition videos and even 3D/multiview videos can be readily captured and browsed by personal computing devices, and conveniently stored and processed with remote cloud resources. More importantly, the users are now actively engaged to be part of a social ecosystem, rather than passively receiving media content. The revolution is being driven further by the deep penetration of 3G/4G wireless networks and smart mobile devices. Coming with highly intuitive interfaces and exceptionally richer multimedia functionalities, they have been seamlessly integrated with online social networking for instant media content generation and sharing.

Below, we list some important milestones in the development of multimedia in the new millennium. We believe that most of the readers of this textbook are familiar with them, as we are all in this Internet age, witnessing its dramatic changes; many readers, particularly the younger generation, would be even more familiar with the use of such multimedia services as YouTube, Facebook, and Twitter than the authors.

2000 WWW size was estimated at over one billion pages. Sony unveiled the first Blu-ray Disc prototypes in October 2000, and the first prototype player was released in April 2003 in Japan.

2001 The first peer-to-peer file sharing (mostly MP3 music) system, Napster, was shut down by court order, but many new peer-to-peer file sharing systems, e.g., Gnutella, eMule, and BitTorrent, were launched in the following years. Coolstreaming was the first large-scale peer-to-peer streaming system that was deployed in the Internet, attracting over one million in 2004. Later years saw the booming of many commercial peer-to-peer TV systems, e.g., PPLive, PPStream, and UUSee, particularly in East Asia. NTT DoCoMo in Japan launched the first commercial 3G wireless network on October 1. 3G then started to be deployed worldwide, promising broadband wireless mobile data transfer for multimedia data.

2003 Skype was released for free peer-to-peer voice over the Internet.

2004 Web 2.0 was recognized as a new way to utilize software developers and end-users use the Web (and is not a technical specification for a new Web). The idea is to promote user collaboration and interaction so as to generate content in a "virtual community," as opposed to simply passively viewing content. Examples include social networking, blogs, wikis, etc. Facebook, the most popular online social network, was founded by Mark Zuckerberg. Flickr, a popular photo hosting and sharing site, was created by Ludicorp, a Vancouver-based company founded by Stewart Butterfield and Caterina Fake.

2005 YouTube was created, providing an easy portal for video sharing, which was purchased by Google in late 2006. Google launched the online map service, with satellite imaging, real-time traffic, and Streetview being added later.

2006 Twitter was created, and rapidly gained worldwide popularity, with 500 million registered users in 2012, who posted 340 million tweets per day. In 2012, Twitter offered the Vine mobile app, which enables its users to create and post short video clips of up to 6 s. Amazon launched its cloud computing platform, Amazon's Web Services (AWS). The most central and well-known of these services are Amazon EC2 and Amazon S3. Nintendo introduced the Wii home video game console, whose remote controller can detect movement in three dimensions.

2007 Apple launched the first generation of iPhone, running the iOS mobile operating system. Its touch screen enabled very intuitive operations, and the associated App Store offered numerous mobile applications. Goolge unveiled Android mobile operating system, along with the founding of the Open Handset Alliance: a consortium of hardware, software, and telecommunication companies devoted to advancing open standards for mobile devices. The first Android-powered phone was sold in October 2008, and Google Play, Android's primary app store, was soon launched. In the following years, tablet computers using iOS, Android, and Windows with larger touch screens joined the eco-system, too.

2009 The first LTE (Long Term Evolution) network was set up in Oslo, Norway, and Stockholm, Sweden, making an important step toward 4G wireless networking. James Cameron's film, Avatar, created a surge on the interest in 3D video.

2010 Netflix, which used to be a DVD rental service provider, migrated its infrastructure to the Amazon AWS cloud computing platform, and became a major online streaming video provider. Master copies of digital films from movie studios are stored on Amazon S3, and each film is encoded into over 50 different versions based on video resolution, audio quality using machines on the cloud. In total, Netflix has over 1 petabyte of data stored on Amazon's cloud. Microsoft introduced Kinect, a horizontal bar with full-body 3D motion capture, facial recognition, and voice recognition capabilities, for its game console Xbox 360.

2012 HTML5 subsumes the previous version, HTML4, which was standardized in 1997. HTML5 is a W3C "Candidate Recommendation." It is meant to provide support for the latest multimedia formats while maintaining consistency for current web browsers and devices, along with the ability to run on low-powered devices such as smartphones and tablets.

