Overview on Image Compression

Jose Oliver[1], Otoniel López[2], M. Martínez-Rach[2], Pablo Piñol[2],
Carlos T. Calafate[1], Manuel P. Malumbres[2]
[1] Technical University of Valencia, Spain
[2] Miguel Hernández University, Spain
e-mail: joliver@disca.upv.es

Introduction

Only a few decades ago, the human-computer interaction was based on a rudimentary text user interface, a convenient method compared to the punch card era, but too tedious and not very appealing for the non-specialist, and thereby not suitable for the mass market. Later on, the multimedia era arrived, with personal computers and other devices having powerful graphic capabilities, plenty of full-coloured pictures shown to the user. Although images made more pleasant the interaction with computers, their use represented a new challenge for electronic engineers; while only a few bytes are needed to represent a text (typically one byte per character in extended ASCII), lots of data must be employed for images, which in a “raw” representation (i.e., uncompressed) for colour images need as many as three bytes per single *pixel* (picture element, each dot forming an image). Thus, there was a clear urge to reduce the amount of bytes required to encode an image, mainly so as to avoid an excessive increase in both memory consumption and network bandwidth required to store and transmit images, which would limit or prevent their use in practice.

In general, in order to exploit the information system resources in an efficient way when dealing with images, compression is almost mandatory. Fortunately, most images are characterized by highly redundant signals (especially natural and synthetic images) since pixels composing an image present high homogeneity, and this redundancy, often called *spatial redundancy*, can be reduced through a compression process, achieving a more compact representation.

Background

In this article, several compression methods are briefly introduced, each focused on efficiently coding a different type of image.

We first consider continuous tone images (e.g., those from a digital camera), where each pixel takes a value in a (non-binary) range - typically any value in the range $[0..2^8-1]$ for a greyscale image. Actually, for colour images, three bytes per pixel are commonly used (one byte per colour component, red, green and blue, which is known as RGB colour space), where each pixel represents a colour from up to 2^{24} possibilities. The well-known standard JPEG (JPEG, 1992) focuses on this type of image. Its sequential mode, widely used throughout entertainment industry, is based on removing information that is hardly perceived by a human viewer (in particular, high-frequency components are less accurately encoded). Although the decompressed image is not equal to the original one if compared pixel by pixel, the perceptual quality could be nearly the same, provided that no heavy compression is applied. An evolution from this standard is the new JPEG 2000 standard (JPEG 2000, 2000), based on the same ideas, which is more efficient and flexible.

This type of image coding process, where recovered data is not the encoded one, is called *lossy compression*. It is important to emphasise that the comparison of lossy encoders cannot be performed based on the final image size alone, but also on its visual quality, in a sort of cost/benefit ratio (usually represented as a rate/distortion curve).

If it is important to recover exactly the original image, e.g. for legal reasons in medical imaging or in image editing, *lossless compression* can be done at the cost of lower compression performance (but no quality loss). Although JPEG 2000 has a lossless mode, specific standards, like JPEG-LS (JPEG-LS, 1997), offer better performance. Another famous lossless image encoder is GIF (CompuserveIncorporated, 1987). Frequently used on the Internet, it is intended for 256-colour images which are previously selected in a palette.

Finally, we deal with binary images where each pixel can take two different values. These images are typically employed in fax transmission, and are compressed using the JBIG standard (JBIG, 1993).

Continuous-tone still image compression

The JPEG standard

The most widely used algorithm for image coding is defined in the JPEG standard (JPEG, 1992). Introduced in the 1980s, and developed by the Joint Photographic Experts Group (from which the standard is named after), nowadays it is the most common way of encoding pictures in a wide range of important applications, such as image transmission on the Internet and image storage for digital cameras. This algorithm is able to encode colour images with an average compression rate of 15:1 with good visual quality (Furth, 1995).

The JPEG standard provides four working modes, three of them are lossy and the other one is lossless (Pennebaker & Mitchell, 1992). All the lossy modes employ the two-dimensional Discrete Cosine Transform (2D-DCT) to analyse the spatial-frequency features of the images so as to store with less precision (or even remove) those frequency components least important for a human observer (according to the human visual system). Another important role of the DCT is to achieve high compactness of the information: after the DCT is applied, a substantial part of the image information is concentrated in only a few transform coefficients, mainly the low frequency ones, and thus it can be represented in an efficient way.

