Preface

Throughout history, pictures have played a major role in communicating information, emotion, and entertainment. Images are now everywhere and the way we exchange and interpret them is central to everyday life.

A very large proportion, approximately 50%, of the human brain's volume is dedicated to vision and related tasks. Comparing this with around 20% for motor functions and only 10% for problem solving, helps to explain why we can build computers that beat us at chess but cannot yet design machines that come close to replicating the human vision system. It also reflects the importance of vision for function and survival, and explains its capacity to entertain, challenge, and inform.

The eye is a relatively simple organ which captures two inverted and distorted impressions of the outside world, created by light reflected from external objects. It is interesting to note that optical images were unknown before the 10th century and it was not until the 17th century that it was realized that "images" existed in the eyes—via light projected onto our retinas.

So why can such a relatively simple receptor be responsible for creating such a rich interface to the outside world? The brain processes the crude images on the retina to detect, assimilate, and predict threats, to produce perceptions of beauty, and to perceive other means for survival and reproduction. These perceptions are not just derived from instantaneous images, but by the temporal and spatial relationships which they represent, placed in the context of a wealth of knowledge obtained from prior experiences. This combination of sensing, stored data, and dynamic neural processing uniquely enables animals, in varying degrees, to perform recognition, prediction, and planning tasks.

Although many contemporary applications use computer vision to aid image understanding (for example in manufacturing, medical diagnosis, robotics, and surveillance), in most cases, the final consumer and interpreter is a human being. In many ways, this captures the essence of communicating pictures—the convergence of information engineering and mathematics with psychology, biology, and the creative arts.

Early visual communications

Visual communication has a very long history, from early cave paintings to Egyptian hieroglyphs, through the inventions of paper, moveable type, photography, and film, to modern art, television, and "new digital media." To give some feel for its beginnings, Figure 1 shows a simple piece of ochre which was discovered in a cave in South Africa. While there is no evidence of drawings in the cave, the markings on the ochre strongly indicate that it was used for drawing—some 200,000 years ago. This is the earliest evidence of humans using tools for visual communication.

FIGURE 1

Drawing Ochre: c. 200,000 BC, South Africa. (Courtesy of Professor Alice Roberts.)

Exploration and exploitation of different media has grown ever more sophisticated over the ensuing 200,000 years. The first evidence for humans engraving patterns dates from the mid stone age (South Africa c. 60,000 BC) and exists in the form of hatched bands on ostrich eggshells. Another equally famous example of prehistoric art, known as Venus of Willendorf, is a stone-carved sculpture representing the female form and dates from the Paleolithic Period (c. 24,000 BC). Later examples of paintings on cave walls, presumed to be a means of recording significant events (but surely also admired for their beauty), can be traced back as far as c. 15,000 BC. Particularly fine examples can be found in Lascaux, France.

Mobility and print

Mobility in image communications started with the need to exchange information. Early Mesopotamian writing systems (c. 4000 BC) used a stylus and clay for recording numbers whereas Chinese tortoise-shell carvings date back even further to c. 6000 BC. Papyrus was first manufactured in Egypt as far back as the third millennium BC and this was rivaled by parchment from c. 600 BC, which was prepared from animal skins. Paper, made from wood-pulp, is attributed to the Chinese c. 200 BC, and provided a cheaper and more flexible alternative to silk. The availability of this new medium and its impact on cheap mass communication was unimaginable at the time.

Printing (leading to the mass reproduction of text and images, typically with ink on paper) was clearly one of the most significant developments in image and textual communications, supporting widespread and consistent dissemination of art and information. Over time, print provided knowledge to the masses and enabled an increased accessibility to what was previously only in the domain of the privileged.

The origins of printing can be found in the cylinder seals used in Mesopotamia around 3500 BC. The invention of woodblock printing is attributed to the Chinese c. 200 AD, as is the introduction of moveable type in 1040.

Perhaps the most significant step forward was taken by Johannes Gutenberg in 1454, with the development of the moveable-type printing press. This high quality process led to the production of the Gutenberg Bible (1455) which rapidly spread across Europe, triggering the Renaissance. Lithography—the use of a flat stone or metal plate to transfer ink to paper—was introduced in 1796 by Alois Senefelder, followed by chromolithography in 1837. It was not until 1907 that screen printing was invented, with phototypesetting in 1960. The printing process has developed dramatically over the past 50 years with the introduction of the laser printer in 1969 and the digital press in 1993.