2013 Sony released its PlayStation 4, a video game console that is to be integrated with Gaikai, a cloud-based gaming service that offers streaming video game content. 4K resolution TV started to be available in the consumer market.

1.3 Multimedia Software Tools: A Quick Scan

For a concrete appreciation of the current state of multimedia software tools available for carrying out tasks in multimedia, we now include a quick overview of software categories and products.

These tools are really only the beginning—a fully functional multimedia project can also call for stand-alone programming as well as just the use of predefined tools to fully exercise the capabilities of machines and the Internet.[2]

In courses we teach using this text, students are encouraged to try these tools, producing full-blown and creative multimedia productions. Yet this textbook is not a "how-to" book about using these tools—it is about understanding the fundamental design principles behind these tools! With a clear understanding of the key multimedia data structures, algorithms, and protocols, a student can make smarter and

[2] See the accompanying website for several interesting uses of software tools. In a typical computer science course in multimedia, the tools described here might be used to create a small multimedia production as a first assignment. Some of the tools are powerful enough that they might also form part of a course project.

advanced use of such tools, so as to fully unleash their potentials, and even improve the tools themselves or develop new tools.

The categories of software tools we examine here are

- Music sequencing and notation
- Digital audio
- Graphics and image editing
- Video editing
- Animation
- Multimedia authoring.

1.3.1 Music Sequencing and Notation

Cakewalk Pro Audio

Cakewalk Pro Audio is a very straightforward music-notation program for "sequencing." The term *sequencer* comes from older devices that stored sequences of notes in the MIDI music language (*events*, in MIDI; see Sect. 6.2).

Finale, Sibelius

Finale and Sibelius are two composer-level notation systems; these programs likely set the bar for excellence, but their learning curve is fairly steep.

1.3.2 Digital Audio

Digital Audio tools deal with accessing and editing the actual sampled sounds that make up audio.

Adobe Audition

Adobe Audition (formerly Cool Edit) is a powerful, popular digital audio toolkit with capabilities (for PC users, at least) that emulate a professional audio studio, including multitrack productions and sound file editing, along with digital signal processing effects.

Sound Forge

Like Audition, Sound Forge is a sophisticated PC-based program for editing WAV files. Sound can be captured through the sound card, and then mixed and edited. It also permits adding complex special effects.

Pro Tools

Pro Tools is a high-end integrated audio production and editing environment that runs on Macintosh computers as well as Windows. Pro Tools offers easy MIDI creation and manipulation as well as powerful audio mixing, recording, and editing software. Full effects depend on purchasing a dongle.

1.3.3 Graphics and Image Editing

Adobe Illustrator

Illustrator is a powerful publishing tool for creating and editing vector graphics, which can easily be exported to use on the Web.

Adobe Photoshop

Photoshop is the standard in a tool for graphics, image processing, and image manipulation. Layers of images, graphics, and text can be separately manipulated for maximum flexibility, and its set of filters permits creation of sophisticated lighting effects.

Adobe Fireworks

Fireworks is software for making graphics specifically for the Web. It includes a bitmap editor, a vector graphics editor, and a JavaScript generator for buttons and rollovers.

Adobe Freehand

Freehand is a text and web graphics editing tool that supports many bitmap formats, such as GIF, PNG, and JPEG. These are *pixel-based* formats, in that each pixel is specified. It also supports *vector-based* formats, in which endpoints of lines are specified instead of the pixels themselves, such as SWF (Adobe Flash). It can also read Photoshop format.

1.3.4 Video Editing

Adobe Premiere

Premiere is a simple, intuitive video editing tool for *nonlinear* editing—putting video clips into any order. Video and audio are arranged in *tracks*, like a musical score.

It provides a large number of video and audio tracks, superimpositions, and virtual clips. A large library of built-in transitions, filters, and motions for clips allows easy creation of effective multimedia productions.

CyberLink PowerDirector

PowerDirector produced by CyberLink Corp. is by far the most popular nonlinear video editing software. It provides a rich selection of audio and video features and special effects and is easy to use. It supports all modern video formats including AVCHD 2.0, 4K Ultra HD, and 3D video. It supports 64-bit video processing, graphics card acceleration, and multiple CPUs. Its processing and preview are much faster than Premiere. However, it is not as “programmable” as Premiere.

