A note from VM Labs, Inc.

Here is the Independent JPEG Group's software, compiled for Merlin. The port was very straightforward, and involved only changing the makefile, configuration files, and adding an "in memory" source (see memdata.c). No Merlin specific optimization has been performed yet, which means that the code runs rather slowly compared to what could be achieved if, for example, the IDCT were re-coded in Merlin assembly language.

This library is provided as-is, as a "quick and dirty" way to get JPEG images up. BE PREPARED TO MODIFY ANY OF YOUR CODE WHICH USES THIS LIBRARY when we release a Merlin specific (and much faster!) JPEG decompresser. Believe me, JPEG can go much faster on Merlin.

----- begin libjpeg's Readme.txt inclusion -----

The Independent JPEG Group's JPEG software

README for release 6a of 7-Feb-96

This distribution contains the sixth public release of the Independent JPEG Group's free JPEG software. You are welcome to redistribute this software and to use it for any purpose, subject to the conditions under LEGAL ISSUES, below.

Serious users of this software (particularly those incorporating it into larger programs) should contact IJG at jpeg-info@uunet.uu.net to be added to our electronic mailing list. Mailing list members are notified of updates and have a chance to participate in technical discussions, etc.

This software is the work of Tom Lane, Philip Gladstone, Luis Ortiz, Jim Boucher, Lee Crocker, Julian Minguillon, George Phillips, Davide Rossi, Ge' Weijers, and other members of the Independent JPEG Group.

IJG is not affiliated with the official ISO JPEG standards committee.

DOCUMENTATION ROADMAP

This file contains the following sections:

OVERVIEW	General description of JPEG and the IJG software.
LEGAL ISSUES	Copyright, lack of warranty, terms of distribution.
REFERENCES	Where to learn more about JPEG.
ARCHIVE LOCATIONS	Where to find newer versions of this software.
RELATED SOFTWARE	Other stuff you should get.
FILE FORMAT WARS	Software *not* to get.
TO DO	Plans for future IJG releases.

Other documentation files in the distribution are:

```
User documentation:
  install.doc       How to configure and install the IJG software.
  usage.doc         Usage instructions for cjpeg, djpeg, jpegtran,
                    rdjpgcom, and wrjpgcom.
  *.1               Unix-style man pages for programs (same info as usage.doc).
  wizard.doc        Advanced usage instructions for JPEG wizards only.
  change.log        Version-to-version change highlights.
Programmer and internal documentation:
  libjpeg.doc       How to use the JPEG library in your own programs.
  example.c         Sample code for calling the JPEG library.
  structure.doc     Overview of the JPEG library's internal structure.
  filelist.doc      Road map of IJG files.
  coderules.doc     Coding style rules --- please read if you contribute code.
```

Please read at least the files install.doc and usage.doc. Useful information can also be found in the JPEG FAQ (Frequently Asked Questions) article. See ARCHIVE LOCATIONS below to find out where to obtain the FAQ article.

If you want to understand how the JPEG code works, we suggest reading one or more of the REFERENCES, then looking at the documentation files (in roughly the order listed) before diving into the code.

OVERVIEW

This package contains C software to implement JPEG image compression and decompression. JPEG (pronounced "jay-peg") is a standardized compression method for full-color and gray-scale images. JPEG is intended for compressing "real-world" scenes; line drawings, cartoons and other non-realistic images are not its strong suit. JPEG is lossy, meaning that the output image is not exactly identical to the input image. Hence you must not use JPEG if you have to have identical output bits. However, on typical photographic images, very good compression levels can be obtained with no visible change, and remarkably high compression levels are possible if you can tolerate a low-quality image. For more details, see the references, or just experiment with various compression settings.

This software implements JPEG baseline, extended-sequential, and progressive compression processes. Provision is made for supporting all variants of these processes, although some uncommon parameter settings aren't implemented yet. For legal reasons, we are not distributing code for the arithmetic-coding variants of JPEG; see LEGAL ISSUES. We have made no provision for supporting the hierarchical or lossless processes defined in the standard.

We provide a set of library routines for reading and writing JPEG image files, plus two sample applications "cjpeg" and "djpeg", which use the library to perform conversion between JPEG and some other popular image file formats. The library is intended to be reused in other applications.

In order to support file conversion and viewing software, we have included considerable functionality beyond the bare JPEG coding/decoding capability; for example, the color quantization modules are not strictly part of JPEG decoding, but they are essential for output to colormapped file formats or colormapped displays. These extra functions can be compiled out of the library if not required for a particular application. We have also included "jpegtran", a utility for lossless transcoding between different JPEG processes, and "rdjpgcom" and "wrjpgcom", two simple applications for inserting and extracting textual comments in JFIF files.

The emphasis in designing this software has been on achieving portability and flexibility, while also making it fast enough to be useful. In particular, the software is not intended to be read as a tutorial on JPEG. (See the REFERENCES section for introductory material.) Rather, it is intended to be reliable, portable, industrial-strength code. We do not claim to have achieved that goal in every aspect of the software, but we strive for it.

We welcome the use of this software as a component of commercial products. No royalty is required, but we do ask for an acknowledgement in product documentation, as described under LEGAL ISSUES.

LEGAL ISSUES

In plain English:

1. We don't promise that this software works. (But if you find any bugs, please let us know!)
2. You can use this software for whatever you want. You don't have to pay us.
3. You may not pretend that you wrote this software. If you use it in a program, you must acknowledge somewhere in your documentation that you've used the IJG code.