Among its four compression modes, the sequential mode is extensively used, being the simplest and most well-known. In this mode, the input image can be a greyscale image or a colour image. Although a digital colour image is commonly represented in the RGB space by an image capture device, it can be stored more efficiently in YCbCr space, in which the luminance component (*Y*) is represented separately from the blue and red chrominance components (*Cb* and *Cr* respectively) (note that the remaining green component, *Cr*, is not needed because it can be inferred from the *Cb*, *Cr* and *Y* components). Since the human eye is more sensitive to brightness information than to colour one, the YCbCr colour representation allows us to reduce the size of the chrominance components without a significant degradation of the perceptual image quality. Thereby, the first step of the JPEG algorithm for colour image coding is a colour space transform from the input colour space to YCbCr, followed by a chrominance downsampling, usually a 4:2:0 subsampling, where the chrominance components are reduced by two in both horizontal and vertical directions (other possible downsampling is 4:2:2, in which only the horizontal direction is reduced by two, or 4:4:4, where there is no subsampling at all).

In the sequential mode, each image component is divided into 8×8 non-overlapping blocks, and they are compressed and transmitted (or stored) in scan order, from left to right, and from top to bottom, so that the decoder can recover the image sequentially, in the same order as it was encoded. Each block is then processed as follows:

(1) The two-dimensional DCT is applied to the entire block (it can be separately applied by using a 1D-DCT, first on the rows and then on the columns). Details on how to compute the DCT can be readily found in the literature (Ghanbari, 2003) (Pennebaker & Mitchell, 1992).

(2) The transform coefficients are then quantised to reduce information, most of all in high frequency components. This step is responsible for introducing information loss in the encoding process. The quantization process is done by dividing each coefficient by an associated constant value from a quantization matrix, rounding the result obtained to the nearest integer. This matrix is defined in a such manner that the higher frequency a coefficient represents, the higher the denominator (quantization value) becomes, and thereby more information reduction takes place.

(3) The DC component of the current block is differentially encoded by using as a reference the DC component of the previous block.

(4) The rest of components are scanned in zig-zag order, from lower to higher frequencies, and joint run-length and entropy coding is done. With the run-length coding, a count of zero-values is performed, while entropy coding is a statistical compression method that encodes symbols by using an amount of bits inversely proportional to the probability of its appearance (i.e., more likely symbols are encoded with fewer bits, and vice versa).

In addition, the JPEG algorithm allows varying the compression ratio by increasing the value of elements in the quantization matrix, at the expense of reducing image quality. This process is called rate control.

The other two lossy modes are the progressive and hierarchical modes, and both are based on the sequential one, adding more features to it. In particular:

(a) In the progressive mode, data from all the blocks is interleaved (by interleaving bit-planes, or coefficients, or both of them (Ghanbari, 2003)) so that a blurry full-resolution image can be displayed when the first bytes are received, and more bytes from the bitstream can be used to obtain more sharpened and defined versions of the image (this feature is known as quality scalability or SNR scalability).

(b) In the hierarchical mode, the original image is first downsampled by two several times, and then encoded and transmitted as in the sequential mode. Then, the encoder transmits the error committed by upsampling by two the low-resolution version of the image, and it repeats this process successively until the original resolution is achieved. By this process, the decoder is able to reconstruct larger images as more bytes are received, approaching the original resolution (this feature is called resolution scalability).

The last operation mode in JPEG is the lossless mode. In this coding method, no block-division and transform is done, being completely different in concept and usage to the lossy modes. In fact, it was introduced as a late addition a few years after the JPEG standard was released.