Adobe After Effects

After Effects is a powerful video editing tool that enables users to add and change existing movies with effects such as lighting, shadows, and motion blurring. It also allows layers, as in Photoshop, to permit manipulating objects independently.

Final Cut Pro

Final Cut Pro is a video editing tool offered by Apple for the Macintosh platform. It allows the input of video and audio from numerous sources, and provides a complete environment, from editing and color correction to the final output of a video file.

1.3.5 Animation

Multimedia APIs

Java3D is an API used by Java to construct and render 3D graphics, similar to the way Java Media Framework handles media files. It provides a basic set of object primitives (cube, splines, etc.) upon which the developer can build scenes. It is an abstraction layer built on top of OpenGL or DirectX (the user can select which), so the graphics are accelerated.

DirectX, a Windows API that supports video, images, audio, and 3D animation, is a common API used to develop multimedia Windows applications such as computer games.

OpenGL was created in 1992 and is still a popular 3D API today. OpenGL is highly portable and will run on all popular modern operating systems, such as UNIX, Linux, Windows, and Macintosh.

Animation Software

Autodesk 3ds Max (formerly 3D Studio Max) includes a number of high-end professional tools for character animation, game development, and visual effects production. Models produced using this tool can be seen in several consumer games, such as for the Sony Playstation.

Autodesk Softimage (previously called Softimage XSI) is a powerful modeling, animation, and rendering package for animation and special effects in films and games.

Autodesk Maya, a competing product to Softimage, is a complete modeling package. It features a wide variety of modeling and animation tools, such as to create realistic clothes and fur. Autodesk Maya runs on Windows, Mac OS, and Linux.

GIF Animation Packages

For a much simpler approach to animation that also allows quick development of effective small animations for the Web, many shareware and other programs permit creating animated GIF images. GIFs can contain several images, and looping through them creates a simple animation.

Linux also provides some simple animation tools, such as `animate`.

1.3.6 Multimedia Authoring

Tools that provide the capability for creating a complete multimedia presentation, including interactive user control, are called *authoring* programs.

Adobe Flash

Flash allows users to create interactive movies by using the score metaphor—a timeline arranged in parallel event sequences, much like a musical score consisting of musical notes. Elements in the movie are called *symbols* in Flash. Symbols are added to a central repository, called a library, and can be added to the movie's timeline. Once the symbols are present at a specific time, they appear on the Stage, which represents what the movie looks like at a certain time, and can be manipulated and moved by the tools built into Flash. Finished Flash movies are commonly used to show movies or games on the Web.

Adobe Director

Director uses a movie metaphor to create interactive presentations. This powerful program includes a built-in scripting language, Lingo, that allows creation of complex interactive movies.[3] The "cast" of characters in Director includes bitmapped sprites, scripts, music, sounds, and palettes. Director can read many bitmapped file formats. The program itself allows a good deal of interactivity, and Lingo, with its own debugger, allows more control, including control over external devices.

Dreamweaver

Dreamweaver is a webpage authoring tool that allows users to produce multimedia presentations without learning any HTML.

1.4 Multimedia in the Future

This textbook emphasizes on the *fundamentals* of multimedia, focusing on the basic and mature techniques that collectively form the foundation of today's multimedia systems. It is however worth noting that multimedia research remains young and is vigorously growing. It brings many exciting topics together, and we will certainly see great innovations that will dramatically change our life in the near future [12].

For example, researchers are interested in camera-based object tracking technology. But while face detection is ubiquitous, with camera software doing a reasonable job of identifying faces in images and video, face detection and object tracking are by no means solved problems today (although for face tracking, combining multiple poses may be a promising direction [13]). As a matter of fact, interest in these topics is somewhat flagging, with need for some new breakthrough. Instead, the current emphasis is on event detection, e.g. for security applications such as a person leaving a bag unattended in an airport.

While shot detection—finding where scene changes exist in video—and video classification have for some time been of interest, new challenges have now arisen in these old subjects due to the abundance of online video that is not professionally edited.