Artistic impression, painting, and perspective

Byzantine and Gothic art from the Early Modern Era and Middle Ages primarily reflects the dominance of the church through the expression of religious stories and icons—depiction of the material world or everyday life was deemed unnecessary at the time. There was extensive use of gold in paintings for religious emphasis and this has been shown through contemporary studies to influence the way eyes are drawn toward or away from key areas (especially when viewed under the prevailing candlelit conditions).

Figures were normally idealized and drawn without perspective. This, however, all changed with the Renaissance (14th–17th centuries) which began in Firenze and then spread to the rest of Europe. During this period, art returned to the material world, with painters such as Holbein depicting real people with emotions, and Filippo Brunelleschi and Leon Battista Alberti formalizing linear perspective to create depth. Leonardo Da Vinci (scientist, engineer, musician, anatomist, cartographer, and painter) was perhaps the most famous contributor to Renaissance art. His painting, Mona Lisa (completed 1519), is worthy of special mention—a half-length portrait of a woman whose facial expression possessed an enigmatic quality that still fascinates today.

The separation of the Church of England from Rome was used, first by Henry VIII in his Great Bible (1539—based on Tyndale's earlier incomplete version) and then by James I in the King James Bible of 1611, to bring an English version of the bible to the population, reinforcing the message of an independent church. Two thousand and five hundred copies of the Great Bible were produced and the King James Bible is widely acknowledged as the best selling book in history.

In the 18th century (the so-called Enlightenment Period), inspiration in art was derived more from our expanding knowledge of the universe and an associated deterministic interpretation of the world. This was then progressively transformed and challenged in the 19th and 20th centuries leading to more abstract forms of art such as Impressionism, Expressionism, Cubism, and Surrealism, all influenced by experiments with perception, the distortion of reality and uncertainty.

Photographic film and motion pictures

The process of creating art on canvas was transformed in 1840 when Louis Daguerre (France) and William Fox Talbot (UK) independently invented the photographic process. Like the printing press 400 years before, this revolutionized the way we see and communicate pictures. At once, it opened up new opportunities for artistic creation as well as, for the first time, enabling communication of the "true" representation of people and events—"the photograph never lies."

A number of other key inventions emerged in the late 19th century that facilitated the creation and public display of stored moving images. Firstly, US inventor Thomas Edison is credited with inventing the electric light bulb (although this was the culmination of many years of work by him and others) in New Jersey in 1879. It was probably Edison who was the first to realize that light was the key to display moving images. Around the same time, George Eastman in Rochester, USA, founded the Eastman Kodak Company and invented roll film, bringing photography into mainstream use. This was used in 1888 by the world's first filmmaker Louis Le Prince and followed in 1895 by the first public motion picture presentation created by the Lumiere brothers in Lyon, France. Louis Lumiere was the first to realize that the grab-advance mechanism of a sewing machine could be used to advance, pause, and expose Eastman's film, illuminating it with one of Edison's light bulbs and projecting the light via a lens (i.e. a back to front camera) onto a screen in front of an audience.

Leon Bouly and Thomas Edison himself were also among the world's first film makers. One of the earliest films by the Lumiere brothers (1896) was called *Arrival of a Train at Ciotat*. This is said to have unnerved the audience due to the realism of the train coming toward them "through" the screen! An impact perhaps more profound or "immersive" at the time than some of today's 3-D movies! We think of widescreen formats as a recent innovation, in cinema since 1953, but only recently in TV. However, the first use of widescreen was in 1897 when Enoch Rector replaced conventional 35 mm film with a 63 mm alternative to form a broader image of a boxing match, in order to show more of the action.

Movies in the early 20th century moved from using technology for capturing reality to story telling and innovations such as special effects, which both mirrored and distorted reality. The development of cinematography followed, exploring close-ups, camera angles, and narrative flow. Sound arrived in the 1920s and color movies in the 1930s.

As Howard Hughes found when making the film *Hells Angels* (1930), communicating pictures can become very expensive. With production costs of about $3 million, the film (which told the story of WWI fighter pilots) and its director were ridiculed by other Hollywood producers, such as Louis Mayer, for its lavish use of real aircraft, props, and cameras. He was vindicated, however, by taking nearly $8 million at the box office, a record at the time.

