TECHNICAL SPECIFICATION

ISO/TS

12033

First edition
2001-11-15

Electronic imaging — Guidance for selection of document image compression methods

Imagerie électronique — Guide pour la sélection des méthodes de compression d'image

---

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years with a view to deciding whether it should be confirmed for a further three years, revised to become an International Standard, or withdrawn. In the case of a confirmed ISO/PAS or ISO/TS, it is reviewed again after six years at which time it has to be either transposed into an International Standard or withdrawn.

Attention is drawn to the possibility that some of the elements of this Technical Specification may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 12033 was prepared by Technical Committee ISO/TC 171, *Document imaging applications*, Subcommittee SC 2, *Application issues*.

Introduction

With respect to the rapid increase of applications using digitization techniques, the role of compression methods has become a factor of growing importance for the management of the volumes of stored data.

The effects of the available compression methods vary greatly, depending on the source documents. For example, an Electronic Image Management (EIM) system configured for scanning and storing continuous tone images will have different image compression requirements as compared to an application involving only text.

Practical methods for analyzing user requirements for image compression in order to select accurate and optimal image compression schemes are complex. It was evidently useful to issue this Technical Specification in order to guide users and system developers in their selection of these methods.

Electronic imaging — Guidance for selection of document image compression methods

1 Scope

This Technical Specification provides information to enable a user or EIM integrator to make an informed decision on selecting compression methods for digital images of business documents. It is designed to provide technical guidance to analyze the type of documents and which compression methods are most suitable for particular documents in order to optimize their storage and use.

For the user, this Technical Specification provides information on image compression methods incorporated in hardware or software in order to help this user during the selection of equipment in which the methods are embedded.

For the equipment or software designer, it provides planning information.

This Technical Specification is applicable only to still images in bit-map mode. It only takes into account compression algorithms based on well-tested mathematical work.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Specification. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this Technical Specification are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 12651:1999, *Electronic imaging — Vocabulary*

ITU-T Recommendation T.4:1999, *Standardization of Group 3 facsimile terminals for document transmission*

ITU-T Recommendation T.6:1988, *Facsimile coding schemes and coding control functions for group 4 facsimile apparatus*

3 Terms and definitions

For the purposes of this Technical Specification, the terms and definitions given in ISO 12651 and the following apply.

3.1
lossless compression
compression algorithm that is capable of recalling all of the original information of a compressed image

3.2
lossy compression
compression algorithm which loses some of the original information during compression, so that the decompressed image is only an approximation of the original

NOTE   This type of algorithm is especially useful in image compression where details can be eliminated, because these details are not perceptible, or are minimally perceptible, to the human eye. In this case, the compression ratio is dramatically increased.

3.3
resolution
number of pixels per unit of length

3.4
dots per inch
dpi
number of dots that a scanner (printer) can scan (print) per inch both horizontally or vertically

3.5
brightness
visual sensation that enables an observer to detect luminance

3.6
contrast
difference between the highest and the lowest densities of an image

3.7
bit level
number of bits used to define a pixel

3.8
luminance
Y
luminous flux emitted from a surface

NOTE   The former term was photometric brightness.

3.9
chrominance
Cr,Cb
colour portion of the video signal including hue and saturation but not brightness

NOTE   Low chroma means the colour picture looks pale or washed out; high chroma means intense colour; black, grey and white have a chrominance equal to zero.

3.10
ITU-T Group 3 and Group 4
standard compression algorithms set by the ITU-T

3.11
Joint Photographic Experts Group
JPEG
popular name of ISO/IEC 10994 standard

3.12
Comité Consultatif International pour le Télégraphe et le Téléphone
CCITT
former name of the International Telecommunication Union – Telecommunication Standardization sector (ITU-T)

3.13
compression ratio
ratio between image size before compression and image size after compression

4   General

In a document imaging system, users are concerned about the quality of archived images, for two reasons: first, because it can affect the imaging system's future in the medium or even long term; and second because they must choose the imaging tools based on an evolving technology.