Extending the conventional 2D video, today's 3D capture technology is fast enough to allow acquiring dynamic characteristics of human facial expression during speech, to synthesize highly realistic facial animation from speech for low-bandwidth applications. Beyond this, multiple views from several cameras or from a single camera under differing lighting can accurately acquire data that gives both the shape and surface properties of materials, thus automatically generating synthetic graphics models. This allows photo-realistic (video-quality) synthesis of virtual actors. Multimedia applications aimed at handicapped persons, particularly those with poor

[3] Therefore, Director is often a viable choice with students for creating a final project in multimedia courses—it provides the desired power without the inevitable pain of using a full-blown C++ program. The competing technology is likely Actionscripts in Flash.

vision and the elderly, are a rich field of endeavor in current research, too. Another related example is *Google Glass*, which, equipped with an optical head-mounted display, enables interactive, smartphone-like information display for its users. Wirelessly connected the Internet, it can also communicate using natural language voice commands. All these make a good step toward *wearable computing* of great potentials.

Online social media, such as YouTube, Facebook, and Twitter, appeared only in the past decade, but are rapidly changing the way for information generation and sharing and even our daily life. Research on social media is likely one of the most important areas under scrutiny, with some 100,000 academic articles produced per year in this area. It leads to a series of interesting new research topics:

Crowdsourcing for multimedia This concept, that the input of a large number of human contributors is made use of in multimedia projects, has experienced a large growth in attention. For example, having people provide tags to aid in understanding the visual content of images and video, such as Amazon's "Mechanical Turk," to outsource such time-consuming tasks as semantic video annotation to a large number of workers who are willing to work for small reward or just for fun. A straightforward use of such large populations is to analyze "sentiment," such as the popularity of a particular brand-name as evidenced by reading several thousand tweets on the subject. Another example is "Digital fashion," which aims to develop smart clothing that can communicate with other such enhanced clothing using wireless communication, so as to artificially enhance human interaction in a social setting. The vision here is to use technology to allow individuals to allow certain thoughts and feelings to be broadcast automatically, for exchange with others equipped with similar technology.

Executable academic papers In science and engineering, one traditional way to communicate findings is by publication of papers in academic journals. A new idea that exploits the completely digital pathway for broadcast of information is called "Executable papers." The idea here is that results discussed in a published paper are often difficult to reproduce. The reason is that datasets being used and programming code working on that data are typically not supplied as part of the publication. The executable papers approach allows the "reader" to interact with and interactively manipulate the data and code, to further understand the findings being presented. Moreover, the concept includes allowing the reader to rerun the code, change parameters, or upload different data.

Animated Lifelike Virtual Agents e.g. virtual educators, in particular as social partners for special needs children; and various other roles that are designed to demonstrate emotion and personality and with a variety of embodiments. The objective is flexibility as opposed to a fixed script.

Behavioral science models can be brought into play to model interaction between people, which can then be extended to enable natural interaction by virtual characters. Such "augmented interaction" applications can be used to develop interfaces between real and virtual humans for tasks such as augmented storytelling.

Each of these application areas pushes the development of computer science generally, stimulates new applications, and fascinates practitioners. The chief leaders of multimedia research have generated several overarching "grand challenge" problems, which act as a type of state-of-the-art for multimedia interests. At present some of these consist of the following:

- Social Event Detection for Social Multimedia: discovering social events planned and attended by people, as indicated by collections of multimedia content that was captured by people and uploaded to social-media sites.
- Search and Hyperlinking of Television Content: finding relevant video segments for a particular subject and generating useful hyperlinks for each of these segments. The underlying idea is that instead of people performing a search and following hyperlinks, this could all be automated intelligently.
- Geo-coordinate Prediction for Social Multimedia: estimating the GPS coordinates of images and videos, using all the data available including tags, audio, and users.
- Violent Scenes Detection in Film: automatically detecting portions of movies depicting violence. Again, all aspects available such as text and audio could be brought into play.
- Preserving Privacy in Surveillance Videos: methods obscuring private information (such as faces on Google Earth), so as to render privacy-sensitive elements of video unrecognizable, while at the same time allowing the video to still be viewable by people and also allow computer vision tasks such as object tracking.
- Spoken Term Web Search: searching for audio content within audio content by using an audio query.
- Question Answering for the Spoken Web: a variant on the above, specifically for matching spoken questions with a collection of spoken answers.
- Soundtrack Selection for Commercials: choosing the most suitable music soundtrack from a list of candidates. The objective here is to use extra features ("meta-data") such as text, descriptive features calculated for audio and for video, webpages, and social tags to help in the task.