In legalese:

it. This software may be referred to only as "the Independent JPEG Group's software".

We specifically permit and encourage the use of this software as the basis of commercial products, provided that all warranty or liability claims are assumed by the product vendor.

ansi2knr.c is included in this distribution by permission of L. Peter Deutsch, sole proprietor of its copyright holder, Aladdin Enterprises of Menlo Park, CA. ansi2knr.c is NOT covered by the above copyright and conditions, but instead by the usual distribution terms of the Free Software Foundation; principally, that you must include source code if you redistribute it. (See the file ansi2knr.c for full details.) However, since ansi2knr.c is not needed as part of any program generated from the IJG code, this does not limit you more than the foregoing paragraphs do.

The configuration script "configure" was produced with GNU Autoconf. It is copyright by the Free Software Foundation but is freely distributable.

It appears that the arithmetic coding option of the JPEG spec is covered by patents owned by IBM, AT&T, and Mitsubishi. Hence arithmetic coding cannot legally be used without obtaining one or more licenses. For this reason, support for arithmetic coding has been removed from the free JPEG software. (Since arithmetic coding provides only a marginal gain over the unpatented Huffman mode, it is unlikely that very many implementations will support it.) So far as we are aware, there are no patent restrictions on the remaining code.

WARNING: Unisys has begun to enforce their patent on LZW compression against GIF encoders and decoders. You will need a license from Unisys to use the included rdgif.c or wrgif.c files in a commercial or shareware application. At this time, Unisys is not enforcing their patent against freeware, so distribution of this package remains legal. However, we intend to remove GIF support from the IJG package as soon as a suitable replacement format becomes reasonably popular.

We are required to state that

> "The Graphics Interchange Format(c) is the Copyright property of CompuServe Incorporated. GIF(sm) is a Service Mark property of CompuServe Incorporated."

REFERENCES

We highly recommend reading one or more of these references before trying to understand the innards of the JPEG software.

The best short technical introduction to the JPEG compression algorithm is

> Wallace, Gregory K. "The JPEG Still Picture Compression Standard", Communications of the ACM, April 1991 (vol. 34 no. 4), pp. 30-44.

(Adjacent articles in that issue discuss MPEG motion picture compression, applications of JPEG, and related topics.) If you don't have the CACM issue handy, a PostScript file containing a revised version of Wallace's article is available at ftp.uu.net, graphics/jpeg/wallace.ps.gz. The file (actually a preprint for an article that appeared in IEEE Trans. Consumer Electronics) omits the sample images that appeared in CACM, but it includes corrections and some added material. Note: the Wallace article is copyright ACM and IEEE, and it may not be used for commercial purposes.

A somewhat less technical, more leisurely introduction to JPEG can be found in "The Data Compression Book" by Mark Nelson, published by M&T Books (Redwood City, CA), 1991, ISBN 1-55851-216-0. This book provides good explanations and example C code for a multitude of compression methods including JPEG. It is an excellent source if you are comfortable reading C code but don't know much about data compression in general. The book's JPEG sample code is far from industrial-strength, but when you are ready to look at a full implementation, you've got one here...

The best full description of JPEG is the textbook "JPEG Still Image Data Compression Standard" by William B. Pennebaker and Joan L. Mitchell, published by Van Nostrand Reinhold, 1993, ISBN 0-442-01272-1. Price US$59.95, 638 pp. The book includes the complete text of the ISO JPEG standards (DIS 10918-1 and draft DIS 10918-2). This is by far the most complete exposition of JPEG in existence, and we highly recommend it.

The JPEG standard itself is not available electronically; you must order a paper copy through ISO or ITU. (Unless you feel a need to own a certified official copy, we recommend buying the Pennebaker and Mitchell book instead; it's much cheaper and includes a great deal of useful explanatory material.) In the USA, copies of the standard may be ordered from ANSI Sales at (212) 642-4900, or from Global Engineering Documents at (800) 854-7179. (ANSI doesn't take credit card orders, but Global does.) It's not cheap: as of 1992, ANSI was charging $95 for Part 1 and $47 for Part 2, plus 7% shipping/handling. The standard is divided into two parts, Part 1 being the actual specification, while Part 2 covers compliance testing methods. Part 1 is titled "Digital Compression and Coding of Continuous-tone Still Images, Part 1: Requirements and guidelines" and has document numbers ISO/IEC IS 10918-1, ITU-T T.81. Part 2 is titled "Digital Compression and Coding of Continuous-tone Still Images, Part 2: Compliance testing" and has document numbers ISO/IEC IS 10918-2, ITU-T T.83.

Extensions to the original JPEG standard are defined in JPEG Part 3, a new ISO document. Part 3 is undergoing ISO balloting and is expected to be approved by the end of 1995; it will have document numbers ISO/IEC IS 10918-3, ITU-T T.84. IJG currently does not support any Part 3 extensions.

The JPEG standard does not specify all details of an interchangeable file format. For the omitted details we follow the "JFIF" conventions, revision 1.02. A copy of the JFIF spec is available from:

      Literature Department
      C-Cube Microsystems, Inc.
      1778 McCarthy Blvd.
      Milpitas, CA 95035
      phone (408) 944-6300, fax (408) 944-6314

A PostScript version of this document is available at ftp.uu.net, file graphics/jpeg/jfif.ps.gz. It can also be obtained by e-mail from the C-Cube mail server, netlib@c3.pla.ca.us. Send the message "send jfif_ps from jpeg" to the server to obtain the JFIF document; send the message "help" if you have trouble.