The digitization process, which by nature transforms an image conveying comprehensible information into a dematerialized one, changes the observer's perception of that image. The observer may consider the image as being improved, though more frequently he considers it degraded. In fact, images undergo a number of successive transformations at different points during the digitization process. At each of these stages, attempts are made to keep the image within acceptable legibility limits, but also to restrict its size to within acceptable economic limits.

The specific role of one of the digitization stages — compression — is to reduce the size of the image. Some compression methods are reversible in that the decompression algorithm restores the initial digital information. These methods are lossless and have no impact on the quality of the image as it is perceived by the human eye. Other methods are lossy, and may cause degradation perceptible to the eye. By adjusting parameters, the user can bring a lossy method within acceptable limits.

While numerous compression methods are described in technical literature, few are stable according to industrial standards. These are based on a limited number of principles: dominance of certain patterns, pattern repetition, and noticeable mathematical properties. In any individual method, the number of parameters the user can modify is small.

The choice of a method and compression parameters are for a large part determined by the characteristics of the document. Obviously, the graphical contents of a document play a key role in determining the method and its parameters. However, other factors characterizing the application context are also very important (see diagram).

A document's graphical contents are themselves important to the digitization process. Thus, a photograph cannot be digitized in the same way if it is in greyscale or based on a "pseudo-grey" process. In the first case, JPEG compression is used, while the second would require ITU or JBIG compression.

Before discussing compression methods, therefore, we need to review the types of documents and how they are represented following digitization. See Figure 1.

5   Type of document and digitization parameters

5.1   General

A document is a set of organized information intended for presentation to a human user. Documents can be a single page or a set of pages, and can contain arbitrary contents types, such as character content, graphical content, and various types of image content.

The following document content may be founded in various types of documents. The classification list hereafter is somewhat arbitrary, but for a given application, these distinctions may be used to understand how to handle a given document.

5.2   Types of documents

Here we will present only those documents (generally called "word processing documents") that are most likely to be archived electronically. These documents include:

— black text on a white background, or less frequently, coloured text or a coloured background;

— photographs, black and white or colour;

— mixed documents containing both text and photographs reproduced by a printing process — black and white or colour.
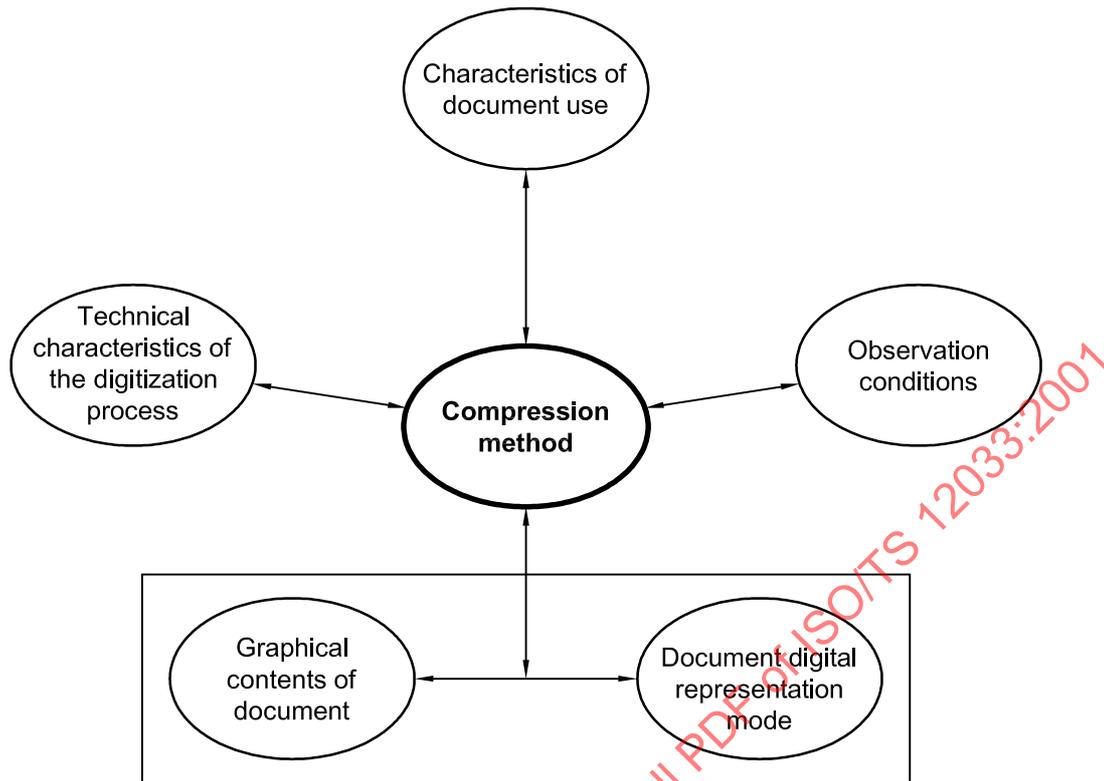
Figure 1 — Interactions with the compression method