By way of comparison, Martin Scorsese's film *The Aviator* (2004), about the life of Hughes, was made with a budget of some $116 million. The film is of particular interest as, for the first 50 min, scenes are rendered only in shades of red and cyan-blue—green objects are rendered as blue. This emulated the early bipack color movies, in particular the Multicolor process, which Hughes himself owned. Scenes in the film depicting events after 1935 emulate the saturated appearance of three-strip Technicolor. It is also important to note that, unlike *Hells Angels*, almost all the flying scenes in *The Aviator* were created using scale models and computer graphics. The film won five academy awards.

As a young man, Hughes set himself three major ambitions: to be the world's best pilot, to be the world's best filmmaker, and to be the richest man in the world. While many might argue with his success in the first two, there is certainly no doubt that he achieved the last one!

Television

Television was perhaps the technology that brought the most dramatic change to the way we acquired information and how we used our leisure time in the 20th century. Television has its roots in the discovery of selenium's photoconductivity (Willoughby Smith, 1873). Early mechanical versions started with German student, Paul Nipkow, who invented the scanning disk–image rasterizer in 1884 (although he never made a working model). This scanning method solved the key problem of how to map a 2-D image onto a 1-D radio signal. Constantin Perskyi is accredited with being the first to use the word "television" in his presentation to the International Electricity Congress in Paris in 1900.

After contributions by DeForest, Korn, and others in the US and Russia, a crude still image was transmitted by Rignoux and Fournier in Paris in 1909. Their scanner employed a rotating mirror with a receiver based on 64 selenium cells. This technology was developed through the early 20th century, culminating in the first public demonstration of TV in 1925 by John Logie Baird in London. AT&T's Bell Telephone Laboratories transmitted halftone still images later the same year. None of these, however, really achieved psychovisually acceptable moving images since they were only transmitted at around five frames per second.

The first demonstration of motion television is attributed to Baird in January 1926 in London, when he presented his 30-line electromechanical system to members of the Royal Institution. Baird's work continued apace, with transmission over fixed and wireless channels. The first transatlantic transmission, between London and New York, was achieved in 1928. It is interesting to note that Baird also worked on low light imaging and 3-D TV at that time!

In the USA, Philo Farnsworth transmitted the first fully electronic television picture in 1927—an image of a straight line—later that year followed by his *Puff of Smoke* sequence. The first experimental TV service was established in Germany in 1929 and later that year a similar operation was established in France. Although Baird's system provided only 240 scanning lines, it did form the basis of the world's first public service, launched by the BBC from Alexandra Palace in 1936. Baird's method was, however, rapidly replaced by Marconi-EMI's "high definition" 405-line all-electronic system.

625-line monochrome broadcasts were introduced in 1964 and the BBC launched its BBC2 service on UHF using a 625-line system with a PAL color system in 1967. 405-line transmissions continued until 1985 when the frequencies were reused for Digital Audio Broadcasting (DAB) and Private Mobile Radio (PMR) services. Color TV was actually introduced much earlier in the US (1953) but was slow to take off due to a lack of content and high prices. The US networks converted to color in 1965, coincident with the introduction of GE's Porta-Color TV set, and it gained popularity rapidly. From the earliest days, operators realized that television broadcasting was expensive in terms of radio bandwidth, especially as frame rates and resolution (or the numbers of scanning lines) increased. This was addressed in a number of ingenious ways—through the use of interlaced fields and through the use of color sub-carriers.

The digital video revolution really took off when television (both satellite and terrestrial) went digital (in the UK) in 1998. This was only made possible by extensive worldwide effort in digital video broadcasting, in particular by the Motion Picture Experts Group (MPEG), chaired by Leonardo Chiariglione. This led to the ubiquitous MPEG-2 Audiovisual Coding framework which still forms the basis for most digital TV in the world. Analog transmission has now been discontinued in much of the world (change-over was completed in the UK in 2012).

The first public high definition television (HDTV) broadcast in the United States occurred in July 1996. Although HDTV broadcasts had been demonstrated in Europe since the early 1990s, the first regular broadcasts were started in 2004 by the Belgian company, Euro1080.

Pervasive media and the internet

Alongside the development of visual media through advances in film and television, the invention of advanced telescopes, microscopes, and computer-based imaging technology allows us to see what used to be too far away, too small, or simply invisible. Now, we can "see" almost everything. But… the challenge is to store and transmit the captured information in an efficient manner without compromising quality. We therefore need to exploit knowledge of the human visual system, to identify redundancy in the pictures and, hence, to compress them efficiently.