The TIFF 6.0 file format specification can be obtained by FTP from sgi.com (192.48.153.1), file graphics/tiff/TIFF6.ps.Z; or you can order a printed copy from Aldus Corp. at (206) 628-6593. The JPEG incorporation scheme found in the TIFF 6.0 spec of 3-June-92 has a number of serious problems. IJG does not recommend use of the TIFF 6.0 design (TIFF Compression tag 6). Instead, we recommend the JPEG design proposed by TIFF Technical Note #2 (Compression tag 7). Copies of this Note can be obtained from sgi.com or from ftp.uu.net:/graphics/jpeg/. It is expected that the next revision of the TIFF spec will replace the 6.0 JPEG design with the Note's design. Although IJG's own code does not support TIFF/JPEG, the free libtiff library uses our library to implement TIFF/JPEG per the Note. libtiff is available from sgi.com:/graphics/tiff/.

ARCHIVE LOCATIONS

The "official" archive site for this software is ftp.uu.net (Internet address 192.48.96.9). The most recent released version can always be found there in directory graphics/jpeg. This particular version will be archived as graphics/jpeg/jpegsrc.v6a.tar.gz. If you are on the Internet, you can retrieve files from ftp.uu.net by standard anonymous FTP. If you don't have FTP access, UUNET's archives are also available via UUCP; contact help@uunet.uu.net for information on retrieving files that way.

Numerous Internet sites maintain copies of the UUNET files. However, only ftp.uu.net is guaranteed to have the latest official version.

You can also obtain this software in DOS-compatible "zip" archive format from the SimTel archives (ftp.coast.net:/SimTel/msdos/graphics/), or on CompuServe in the Graphics Support forum (GO CIS:GRAPHSUP), library 12 "JPEG Tools". Again, these versions may sometimes lag behind the ftp.uu.net release.

The JPEG FAQ (Frequently Asked Questions) article is a useful source of general information about JPEG. It is updated constantly and therefore is not included in this distribution. The FAQ is posted every two weeks to Usenet newsgroups comp.graphics.misc, news.answers, and other groups. You can always obtain the latest version from the news.answers archive at rtfm.mit.edu. By FTP, fetch /pub/usenet/news.answers/jpeg-faq/part1 and .../part2. If you don't have FTP, send e-mail to mail-server@rtfm.mit.edu with body

```
      send usenet/news.answers/jpeg-faq/part1
      send usenet/news.answers/jpeg-faq/part2
```

RELATED SOFTWARE

Numerous viewing and image manipulation programs now support JPEG. (Quite a few of them use this library to do so.) The JPEG FAQ described above lists some of the more popular free and shareware viewers, and tells where to obtain them on Internet.

If you are on a Unix machine, we highly recommend Jef Poskanzer's free PBMPLUS image software, which provides many useful operations on PPM-format image files. In particular, it can convert PPM images to and from a wide range of other formats. You can obtain this package by FTP from ftp.x.org (contrib/pbmplus*.tar.Z) or ftp.ee.lbl.gov (pbmplus*.tar.Z). There is also a newer update of this package called NETPBM, available from wuarchive.wustl.edu under directory /graphics/graphics/packages/NetPBM/. Unfortunately PBMPLUS/NETPBM is not nearly as portable as the IJG software is; you are likely to have difficulty making it work on any non-Unix machine.

A different free JPEG implementation, written by the PVRG group at Stanford, is available from havefun.stanford.edu in directory pub/jpeg. This program is designed for research and experimentation rather than production use; it is slower, harder to use, and less portable than the IJG code, but it is easier to read and modify. Also, the PVRG code supports lossless JPEG, which we do not.

FILE FORMAT WARS

Some JPEG programs produce files that are not compatible with our library. The root of the problem is that the ISO JPEG committee failed to specify a concrete file format. Some vendors "filled in the blanks" on their own, creating proprietary formats that no one else could read. (For example, none of the early commercial JPEG implementations for the Macintosh were able to exchange compressed files.)

The file format we have adopted is called JFIF (see REFERENCES). This format has been agreed to by a number of major commercial JPEG vendors, and it has become the de facto standard. JFIF is a minimal or "low end" representation. We recommend the use of TIFF/JPEG (TIFF revision 6.0 as modified by TIFF Technical Note #2) for "high end" applications that need to record a lot of additional data about an image. TIFF/JPEG is fairly new and not yet widely supported, unfortunately.

The upcoming JPEG Part 3 standard defines a file format called SPIFF. SPIFF is interoperable with JFIF, in the sense that most JFIF decoders should be able to read the most common variant of SPIFF. SPIFF has some technical advantages over JFIF, but its major claim to fame is simply that it is an official standard rather than an informal one. At this point it is unclear whether SPIFF will supersede JFIF or whether JFIF will remain the de-facto standard. IJG intends to support SPIFF once the standard is frozen, but we have not decided whether it should become our default output format or not. (In any case, our decoder will remain capable of reading JFIF indefinitely.)

Various proprietary file formats incorporating JPEG compression also exist. We have little or no sympathy for the existence of these formats. Indeed, one of the original reasons for developing this free software was to help force convergence on common, open format standards for JPEG files. Don't use a proprietary file format!