5.3 Document classification and digitization

5.3.1 General

For the purpose of determining a compression scheme, documents may be described in the following five ways. For each type of document, digitization methods are briefly described.

5.3.2 Black and white documents

Digitizing pages printed in black and white (primarily text) generates bi-level images where each pixel is represented by a bit. This form of representation can also be applied to images in text documents with a coloured background or characters, as well as to line drawings.

The most important digitization parameter is resolution.

Resolution must be determined according to visual perception needs and on the limits of the complete imaging process (e.g. 200 dpi for word processing documents, 300 dpi for digitized books).

There are also other parameters, related to image processing, which vary according to the kind of image. If we know, for example, that the images to be digitized are text, we will try to produce black characters that are sharply defined against a white background. Thus, we have brightness (adjusting the colour of a pixel against a threshold) and contrast parameters (adjusting the colour of a pixel against that of the surrounding pixels).

5.3.3 Greyscale documents

This form of representation is applied to photographic documents, printed on paper from a black and white film.

Digitization changes an initially continuous document into a matrix of pixels whose intensity is encoded in a range of levels. Thus, 8-bit encoding produces 256 greyscales.

The number of greyscales or the bit level must be determined according to visual perception needs and the limits of the complete imaging process.

5.3.4 Pseudo-grey documents

This category includes images that simulate grey using a variable arrangement of black and white pixels. There can be two cases:

1) the source document is a photographic reproduction in a text; it was produced using a printing technique and is itself a pseudo-grey document (rastering uses black pixels of variable size);

2) the source document is a true photograph, but was digitized in pseudo-grey for performance reasons: to reduce the storage volume or transmission times on a network (the "half-tone" technique involves arranging a variable number of black pixels inside a matrix of fixed size).

5.3.5 Colour documents

This form of representation is applied to photographic documents, printed on paper from a colour film. Another application is digital colour capture of business documents where yellow highlights, colour boxes, pencil, red pen, etc. is a part of the information capture integrity.

Colour documents are intended to be restored in colour, but may also be reproduced in greyscale.

Colour representation is based on the neuro-physiological properties of the human eye, notably the "visual tri-variance" principle, which states that all colours can be produced by combining the three primary colours. Thus, a colour can be represented by three co-ordinates in a vector space based on primary colours, or by linear combinations of these co-ordinates.

The colour space most frequently adopted uses Red, Green and Blue. These colours are differentiated by the retinal cones in the eye. Another colour space decouples the variables into one "luminance" variable, and two "chrominance" variables. This colour space is used to transmit TV signal.

In a digitized colour image, each pixel is represented by assembling three components corresponding to the primary colours. The bit level adopted for a component determines the quality of hues; the standard of 8 bits per component can represent $256^3 = 16$ million different colours. Representations on a total of 8 bits sent by data communication networks are also fairly frequent.

5.3.6 Mixed documents

Many documents to be archived are composed of pages of text containing graphic elements and/or photographic images. There is no completely satisfactory way of representing this type of document:

⎯ a bi-level representation would make illustrations illegible;

⎯ a greyscale or colour representation to preserve illustrations would provide the best quality, but would make storage volumes disproportionately large with respect to the importance of the illustrations (one must be aware of the possible trade-offs between resolution and bit level of greyscale or colour image files);

⎯ a half-tone representation would decrease the legibility of the characters.