The JPEG 2000 standard

The JPEG 2000 standard (JPEG 2000, 2000) was proposed at the beginning of the new millennium, two decades after the JPEG standard was released, aiming at replacing it. Much research had been done in the field of image coding, and this new standard was intended to collect and apply much of it. During these years, a new mathematical tool, called Discrete Wavelet Transform (DWT), had aroused great interest in the field of image coding, mainly because it achieves high compactness of energy in the lower frequency subbands for natural images, which is extremely useful in image compression. Moreover, it allows resolution scalability in a natural way since more wavelet subbands are used to progressively enlarge the low frequency subbands (Mallat, 1989). For these reasons, the JPEG 2000 standard replaced the use of the DCT by the DWT.

Another advantage of the DWT compared to the DCT is that there are computationally efficient algorithms to apply the two-dimensional transform to the entire image as a whole, with no need to partition it into smaller blocks. One of the side effects of the block processing in the DCT, and perhaps the main problem of JPEG, is that blocking artifacts appear with moderate to high compression ratios. According to the human visual system, block edges are easily identified and, hence, the visual image quality is highly degraded. Moreover, when small blocks are independently encoded, spatial redundancy is not optimally removed from the image. The use of the wavelet transform also solves these problems.

As a result of the replacement of the DCT by the DWT, the JPEG 2000 standard offers superior coding performance, especially at high compression rates, where blocking artifacts are more noticeable. Despite of the bit rate saving of the new standard, maybe its main feature is a different one: it produces an extremely flexible and versatile bitstream. First, the user can indicate exactly the desired file size of the compressed image, while in the original JPEG only a quality parameter can be used (and the final file size is unknown a priori). Besides, after encoding the wavelet coefficients, the bitstream is reorganised to achieve different types of resolution and quality scalability, with almost no loss of coding performance (observe that the progressive and hierarchical modes are less efficient than the sequential one in JPEG).

Let us now briefly describe the image coding process of JPEG 2000. After a colour space transform (if needed), each image component is transformed by the DWT, which is applied to the entire component as a whole (unless there are memory limitations), and the resulting transform coefficients are quantized to obtain integer values. Then, the Embedded Block Coding with Optimized Truncation (EBCOT) algorithm (Taubman, 2000) is used to encode the coefficients in blocks (typically of 64×64 coefficients each block). Although block coding is also used in JPEG 2000, blocking artifacts do not appear because the image transform and quantization is applied to the entire image, and the block size is considerably bigger than in JPEG. The EBCOT algorithm consists in a bit-plane encoder (with three passes per bit-plane), followed by an optimization algorithm based on the Lagrange multiplier method (Everett, 1963), which is used to achieve the target file size in an optimal way, and a bitstream reorganization to fulfil the desired scalability.

Moreover, the original JPEG 2000 algorithm is able to work in lossless mode as described above, simply by utilizing a reversible integer-to-integer wavelet transform, which ensures the exact reversibility of the transform coefficient (avoiding possible rounding errors), and by skipping the quantization stages.

Specific lossless image encoders

Most lossless coders are based on entropy coding and predictive techniques. Predictive schemes try to calculate every sample from the previously encoded samples, which are available to the encoder and the decoder. In lossless image coding, prediction is usually done from nearby pixels. After computing a prediction, the residual pixel is entropy encoded as the difference between the prediction and the original value. Basically, this is how the lossless mode of the JPEG standard works, a process that differs significantly from the lossy modes defined by that same standard.

Clearly, the better a prediction is, the lower the residual error becomes. In order to improve the prediction, the CALIC scheme (Wu & Memon, 1996) uses a very large amount of contexts (i.e., different statistics for the entropy coding depending on the neighbour pixels), achieving one of the best coding performances among the continuous-tone still image lossless encoders. The high number of contexts makes CALIC quite complex, and therefore a simplification of CALIC was adopted as the JPEG-LS standard (JPEG-LS, 1997), which aims at improving and replacing the lossless mode of the original JPEG standard. The core algorithm of JPEG-LS is a simplified version of CALIC called LOCO-I (Weinberger, Seroussi & Sapiro, 2000), whose performance is close to CALIC with lower complexity. It is able to provide both lossless and "near lossless" image compression. Moreover, it is more efficient and much faster than JPEG 2000 working in lossless mode.