TO DO

In future versions, we are considering supporting some of the upcoming JPEG Part 3 extensions --- principally, variable quantization and the SPIFF file format.

Tuning the software for better behavior at low quality/high compression settings is also of interest. The current method for scaling the quantization tables is known not to be very good at low Q values.

As always, speeding things up is high on our priority list.

Please send bug reports, offers of help, etc. to jpeg-info@uunet.uu.net.

----- end Readme.txt inclusion -----
----- begin libjpeg.txt inclusion -----

USING THE IJG JPEG LIBRARY

This file describes how to use the IJG JPEG library within an application program. Read it if you want to write a program that uses the library.

The file example.c provides heavily commented skeleton code for calling the JPEG library. Also see jpeglib.h (the include file to be used by application programs) for full details about data structures and function parameter lists. The library source code, of course, is the ultimate reference.

Note that there have been *major* changes from the application interface presented by IJG version 4 and earlier versions. The old design had several inherent limitations, and it had accumulated a lot of cruft as we added features while trying to minimize application-interface changes. We have sacrificed backward compatibility in the version 5 rewrite, but we think the improvements justify this.

TABLE OF CONTENTS

Memory management
Library compile-time options
Portability considerations
Notes for MS-DOS implementors

You should read at least the overview and basic usage sections before trying to program with the library. The sections on advanced features can be read if and when you need them.

OVERVIEW

Functions provided by the library

The IJG JPEG library provides C code to read and write JPEG-compressed image files. The surrounding application program receives or supplies image data a scanline at a time, using a straightforward uncompressed image format. All details of color conversion and other preprocessing/postprocessing can be handled by the library.

The library includes a substantial amount of code that is not covered by the JPEG standard but is necessary for typical applications of JPEG. These functions preprocess the image before JPEG compression or postprocess it after decompression. They include colorspace conversion, downsampling/upsampling, and color quantization. The application indirectly selects use of this code by specifying the format in which it wishes to supply or receive image data. For example, if colormapped output is requested, then the decompression library automatically invokes color quantization.

A wide range of quality vs. speed tradeoffs are possible in JPEG processing, and even more so in decompression postprocessing. The decompression library provides multiple implementations that cover most of the useful tradeoffs, ranging from very-high-quality down to fast-preview operation. On the compression side we have generally not provided low-quality choices, since compression is normally less time-critical. It should be understood that the low-quality modes may not meet the JPEG standard's accuracy requirements; nonetheless, they are useful for viewers.

A word about functions *not* provided by the library. We handle a subset of the ISO JPEG standard; most baseline, extended-sequential, and progressive JPEG processes are supported. (Our subset includes all features now in common use.) Unsupported ISO options include:

* Hierarchical storage
* Lossless JPEG
* Arithmetic entropy coding (unsupported for legal reasons)
* DNL marker
* Nonintegral subsampling ratios

We support both 8- and 12-bit data precision, but this is a compile-time choice rather than a run-time choice; hence it is difficult to use both precisions in a single application.

By itself, the library handles only interchange JPEG datastreams --- in particular the widely used JFIF file format. The library can be used by surrounding code to process interchange or abbreviated JPEG datastreams that are embedded in more complex file formats. (For example, this library is used by the free LIBTIFF library to support JPEG compression in TIFF.)

Outline of typical usage

The rough outline of a JPEG compression operation is:

```
Allocate and initialize a JPEG compression object
Specify the destination for the compressed data (eg, a file)
Set parameters for compression, including image size & colorspace
jpeg_start_compress(...);
while (scan lines remain to be written)
      jpeg_write_scanlines(...);
jpeg_finish_compress(...);
Release the JPEG compression object
```

A JPEG compression object holds parameters and working state for the JPEG library. We make creation/destruction of the object separate from starting or finishing compression of an image; the same object can be re-used for a series of image compression operations. This makes it easy to re-use the same parameter settings for a sequence of images. Re-use of a JPEG object also has important implications for processing abbreviated JPEG datastreams, as discussed later.

The image data to be compressed is supplied to jpeg_write_scanlines() from in-memory buffers. If the application is doing file-to-file compression, reading image data from the source file is the application's responsibility. The library emits compressed data by calling a "data destination manager", which typically will write the data into a file; but the application can provide its own destination manager to do something else.

Similarly, the rough outline of a JPEG decompression operation is:

```
Allocate and initialize a JPEG decompression object
Specify the source of the compressed data (eg, a file)
Call jpeg_read_header() to obtain image info
Set parameters for decompression
jpeg_start_decompress(...);
while (scan lines remain to be read)
      jpeg_read_scanlines(...);
jpeg_finish_decompress(...);
Release the JPEG decompression object
```

This is comparable to the compression outline except that reading the datastream header is a separate step. This is helpful because information about the image's size, colorspace, etc is available when the application selects decompression parameters. For example, the application can choose an output scaling ratio that will fit the image into the available screen size.

The decompression library obtains compressed data by calling a data source manager, which typically will read the data from a file; but other behaviors can be obtained with a custom source manager. Decompressed data is delivered into in-memory buffers passed to jpeg_read_scanlines().

It is possible to abort an incomplete compression or decompression operation by calling jpeg_abort(); or, if you do not need to retain the JPEG object, simply release it by calling jpeg_destroy().

JPEG compression and decompression objects are two separate struct types. However, they share some common fields, and certain routines such as jpeg_destroy() can work on either type of object.