Solutions to these challenges can be difficult, but the impact can be enormous, not only to the IT industry, but also to everyone, as we all live in a digital multimedia world. We want this textbook to bring valuable knowledge about multimedia to you, and hope you enjoy it and perhaps even contribute to this promising field (maybe for some of the topics listed above, or beyond) in your future career!

1.5 Exercises

1. Using your own words, describe what is "multimedia"? Is multimedia simply a collection of different types of media?
2. Identify three novel multimedia applications. Discuss why you think these are novel and their potential impact.
3. Discuss the relation between multimedia and hypermedia.
4. Briefly explain, in your own words, "Memex" and its role regarding hypertext. Could we carry out the Memex task today? How do you use Memex ideas in your own work?

5. Discover a current media input, storage, or playback device that is analog. Is it necessary to convert to digital? What are the pros and cons to be analog or digital?
6. Your task is to think about the transmission of smell over the Internet. Suppose we have a smell sensor at one location and wish to transmit the *Aroma Vector* (say) to a receiver to reproduce the same sensation. You are asked to design such a system. List three key issues to consider and two applications of such a delivery system. *Hint*: Think about medical applications.
7. Tracking objects or people can be done by both sight and sound. While vision systems are precise, they are relatively expensive; on the other hand, a pair of microphones can detect a person's *bearing* inaccurately but cheaply. Sensor *fusion* of sound and vision is thus useful. Surf the Web to find out who is developing tools for video conferencing using this kind of multimedia idea.
8. *Non-photorealistic* graphics means computer graphics that do well enough without attempting to make images that look like camera images. An example is conferencing. For example, if we track lip movements, we can generate the right animation to fit our face. If we do not much like our own face, we can substitute another one—facial-feature modeling can map correct lip movements onto another model. See if you can find out who is carrying out research on generating avatars to represent conference participants' bodies.
9. Watermarking is a means of embedding a hidden message in data. This could have important legal implications: Is this image copied? Is this image doctored? Who took it? Where? Think of "messages" that could be sensed while capturing an image and secretly embedded in the image, so as to answer these questions. (A similar question derives from the use of cell phones. What could we use to determine who is putting this phone to use, and where, and when? This could eliminate the need for passwords or others using the phone you lost.)

References

1. B. Newhall, *The History of Photography: From 1839 to the Present*, The Museum of Modern Art (1982)
2. T. Gustavson, G. Eastman House, *Camera: A History of Photography from Daguerreotype to Digital* (Sterling Signature, New York, 2012)
3. A. Koenigsberg, *The Patent History of the Phonograph*, (APM Press, Englewood, 1991), pp. 1877–1912
4. L.M. David Jr., *Sound Recording: The Life Story of a Technology*, (Johns Hopkins University Press, Baltimore, 2006)
5. Q.D. Bowers, K. Fuller-Seeley. *One Thousand Nights at the Movies: An Illustrated History of Motion Pictures*, (Whitman Publishing, Atlanta, 2012), pp. 1895–1915
6. T.K. Sarkar, R. Mailloux, A.O. Arthur, M. Salazar-Palma, D.L. Sengupta, *History of Wireless*, (Wiley-IEEE Press, Hoboken, 2006)
7. M. Hilmes, J. Jacobs, *The Television History Book (Television, Media and Cultural Studies)*, (British Film Institute, London, 2008)
8. N. Yankelovitch, N. Meyrowitz, A. van Dam, Reading and writing the electronic book, in *Hypermedia and Literary Studies*, ed. by P. Delany, G.P. Landow (MIT Press, Cambridge, 1991)
9. V. Bush, in *As We May Think*, (The Atlantic Monthly, Boston, 1945)

10. D. Engelbart, H. Lehtman, *Working Together*, (BYTE Magazine, Penticton, 1988), pp. 245–252
11. J. Duckett, *HTML and CSS: Design and Build Websites*, (Wiley, Hoboken, 2011)
12. K. Nahrstedt, R. Lienhart, M. Slaney, Special issue on the 20th anniversary of ACM SIGMM. ACM Trans. Multimedia Comput. Commun. Appl. (TOMCCAP), (2013)
13. A.D. Bagdanov, A.D. Bimbo, F. Dini, G. Lisanti, I. Masi, Posterity logging of face imagery for video surveillance. IEEE Multimedia **19**(4), 48–59 (2012)