Lossless image coding is also used to transmit bitmaps on the Internet in well-known services such as the World Wide Web. Thus, the Graphics Interchange Format (GIF) (CompuserveIncorporated, 1987) was introduced to losslessly encode images formed by 256 colours, which are previously selected and stored in a colour table (also known as colour palette) from 2^{24} different possibilities (all the colours generated from a 24-bit RGB colour space). GIF images are compressed with the Lempel-Ziv-Welch (LZW) lossless encoder (Welch, 1984). The fact that LZW was a patented algorithm caused that the PNG format to be proposed as a royalty-free alternative to GIF, with an additional support for palettes of 24-bit RGB colours (PNG, 2003).

Binary image compression

The JBIG standard (JBIG, 1993) was released by the Joint Bi-level Image Experts Group as an efficient solution to encode images with only two possible values for each pixel (known as binary or bi-level images). One of the main applications of this standard is fax transmission. JBIG uses a special entropy coder, the Q-coder (Pennebaker, Mitchell, Langdon, Arps, 1988), as a basis to transmit the bi-level pixels in a lossless way.

An advanced version of this standard, called JBIG 2 (JBIG 2, 2001), was released in 2001. This new version is suitable for lossless, lossy and perceptually lossless compression, and provides large increases in coding performance. This new standard is able to take advantage of image segmentation by separating an image into text, picture, graphical regions, etc., and encoding each different segment with the most appropriate algorithm.

Future trends

One of the greatest doubts about the future of image compression is if JPEG 2000 will be able to become a widespread standard as occurred with the previous JPEG standard. At the time of writing this chapter, in early 2008, the support of images encoded with JPEG 2000 was not generalised in the World Wide Web, mainly because the most popular browsers (such as Internet Explorer and Firefox) do not support them yet. In order to improve the use of JPEG 2000 in interactive networks with client/server protocols, a new part of this standard has been recently released. This is Part 9, and is called JPIP. In this extension, a client can control and optimise the data flow downloaded from the server to meet the needed requirements.

Other new parts of the JPEG 2000 standard include Part 8 (JPSEC), which addresses security and content protection issues, Part 10 (JP3D), dealing with volumetric and 3D image coding, and Part 11 (JPWL), for image transmission on wireless networks, including: (a) protection of the bitstream against transmission errors (Part 1 only conceals errors), (b) detection of sensible areas of the bitstream, which will be protected more intensively, and (c) detection of errors that could not be corrected by the error protection decoder.

As for digital cameras, an increase of the native support of JPEG 2000 is expected because the specific hardware is becoming cheaper (e.g., Analog Devices, Inc has recently lowered the price of their JPEG 2000 chip from USD 30 to USD 8). On the other hand, JPEG 2000 is the standard adopted by the Digital Cinema industry for media content distribution (each movie frame is encoded with a JPEG 2000 based encoder).

The main drawback of JPEG 2000 implementations is that they are computationally-intensive and need a lot of memory to work, especially if compared to the sequential mode of JPEG. Thereby, alternative methods have recently been proposed (Oliver & Malumbres, 2006; Cho & Pearlman, 2007) to reduce the computational requirements while trying to preserve coding efficiency, many times at the cost of a less flexible bitstream than in JPEG 2000.

Conclusion

In this chapter we have surveyed the main existing standards for digital image compression. When dealing with continuous-tone images, JPEG is without doubt the most commonly employed compression method, being widely used in digital cameras and on the web. The JPEG 2000 standard aims at replacing it, although its use in these fields is still limited, and the success of this standard in the mass market is an open question. On the other hand, lossless coding is required at times (e.g., in medical imaging for legal issues, or in image editing, to avoid that accumulative errors from successive editions progressively damage the image quality). Hence, both JPEG and JPEG 2000 have lossless modes, though JPEG-LS, the specific standard for lossless image coding, is a better option in most cases because it is faster and more efficient. Finally, if the image to encode is not continuous-toned but binary instead, the standard solution is JBIG, which has been employed in several important applications like fax transmission.

References

Cho, Y. Pearlman, W. (2007), Hierarchical Dynamic Range Coding of Wavelet Subbands for Fast and Efficient Image Decompression, *IEEE Transactions on Image Processing*, 16(8), 2005-2015.