The JPEG library has no static variables: all state is in the compression or decompression object. Therefore it is possible to process multiple compression and decompression operations concurrently, using multiple JPEG objects.

Both compression and decompression can be done in an incremental memory-to-memory fashion, if suitable source/destination managers are used. See the section on "I/O suspension" for more details.

BASIC LIBRARY USAGE

Data formats

Before diving into procedural details, it is helpful to understand the image data format that the JPEG library expects or returns.

The standard input image format is a rectangular array of pixels, with each pixel having the same number of "component" or "sample" values (color channels). You must specify how many components there are and the colorspace interpretation of the components. Most applications will use RGB data (three components per pixel) or grayscale data (one component per pixel). PLEASE NOTE THAT RGB DATA IS THREE SAMPLES PER PIXEL, GRAYSCALE ONLY ONE. A remarkable number of people manage to miss this, only to find that their programs don't work with grayscale JPEG files.

There is no provision for colormapped input. JPEG files are always full-color or full grayscale (or sometimes another colorspace such as CMYK). You can feed in a colormapped image by expanding it to full-color format. However JPEG often doesn't work very well with source data that has been colormapped, because of dithering noise. This is discussed in more detail in the JPEG FAQ and the other references mentioned in the README file.

Pixels are stored by scanlines, with each scanline running from left to right. The component values for each pixel are adjacent in the row; for example, R,G,B,R,G,B,R,G,B,... for 24-bit RGB color. Each scanline is an array of data type JSAMPLE --- which is typically "unsigned char", unless you've changed jmorecfg.h. (You can also change the RGB pixel layout, say to B,G,R order, by modifying jmorecfg.h. But see the restrictions listed in that file before doing so.)

A 2-D array of pixels is formed by making a list of pointers to the starts of scanlines; so the scanlines need not be physically adjacent in memory. Even if you process just one scanline at a time, you must make a one-element pointer array to conform to this structure. Pointers to JSAMPLE rows are of type JSAMPROW, and the pointer to the pointer array is of type JSAMPARRAY.

The library accepts or supplies one or more complete scanlines per call. It is not possible to process part of a row at a time. Scanlines are always processed top-to-bottom. You can process an entire image in one call if you have it all in memory, but usually it's simplest to process one scanline at a time.

For best results, source data values should have the precision specified by BITS_IN_JSAMPLE (normally 8 bits). For instance, if you choose to compress data that's only 6 bits/channel, you should left-justify each value in a byte before passing it to the compressor. If you need to compress data that has more than 8 bits/channel, compile with BITS_IN_JSAMPLE = 12. (See "Library compile-time options", later.)

The data format returned by the decompressor is the same in all details, except that colormapped output is supported. (Again, a JPEG file is never colormapped. But you can ask the decompressor to perform on-the-fly color quantization to deliver colormapped output.) If you request colormapped output then the returned data array contains a single JSAMPLE per pixel; its value is an index into a color map. The color map is represented as a 2-D JSAMPARRAY in which each row holds the values of one color component, that is, colormap[i][j] is the value of the i'th color component for pixel value (map index) j. Note that since the colormap indexes are stored in JSAMPLEs, the maximum number of colors is limited by the size of JSAMPLE (ie, at most 256 colors for an 8-bit JPEG library).

Compression details

Here we revisit the JPEG compression outline given in the overview.

1. Allocate and initialize a JPEG compression object.

A JPEG compression object is a "struct jpeg_compress_struct". (It also has a bunch of subsidiary structures which are allocated via malloc(), but the application doesn't control those directly.) This struct can be just a local variable in the calling routine, if a single routine is going to execute the whole JPEG compression sequence. Otherwise it can be static or allocated from malloc().

You will also need a structure representing a JPEG error handler. The part of this that the library cares about is a "struct jpeg_error_mgr". If you are providing your own error handler, you'll typically want to embed the jpeg_error_mgr struct in a larger structure; this is discussed later under "Error handling". For now we'll assume you are just using the default error handler. The default error handler will print JPEG error/warning messages on stderr, and it will call exit() if a fatal error occurs.

You must initialize the error handler structure, store a pointer to it into the JPEG object's "err" field, and then call jpeg_create_compress() to initialize the rest of the JPEG object.

Typical code for this step, if you are using the default error handler, is

```
struct jpeg_compress_struct cinfo;
struct jpeg_error_mgr jerr;
...
cinfo.err = jpeg_std_error(&jerr);
```

```
jpeg_create_compress(&cinfo);
```

jpeg_create_compress allocates a small amount of memory, so it could fail if you are out of memory. In that case it will exit via the error handler; that's why the error handler must be initialized first.

2. Specify the destination for the compressed data (eg, a file).

As previously mentioned, the JPEG library delivers compressed data to a "data destination" module. The library includes one data destination module which knows how to write to a stdio stream. You can use your own destination module if you want to do something else, as discussed later.

If you use the standard destination module, you must open the target stdio stream beforehand. Typical code for this step looks like:

```
FILE * outfile;
...
if ((outfile = fopen(filename, "wb")) == NULL) {
    fprintf(stderr, "can't open %s\n", filename);
    exit(1);
}
jpeg_stdio_dest(&cinfo, outfile);
```

where the last line invokes the standard destination module.

WARNING: it is critical that the binary compressed data be delivered to the output file unchanged. On non-Unix systems the stdio library may perform newline translation or otherwise corrupt binary data. To suppress this behavior, you may need to use a "b" option to fopen (as shown above), or use setmode() or another routine to put the stdio stream in binary mode. See cjpeg.c and djpeg.c for code that has been found to work on many systems.