Video is now one of the main drivers for advances in communication technology and video content accounts for the majority of internet traffic. Real Networks were one of the pioneers of streaming media, introducing RealPlayer in 1997. Streaming was first incorporated into Windows media Player 6.4 in 1999 and Apple introduced streaming in Quicktime 4 in 1999. The demand for a common streaming format was settled through widespread adoption of Adobe Flash which is used by sites such as YouTube.

Internet video has been enabled by advances in video compression, most notably H.264/AVC, which was developed by the ITU-T Video Coding Experts Group (VCEG) together with the International Organization for Standardization (ISO)/International

Electrotechnical Commission (IEC) Moving Picture Experts Group (MPEG). H.264/AVC was co-chaired by Gary Sullivan and Thomas Wiegand and ended up doubling the performance of its predecessor, MPEG-2.

Sites like YouTube have transformed how we access and share video. Created in 2005 by three former PayPal employees, it enabled users to upload, view, and share videos. It uses Adobe Flash Video and HTML5 technology with advanced codecs such as H.264/AVC to display a wide variety of content such as movie clips, TV clips, music videos, amateur footage, and blogs. Google bought YouTube in 2006 for $1.65 billion. At the time of writing, it is reported that over 6 billion hours of video are watched each month on YouTube, with over 25% of this on mobile devices.[1] In 2013 YouTube and Netflix together accounted for approximately half of North America's peak download traffic.[2]

It is currently a very exciting and challenging time for video compression. The predicted growth in demand for bandwidth, driven largely by video applications, is probably greater now than it has ever been. There are four primary drivers for this:

1. Recently introduced formats such as 3-D and multiview, coupled with pressures for increased dynamic range, spatial resolution and frame rate, all require increased bit rate to deliver improved levels of immersion or interactivity.
2. Video-based web traffic continues to grow and dominate the internet through social networking and catch-up TV. In recent years YouTube has accounted for 27% of all video traffic and, by 2015, it is predicted that there will be 700 billion minutes of video downloaded.
3. User expectations continue to drive flexibility and quality, with a move from linear to non-linear delivery. Users are demanding "my-time" rather than "prime time" viewing.
4. New services, in particular mobile delivery through 4G/LTE to smart phones, has led some mobile network operators to predict that the demand for bandwidth will double every year for the next 10 years! While this can, to some extent, be addressed through efficiency improvements in network and physical layer technology, because video is the prime driver, the role of video compression is also enormously important.

The recent and ongoing standards activity, referred to as *High Efficiency Video Coding* (HEVC) chaired by Jens-Rainer Ohm, has driven the video community to challenge existing methods, delivering still further improvements. HEVC clearly represents the immediate future for compression and has produced impressive rate-quality gains. In the longer term, alternative approaches—perhaps merging conventional compression techniques with computer graphics—have the potential to create a new content-driven rate–quality optimization framework for video compression.

---

[1] YouTube Statistics: http://youtube.com/yt/press/en-GB/statistics.html.

[2] Sandvine Global Internet Phenomenon Report 1H 2013: https://www.sandvine.com/trends/global-internet-phenomena/.

This book

The importance of video compression and communications has increased enormously in recent years as more and more products and services are being designed around video content. And, with the emergence of new immersive services, coupled with increased user numbers and expectations, the tension between available bandwidth and picture quality has never been greater. Graduates with experience in video compression are, therefore, highly sought-after in industry across the world and most leading EEE, ECE, and CS departments now include coverage of this topic in their programmes.

Several good books have been written on image and video compression but few have taken the holistic approach presented here. *Communicating Pictures* is targeted primarily at 4th (Senior) year and Masters courses on image and/or video compression or coding. I have tried to balance theory with practice, so I hope it will also be of significant value to practitioners and those wishing to update their skills in industry.

In *Communicating Pictures* I have tried to place the topic in an application framework while also providing some historical context which I hope the reader will find interesting. The book is intended to provide sufficient mathematical background to enable a proper understanding of the material, but without detracting from the design and implementation challenges that are faced. The linkage between coding and assessment methods and our understanding of visual perception is also a key and is emphasized throughout. An algorithmic approach is adopted where appropriate, to capture the essence of the methods presented. While I have, on the whole, kept algorithmic developments independent of specific standards, I have used current standards (H.264/AVC and HEVC) as a reference where appropriate, especially in the context of hybrid motion compensated, block-based compression.