In mixed documents, text is considered more important, so a bi-level representation would be used to draw black characters on a white background. The photos would either be lost, or would have to be separated from the text for appropriate representation. In most cases, text and photos can be automatically and successfully separated using segmentation algorithms. Sometimes, segmentation can lead to loss of information (such as captions under photos, or unusual typographic arrangements).

6   Compression methods and standards

6.1   RLE compression (Run-Length Encoding)

This method is accepted by most graphic image formats.

This method takes into account runs of identical symbols inside data streams (which would be characters in an ASCII text). Each data stream is encoded with the number of occurrences of the repetitive elements and the length of the stream.

An RLE algorithm can operate at the bit, byte or pixel level. The basic algorithm works one line at a time, but some variations can also work vertically, taking into account repeating characters in adjacent lines. The RLE method is normally lossless, although to improve efficiency, some variations drop lower-order bits, resulting in loss.

This method is not very efficient for texts and complex photos, because there are few long sequences. It is most efficient for images with large areas of uniform colour.

6.2   LZW compression (Lempel-Ziv-Welch)

Variations of this method are found in several image formats.

The method seeks runs of repetitive elements not in a rigid manner, but by referring to a dictionary of data streams previously encountered which is enriched as the algorithm processes an image (for text, the dictionary is initialized with the ASCII alphabet).

6.3   ITU-T algorithms

6.3.1   General

ITU-T has defined a series of protocols for transmitting images via facsimile. These protocols are officially named T.4 and T.6, but are popularly known as the Group 3 and Group 4 methods. The compression methods used in archiving are variations of ITU-T. ITU-T may contain end-of-line and end-of-message codes to simplify fax transmissions. These codes are superfluous when these methods are used for archiving.

ITU-T compression is based on the variations of the Huffman algorithm.

ITU-T defines three fax standards, which are used for compressing bi-level images:

⎯ Group 3 Modified Huffman (MH): a one-dimensional compression method (G3 1D);

⎯ Group 3 Modified Read (MR): a two-dimensional compression method (G3 2D);

⎯ Group 4 Modified MR (MMR): a two-dimensional compression method (G4).

6.3.2   Group 3 one-dimensional method (G3 1D)

The Group 3 one-dimensional method (G3 1D) is a variation of the Huffman algorithm. In a bi-level image, each scanned line alternates variable-length zones, composed of black or white pixels. The Group 3 encoder determines the length of each black or white zone, called the run length, and looks up the corresponding code words in the Huffmann table.

Compression occurs because the code words are shorter than the zones they represent. Each code word represents a zone length corresponding to either white or black.

Group 3 is the basic compression algorithm used in Group 3 fax transmission.

The length of the code words was determined when the method was created, based on static observations of typed and hand-written documents. Run lengths with a high probability of occurrence were assigned the shortest code words.

NOTE   Although ITU-T compression was initially designed for text documents; it can also be applied to raster photos, although it is less efficient.

Sequences of pixels are represented by two types of code words: configuration code words and termination code words. Configuration code words represent long zones and termination code words represent short zones. A zone with a length of between 0 and 63 bits is encoded in a termination code word. A stream of between 64 and 2 623 bits is encoded in a configuration code word corresponding to the quotient of the length divided by 64; a termination code word can be added for the remainder. A stream with a length of over 2 623 bits is encoded as a series of configuration code words to which a termination code word can be added.

This one-dimensional encoding scheme eliminates redundancy only within each scan line, left to right. It does not reduce the redundancy between scan lines, up and down.

6.3.3   Group 3 two-dimensional method (G3 2D) and Group 4 method

Where the Group 3 one-dimensional method deals with each line of an image individually, the Group 3 two-dimensional method takes advantage of consistencies between two successive lines in the same image that are often very similar.

G3 2D is defined as an option of Group 3, which restricts itself to a small number of lines inserted between "one-dimensional" lines. Group 4 uses the same algorithm.

Like G3 1D, the G3 2D algorithm uses breakpoints that separate different colours in a single line ("mutant elements"). In creating an encoded representation of the image, the algorithm takes into account the mutant elements not only in a single line, but also in two adjacent lines. Thus, in addition to the code words used in G3 1D, the G3 2D and G4 methods use code words representing the distance and relative arrangement of mutant elements in two or more adjacent lines.