You can select the data destination after setting other parameters (step 3), if that's more convenient. You may not change the destination between calling jpeg_start_compress() and jpeg_finish_compress().

3. Set parameters for compression, including image size & colorspace.

You must supply information about the source image by setting the following fields in the JPEG object (cinfo structure):

```
image_width       Width of image, in pixels
image_height            Height of image, in pixels
input_components  Number of color channels (samples per pixel)
in_color_space          Color space of source image
```

The image dimensions are, hopefully, obvious. JPEG supports image dimensions of 1 to 64K pixels in either direction. The input color space is typically RGB or grayscale, and input_components is 3 or 1 accordingly. (See "Special color spaces", later, for more info.) The in_color_space field must be assigned one of the J_COLOR_SPACE enum constants, typically JCS_RGB or JCS_GRAYSCALE.

JPEG has a large number of compression parameters that determine how the image is encoded. Most applications don't need or want to know about all these parameters. You can set all the parameters to reasonable defaults by calling jpeg_set_defaults(); then, if there are particular values you want to change, you can do so after that. The "Compression parameter selection" section tells about all the parameters.

You must set in_color_space correctly before calling jpeg_set_defaults(), because the defaults depend on the source image colorspace. However the other three source image parameters need not be valid until you call jpeg_start_compress(). There's no harm in calling jpeg_set_defaults() more than once, if that happens to be convenient.

Typical code for a 24-bit RGB source image is

```
cinfo.image_width = Width;    /* image width and height, in pixels */
cinfo.image_height = Height;
cinfo.input_components = 3;   /* # of color components per pixel */
cinfo.in_color_space = JCS_RGB; /* colorspace of input image */

jpeg_set_defaults(&cinfo);
/* Make optional parameter settings here */
```

4. jpeg_start_compress(...);

After you have established the data destination and set all the necessary source image info and other parameters, call jpeg_start_compress() to begin a compression cycle. This will initialize internal state, allocate working storage, and emit the first few bytes of the JPEG datastream header.

Typical code:

```
jpeg_start_compress(&cinfo, TRUE);
```

The "TRUE" parameter ensures that a complete JPEG interchange datastream will be written. This is appropriate in most cases. If you think you might want to use an abbreviated datastream, read the section on abbreviated datastreams, below.

Once you have called jpeg_start_compress(), you may not alter any JPEG parameters or other fields of the JPEG object until you have completed the compression cycle.

5. while (scan lines remain to be written)
      jpeg_write_scanlines(...);

Now write all the required image data by calling jpeg_write_scanlines() one or more times. You can pass one or more scanlines in each call, up to the total image height. In most applications it is convenient to pass just one or a few scanlines at a time. The expected format for the passed data is discussed under "Data formats", above.

Image data should be written in top-to-bottom scanline order. The JPEG spec contains some weasel wording about how top and bottom are application-defined terms (a curious interpretation of the English language...) but if you want your files to be compatible with everyone else's, you WILL use top-to-bottom order. If the source data must be read in bottom-to-top order, you can use the JPEG library's virtual array mechanism to invert the data efficiently. Examples of this can be found in the sample application cjpeg.

The library maintains a count of the number of scanlines written so far in the next_scanline field of the JPEG object. Usually you can just use this variable as the loop counter, so that the loop test looks like "while (cinfo.next_scanline < cinfo.image_height)".

Code for this step depends heavily on the way that you store the source data. example.c shows the following code for the case of a full-size 2-D source array containing 3-byte RGB pixels:

```
	JSAMPROW row_pointer[1];	/* pointer to a single row */
	int row_stride;			/* physical row width in buffer */

	row_stride = image_width * 3;	/* JSAMPLEs per row in image_buffer */

	while (cinfo.next_scanline < cinfo.image_height) {
	    row_pointer[0] = & image_buffer[cinfo.next_scanline * row_stride];
	    jpeg_write_scanlines(&cinfo, row_pointer, 1);
	}
```

jpeg_write_scanlines() returns the number of scanlines actually written. This will normally be equal to the number passed in, so you can usually ignore the return value. It is different in just two cases:

* If you try to write more scanlines than the declared image height, the additional scanlines are ignored.
* If you use a suspending data destination manager, output buffer overrun will cause the compressor to return before accepting all the passed lines. This feature is discussed under "I/O suspension", below. The normal stdio destination manager will NOT cause this to happen.

In any case, the return value is the same as the change in the value of next_scanline.

6. jpeg_finish_compress(...);

After all the image data has been written, call jpeg_finish_compress() to complete the compression cycle. This step is ESSENTIAL to ensure that the last bufferload of data is written to the data destination. jpeg_finish_compress() also releases working memory associated with the JPEG object.

Typical code:

```
	jpeg_finish_compress(&cinfo);
```

If using the stdio destination manager, don't forget to close the output stdio stream if necessary.

If you have requested a multi-pass operating mode, such as Huffman code optimization, jpeg_finish_compress() will perform the additional passes using data buffered by the first pass. In this case jpeg_finish_compress() may take quite a while to complete. With the default compression parameters, this will not happen.

It is an error to call jpeg_finish_compress() before writing the necessary total number of scanlines. If you wish to abort compression, call jpeg_abort() as discussed below.