CompuserveIncorporated (1987), *Graphics Interchange Format specification version 89a*, http://www.w3.org/Graphics/GIF/spec-gif89a.txt.

Everett, H. (1963). Generalized Lagrange multiplier method for solving problems of optimum allocation of resources, *Operations Research*, 11(3), 399-417.

Furht, B. (1995). A survey of multimedia compression techniques and standards. Part I: JPEG standard, *Real-Time Imaging*, 1-49.

Ghanbari, M. (2003). “Standard Codecs: Image Compression to Advanced Video Coding”, The Institution of Electrical Engineers.

JBIG Standard (1993). *ISO/IEC 11544 (ITU-T Recommendation T.82), 1993.*

JBIG 2 Standard (2001). *ISO/IEC 14492 (ITU-T Recommendation T.88).*

JPEG Standard (1992). *ISO/IEC 10918-1 (ITU-T Recommendation T.81)*.

JPEG 2000 Standard (2000). *ISO/IEC 15444-1, (ITU-T Recommendation T.800)*.

JPEG-LS Standard (1997). *ISO/IEC 14495-1 (ITU-T Recommendation T.87)*.

Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674-693.

Oliver, J. Malumbres, M.P. (2006), Low-Complexity Multiresolution Image Compression Using Wavelet Lower Trees, *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11), 1437-1444.

Pennebaker, W.B. Mitchell, J.L. Langdon, G.G. Arps Jr., R. B. (1988), An overview of the basic principles of the Q-Coder adaptive binary arithmetic coder, *IBM Journal of Research and Development*, 32(6), 717.

Pennebaker, W.B., Mitchell, J.L. (1992). “JPEG Still Image Data Compression Standard”, *Van Nostrand Reinhold, International Thomson Publishing*.

Portable Network Graphics (PNG) Specification (Second Edition) (2003), Information technology- Computer graphics and image processing- Portable Network Graphics (PNG): Functional specification. *ISO/IEC 15948.*

Taubman, D. (2000). High Performance Scalable Image Compression with EBCOT, *IEEE Transactions on Image Processing*, 9(7), 1158-1170.

Weinberger, M. Seroussi, G. Sapiro, G. (2000). The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS, *IEEE Transactions on Image Processing*, 9(8), 1309-1324.

Welch, T.A. (1984). A Technique for High-Performance Data Compression, *Computer*, 17(6), 8-19.

Wu, X., Memon, N.D. (1996). Context Based, Adaptive, Lossless Image Coding, *IEEE Transactions on Communications*, 45(5), 437-444.

Terms and definitions

Chrominance: The difference between a certain colour component and the luminance component is called chrominance. The *Cb* and *Cr* components refer to the blue and red chrominances, increased by 0.5 and rescaled by 2 and 1.6 respectively (Pennebaker & Mitchell, 1992)

DC component: When the two-dimensional Discrete Cosine Transform is computed, the DC coefficient refers to the mean value of the pixels in the block, usually scaled according to the normalization used in the transform (in the DCT transform applied in JPEG, the DC is the mean value multiplied by 8).

Entropy coding: An entropy coder is a general lossless data compression method that encodes symbols by using an amount of bits inversely proportional to the probability of the symbols.

Human visual system (HVS): In order to improve coding performance, a lossy encoder removes the information that is not seen by the human eye. The human visual system is the part of the nervous system that allows us to see. It has been modelled to determine its behaviour, and hence to identify the information that can be removed without noticeable artifacts.

Lossless coding: In lossless image compression, the decoded pixels are exactly the same as those that were encoded. For this reason, we consider that there is no loss of data.

Lossy coding: An image encoder can modify a source image in order to achieve higher compression ratios, while trying to keep the perceived quality unaltered. This is the case of lossy image compression, in which the decoded pixels do not maintain the same values as when they were encoded.

Luminance: The brightness of an image is determined by the luminance component (usually referred to as Y). It is the main component in an YCbCr colour space because the HVS is more sensitive to this component than to chrominance components. It can be computed from the RGB components as $Y = 0.2126\ R + 0.7152\ G + 0.0722\ B$.