After an introduction to the motivation for compression and the description of a generic coding framework in Chapter 1, the book describes some basic concepts that underpin the topic in Chapters 2–4. Chapter 2 looks at the perceptual aspects of visual information, characterizing the limits and attributes of our visual system. This chapter concludes with a comprehensive listing of these attributes and how we can exploit redundancy to reduce bit rate and improve rate–quality performance. Chapter 3 provides a grounding in some of the more fundamental mathematical concepts that underpin discrete time signal and image processing, including: signal statistics, sampling, filters, transforms, information theory, quantization, and linear prediction. Chapter 4 looks, in more detail, at the practical basics, covering acquisition formats, coding structures, and color spaces, as well as providing an introduction to quality assessment and rate–distortion theory.

The most common basis for decorrelating still images is the block transform. This approach is described in Chapter 5, where decorrelating transforms are derived and their properties investigated. The primary focus of this chapter is on the workhorse of compression, the Discrete Cosine Transform (DCT), including associated algorithmic and implementation issues. Coverage of lossless coding is given in Chapter 6—both in its own right and in association with the entropy coding stages of lossy encoders.

This culminates in an operational description of a full image encoder–decoder (or codec) using JPEG as an example. An important alternative to block-based transform coding, based on wavelet filter banks, is addressed in Chapter 7. This explains the benefits and practical issues of the approach such as boundary extension, filter selection, and quantization.

Chapter 8 begins the journey into coding moving pictures, with a detailed description of motion estimation. The focus here is on the trade-offs between accuracy and complexity and a range of fast algorithms are described algorithmically and compared in terms of their efficiencies. This leads on to Chapter 9 which integrates the concepts and methods from previous chapters into a description of the ubiquitous hybrid motion compensated block-based codec. This chapter also introduces many of the codec features and refinements, such as sub-pixel and multiple reference frame estimation, variable block sizes, loop filtering, and intra-prediction, that have facilitated significant coding gains in recent years.

Chapter 10 addresses the important and challenging topic of measuring and managing the quality of digital pictures. This is related, throughout, to human perception, starting with an analysis of the advantages and disadvantages of mean squared error, then continuing to described various perceptually inspired objective metrics. This chapter also includes a detailed description of how to conduct subjective assessment trials, including aspects related to subject screening, test environment, test conditions, and post test analysis. Rate–Quality Optimization (RQO) is also covered here. It is only through effective RQO that modern codecs can make informed local mode decisions during coding.

Now that all of the components of the video codec have been described in Chapters 1–10. Chapter 11 offers insight into the problems associated with the transmission of coded video in real time over lossy and bandwidth-limited networks (wireless and fixed). It explains the conflicts between conventional video compression, based on variable length coding and spatio-temporal prediction, and the requirements for error-resilient transmission. It describes how the presence of packet and bit errors will influence a codec's performance and how we can mitigate their impact through packetization strategies, cross layer optimization, resilient entropy coding, pyramid vector quantization, and through the addition of source and channel coding redundancy.

The final two Chapters 12 and 13, provide a view of the current state of the art in standards and coding efficiency. Chapter 12 explains how the methods described throughout the book have been integrated into today's standards, specifically H.264/AVC and the most recent H.265/HEVC standard. The book concludes in Chapter 13 with a glimpse of how we might code and communicate pictures in the future, including examples of synthesis-based parametric and region-based compression methods. These are the topic of ongoing research and are candidates for future standardization.

I hope you find this book clear and concise. I have tried to make it informative and relevant, and to present concepts in a style that is rigorous yet interesting and accessible. I have included numerous worked examples together with many additional tutorial questions—hopefully these will help to reinforce the important messages and place them in a practical context.

I have always found picture processing to be a popular topic that students enjoy, perhaps because it so intimately relates technology to our everyday personal experiences. This connection between the visual senses and technology, where psychology, biology, and aesthetics meet mathematics, engineering, and computer science, is what makes the subject so fascinating. Also, perhaps, because we can create the illusion of high quality while throwing away most of the original data.

I hope you will benefit from and have pleasure reading this book. Enjoy the magic!

Additional Resources

For the solutions manual, a list of errata and selected high quality colour images, please visit the accompanying resources website at http://booksite.elsevier.com/9780124059061 or see the electronic version of this text.

A software (Matlab) based teaching aid has also been developed at the University of Bristol by Steve Ierodiaconou, Aaron Zhang, Alex Mackin, Paul Hill and myself. This provides a range of demonstrations that will enable students to better understand some of the issues associated with image and video compression. It can be accessed via the above website or directly from www.bristol.ac.uk/vi-lab.