After completing a compression cycle, you may dispose of the JPEG object as discussed next, or you may use it to compress another image. In that case return to step 2, 3, or 4 as appropriate. If you do not change the destination manager, the new datastream will be written to the same target. If you do not change any JPEG parameters, the new datastream will be written with the same parameters as before. Note that you can change the input image dimensions freely between cycles, but if you change the input colorspace, you should call jpeg_set_defaults() to adjust for the new colorspace; and then you'll need to repeat all of step 3.

7. Release the JPEG compression object.

When you are done with a JPEG compression object, destroy it by calling jpeg_destroy_compress(). This will free all subsidiary memory. Or you can call jpeg_destroy() which works for either compression or decompression objects --- this may be more convenient if you are sharing code between compression and decompression cases. (Actually, these routines are equivalent except for the declared type of the passed pointer. To avoid gripes from ANSI C compilers, jpeg_destroy() should be passed a j_common_ptr.)

If you allocated the jpeg_compress_struct structure from malloc(), freeing it is your responsibility --- jpeg_destroy() won't. Ditto for the error handler structure.

Typical code:

```
	jpeg_destroy_compress(&cinfo);
```

8. Aborting.

If you decide to abort a compression cycle before finishing, you can clean up in either of two ways:

* If you don't need the JPEG object any more, just call jpeg_destroy_compress() or jpeg_destroy() to release memory. This is legitimate at any point after calling jpeg_create_compress() --- in fact, it's safe even if jpeg_create_compress() fails.

* If you want to re-use the JPEG object, call jpeg_abort_compress(), or jpeg_abort() which works on both compression and decompression objects. This will return the object to an idle state, releasing any working memory. jpeg_abort() is allowed at any time after successful object creation.

Note that cleaning up the data destination, if required, is your responsibility.

Decompression details

Here we revisit the JPEG decompression outline given in the overview.

1. Allocate and initialize a JPEG decompression object.

This is just like initialization for compression, as discussed above, except that the object is a "struct jpeg_decompress_struct" and you call jpeg_create_decompress(). Error handling is exactly the same.

Typical code:

```
struct jpeg_decompress_struct cinfo;
struct jpeg_error_mgr jerr;
...
cinfo.err = jpeg_std_error(&jerr);
jpeg_create_decompress(&cinfo);
```

(Both here and in the IJG code, we usually use variable name "cinfo" for both compression and decompression objects.)

2. Specify the source of the compressed data (eg, a file).

As previously mentioned, the JPEG library reads compressed data from a "data source" module. The library includes one data source module which knows how to read from a stdio stream. You can use your own source module if you want to do something else, as discussed later.

If you use the standard source module, you must open the source stdio stream beforehand. Typical code for this step looks like:

```
FILE * infile;
...
if ((infile = fopen(filename, "rb")) == NULL) {
    fprintf(stderr, "can't open %s\n", filename);
    exit(1);
}
jpeg_stdio_src(&cinfo, infile);
```

where the last line invokes the standard source module.

WARNING: it is critical that the binary compressed data be read unchanged. On non-Unix systems the stdio library may perform newline translation or otherwise corrupt binary data. To suppress this behavior, you may need to use a "b" option to fopen (as shown above), or use setmode() or another routine to put the stdio stream in binary mode. See cjpeg.c and djpeg.c for code that has been found to work on many systems.

You may not change the data source between calling jpeg_read_header() and jpeg_finish_decompress(). If you wish to read a series of JPEG images from a single source file, you should repeat the jpeg_read_header() to jpeg_finish_decompress() sequence without reinitializing either the JPEG object or the data source module; this prevents buffered input data from being discarded.

3. Call jpeg_read_header() to obtain image info.

Typical code for this step is just

```
        jpeg_read_header(&cinfo, TRUE);
```

This will read the source datastream header markers, up to the beginning of the compressed data proper. On return, the image dimensions and other info have been stored in the JPEG object. The application may wish to consult this information before selecting decompression parameters.

More complex code is necessary if

* A suspending data source is used --- in that case jpeg_read_header() may return before it has read all the header data. See "I/O suspension", below. The normal stdio source manager will NOT cause this to happen.
* Abbreviated JPEG files are to be processed --- see the section on abbreviated datastreams. Standard applications that deal only in interchange JPEG files need not be concerned with this case either.

It is permissible to stop at this point if you just wanted to find out the image dimensions and other header info for a JPEG file. In that case, call jpeg_destroy() when you are done with the JPEG object, or call jpeg_abort() to return it to an idle state before selecting a new data source and reading another header.

4. Set parameters for decompression.

jpeg_read_header() sets appropriate default decompression parameters based on the properties of the image (in particular, its colorspace). However, you may well want to alter these defaults before beginning the decompression. For example, the default is to produce full color output from a color file. If you want colormapped output you must ask for it. Other options allow the returned image to be scaled and allow various speed/quality tradeoffs to be selected. "Decompression parameter selection", below, gives details.

If the defaults are appropriate, nothing need be done at this step.

Note that all default values are set by each call to jpeg_read_header(). If you reuse a decompression object, you cannot expect your parameter settings to be preserved across cycles, as you can for compression. You must set desired parameter values each time.

5. jpeg_start_decompress(...);

Once the parameter values are satisfactory, call jpeg_start_decompress() to begin decompression. This will initialize internal state, allocate working memory, and prepare for returning data.