6.4   JBIG compression

JBIG is the abbreviation of Joint Bi-level Image Group. As its name indicates, this method is used for bi-level images. It is used primarily for text (like T.4 and T.6), though it can also be used for raster photos in printed documents (unlike T.4 and T.6). According to its authors, JBIG is as efficient as T.4 and T.6 for pure text, and 2 to 30 times more efficient for raster photos. Like T.4 and T.6, JBIG is lossless.

The method uses progressive encoding, which manipulates resolution. This encoding system initially transmits images in low resolution (e.g. 25 dpi). Then the resolution is progressively doubled until the resolution of the original image is obtained. There are two advantages to this progressive method: 1) it analyses images with just the necessary degree of detail; 2) it can adapt the resolution of an image according to the characteristics of the output peripheral or the perception needs of the human observer (e.g. transmission may be interrupted as soon as the image is recognized).

ISO/IEC 11544 describes the JBIG method as an assembly of procedure blocks. Most of these blocks have the dual purpose of saving encoding and accelerating processing. Encoding is saved when regions of uniform colour, repetitiveness, and similarities between the low and high-resolution image are detected.

6.5   JPEG compression

6.5.1   General

Generally, images are highly correlated from one pixel to the next. That is, if one pixel is a certain shade of grey, odds are that its neighbours are approximately that same shade. This means that some of the redundant information can be lost without a serious impact on the image's legibility. JPEG standard uses this method.

JPEG covers a family of algorithms and two types of compression.

One type is lossless, but is not very efficient. The more frequently used system is lossy. It contains several steps, some of which are the source of loss. The heart of this system is a mathematical transform known as the Discrete Cosine Transform (DCT).

6.5.2 Discrete Cosine Transform (DCT)

The DCT is applied to each element of each block (a block is an 8×8 pixel extracted from the digitized image). DCT transforms the block into a 64 frequency coefficients.

This transform has the interesting property of concentrating information in a small number of coefficients, with most other coefficients having a value close to zero.

6.5.3 JPEG steps

The JPEG method has several steps:

1) Discrete Cosine Transform;

2) a truncation induced by the digital representation of the data (very low-order values are rounded to zero);

3) the "quantization" transform;

4) "post-entropic encoding", which superimposes the Huffman coding technique or arithmetic compression to JPEG compression.

Except for DCT, the most important step in JPEG is quantization.

Quantization involves applying a linear transform to the coefficients obtained by the DCT. This quantization is used to eliminate frequencies with small contribution (mostly high frequencies) as well as to represent coarsely remaining frequencies.

A quantization matrix controls both the compression ratio obtained and image degradation. It is possible to fix a "level of loss."

The JPEG experts did not define the quantization matrix strictly in the standard. They simply gave an example adapted to a 720×576-pixel television screen. This example quantization matrix is based on psycho-visual testing with photographs.

6.5.4 Components of JPEG

JPEG primarily describes the principles of compression. Some, like DCT, are specific to JPEG, while others, such as the Huffman coding technique, are not.

Although JPEG does not impose a colour space, it can take advantage of the eye's most acute sensitivity to chromatic and luminance variations. Thus, in a Y, Cr, Cb encoding, it can sub-sample chrominance information with respect to luminance.

JPEG also introduces aspects related to image display dynamics, which are not directly linked to a compression principle. These subsidiary aspects affect the visual perception of the digitized image making them key criteria for JPEG.

— Interlacing: JPEG can transmit Y, Cr and Cb components in three sets, or interlaced.

— Progressive DCT encoding: the image is transmitted several times, with quality improved with each transmission.

The second principle can be used in interactive search systems: most images, which are not considered pertinent, are transmitted with minimal quality, while only those specifically sought are sent with maximum quality.

6.6   Fractal compression

This method is based on fractal geometry, attributed to the French mathematician Benoît Mandelbrot. Unlike analytical geometry which approaches complex curves using infinitesimals composed of line intervals, fractal geometry considers that the division of a curve (or a surface) into smaller elements does not reduce its complexity. A special category of fractals is created by the infinite repetition of the same pattern inside elements subdivided to infinity in a given curve.

Image compression using the fractal method is the opposite of generating fractal images. Instead of creating an image from a given transform, the purpose of compression is to find a series of transformations with which to approach a given digital image.

By analogy with fractal geometry, the basic purpose is to find similarities between the image elements obtained at different levels of subdivision: if all the small elements can be considered similar to large elements, they do not have to be described in detail, and compression is achieved.

In the fractal method, the algorithm seeks to compare image elements using matrix transforms. The algorithm includes an image subdivision principle and a metric definition which allows it to determine how similar the elements in an image are.

6.7   Wavelet compression

This method is based on research in mathematics expanding on Fourier's work in the 19th century.

A Fourier series can represent any periodic function as a linear combination of sinusoidal functions with variable frequencies. Modifications to the Fourier transform are needed to adapt to the representation of discrete valued functions such as those generated by digitization, and to deal with non-periodic functions. This leads to the DCT, the basis of JPEG. However, a Fourier transform is ill-suited for pronounced discontinuities in an image.

Wavelets form families of mathematical non-sinusoidal functions. These functions can have peaks and be rapidly decreasing, which means that they can be linked to non-periodic local characteristics of the image.

As in the DCT method, wavelet compression involves replacing the values generated by digitization with linear combination coefficients. The compression effect is achieved through the presence of near-zero coefficients.

7   Selecting compression parameters

7.1   Pertinence of compression

Since compression may or may not result in loss, our first inclination would be to use a lossless method. However, we are immediately faced with two problems:

1)   Currently, only ITU-T G3 and G4 and JPEG are standardized and industrialized on a large scale.

2)   The first difficulty would lead us to prefer JPEG for photographs. However, lossless JPEG is significantly less efficient than lossy JPEG.

Therefore, in the case of JPEG, it is most important to decide whether compression is even pertinent.

Serious consideration of the image's role in the imaging system should answer this question.

There can be two extreme situations:

1)   either the source document has a high aesthetic value, requiring images of the highest possible quality following digitization;

2)   or the document's contents are more important than its appearance, so a compromise between quality and efficiency can be reached as long as the digitized document remains legible.

Situation 1 would typically exclude lossy compression, while situation 2 allows compression even with loss, as a means of reducing the size of an image.

To a certain extent, the imaging system architecture can determine whether quality or efficiency is more important. Namely, one of the following decisions could be reached:

— to store uncompressed documents for editorial reproduction, and distribute compressed documents when needed for display on a computer screen (this is the principle of the "image catalogue", as opposed to a true illustrated document);

— to isolate some of the documents which will be digitized with higher-than-average quality; i.e. using lossless compression;

— to archive both paper and digitized documents so that only those paper documents requiring high quality reproduction are digitized when required.

7.2   Selecting a compression method

Once the decision to compress an image has been made, the user must choose the method best suited to each type of document in the collection. Table 1 summarizes the possible choices.

Table 1 — Compression methods

Type of document	Compression methods	Typical size before compression	Typical compression ratio	Typical size after compression
Text Line drawings Raster drawings	Group IV	1 MB	30	30 kB
Text Line drawings Raster photos	JBIG	1 MB	40	25 kB
Photos on photographic media (simplified shapes and colours)	RLE	100 kB	5	20 kB
Photos on photographic media (simplified shapes and colours)	LZW	100 kB	5	20 kB
Photos on photographic media (real-world images)	JPEG	40 MB	20	200 kB
Photos on photographic media (real-world images)	Fractals	40 MB	100	40 kB
Photos on photographic media (real-world images)	Wavelets	40 MB	100	40 kB

7.3   Adjusting JPEG compression

The compression ratio cannot be directly defined in JPEG. Moreover, such an adjustment would be senseless since the compression ratio in all compression methods varies according to the image's characteristics.

However, JPEG does allow the user to modify the compression ratio indirectly by adjusting the quality level. Obviously, the quality level is inversely proportional to the compression ratio (the highest level of quality has the lowest compression ratio).

There is no universal definition of quality. The concept is specific to each compression hardware or software product. The quality level is generally defined on a scale of 0 % to 100 %.