Typical code is just

```
        jpeg_start_decompress(&cinfo);
```

If you have requested a multi-pass operating mode, such as 2-pass color quantization, jpeg_start_decompress() will do everything needed before data output can begin. In this case jpeg_start_decompress() may take quite a while to complete. With a single-scan (non progressive) JPEG file and default decompression parameters, this will not happen; jpeg_start_decompress() will return quickly.

After this call, the final output image dimensions, including any requested scaling, are available in the JPEG object; so is the selected colormap, if colormapped output has been requested. Useful fields include

```
	output_width		image width and height, as scaled
	output_height
	out_color_components	# of color components in out_color_space
	output_components	# of color components returned per pixel
	colormap		the selected colormap, if any
	actual_number_of_colors		number of entries in colormap
```

output_components is 1 (a colormap index) when quantizing colors; otherwise it equals out_color_components. It is the number of JSAMPLE values that will be emitted per pixel in the output arrays.

Typically you will need to allocate data buffers to hold the incoming image. You will need output_width * output_components JSAMPLEs per scanline in your output buffer, and a total of output_height scanlines will be returned.

Note: if you are using the JPEG library's internal memory manager to allocate data buffers (as djpeg does), then the manager's protocol requires that you request large buffers *before* calling jpeg_start_decompress(). This is a little tricky since the output_XXX fields are not normally valid then. You can make them valid by calling jpeg_calc_output_dimensions() after setting the relevant parameters (scaling, output color space, and quantization flag).

```
6. while (scan lines remain to be read)
	jpeg_read_scanlines(...);
```

Now you can read the decompressed image data by calling jpeg_read_scanlines() one or more times. At each call, you pass in the maximum number of scanlines to be read (ie, the height of your working buffer); jpeg_read_scanlines() will return up to that many lines. The return value is the number of lines actually read. The format of the returned data is discussed under "Data formats", above. Don't forget that grayscale and color JPEGs will return different data formats!

Image data is returned in top-to-bottom scanline order. If you must write out the image in bottom-to-top order, you can use the JPEG library's virtual array mechanism to invert the data efficiently. Examples of this can be found in the sample application djpeg.

The library maintains a count of the number of scanlines returned so far in the output_scanline field of the JPEG object. Usually you can just use this variable as the loop counter, so that the loop test looks like "while (cinfo.output_scanline < cinfo.output_height)". (Note that the test should NOT be against image_height, unless you never use scaling. The image_height field is the height of the original unscaled image.) The return value always equals the change in the value of output_scanline.

If you don't use a suspending data source, it is safe to assume that jpeg_read_scanlines() reads at least one scanline per call, until the bottom of the image has been reached.

If you use a buffer larger than one scanline, it is NOT safe to assume that jpeg_read_scanlines() fills it. (The current implementation won't return more than cinfo.rec_outbuf_height scanlines per call, no matter how large a buffer you pass.) So you must always provide a loop that calls jpeg_read_scanlines() repeatedly until the whole image has been read.

7. jpeg_finish_decompress(...);

After all the image data has been read, call jpeg_finish_decompress() to complete the decompression cycle. This causes working memory associated with the JPEG object to be released.

Typical code:

```
jpeg_finish_decompress(&cinfo);
```

If using the stdio source manager, don't forget to close the source stdio stream if necessary.

It is an error to call jpeg_finish_decompress() before reading the correct total number of scanlines. If you wish to abort compression, call jpeg_abort() as discussed below.

After completing a decompression cycle, you may dispose of the JPEG object as discussed next, or you may use it to decompress another image. In that case return to step 2 or 3 as appropriate. If you do not change the source manager, the next image will be read from the same source.

8. Release the JPEG decompression object.

When you are done with a JPEG decompression object, destroy it by calling jpeg_destroy_decompress() or jpeg_destroy(). The previous discussion of destroying compression objects applies here too.

Typical code:

```
jpeg_destroy_decompress(&cinfo);
```

9. Aborting.

You can abort a decompression cycle by calling jpeg_destroy_decompress() or jpeg_destroy() if you don't need the JPEG object any more, or jpeg_abort_decompress() or jpeg_abort() if you want to reuse the object. The previous discussion of aborting compression cycles applies here too.

Mechanics of usage: include files, linking, etc

Applications using the JPEG library should include the header file jpeglib.h to obtain declarations of data types and routines. Before including jpeglib.h, include system headers that define at least the typedefs FILE and size_t. On ANSI-conforming systems, including <stdio.h> is sufficient; on older Unix systems, you may need <sys/types.h> to define size_t.

If the application needs to refer to individual JPEG library error codes, also include jerror.h to define those symbols.

jpeglib.h indirectly includes the files jconfig.h and jmorecfg.h. If you are installing the JPEG header files in a system directory, you will want to install all four files: jpeglib.h, jerror.h, jconfig.h, jmorecfg.h.

The most convenient way to include the JPEG code into your executable program is to prepare a library file ("libjpeg.a", or a corresponding name on non-Unix machines) and reference it at your link step. If you use only half of the library (only compression or only decompression), only that much code will be included from the library, unless your linker is hopelessly brain-damaged. The supplied makefiles build libjpeg.a automatically (see install.doc).

On some systems your application may need to set up a signal handler to ensure that temporary files are deleted if the program is interrupted. This is most critical if you are on MS-DOS and use the jmemdos.c memory manager back end; it will try to grab extended memory for temp files, and that space will NOT be freed automatically. See cjpeg.c or djpeg.c for an example signal handler.

It may be worth pointing out that the core JPEG library does not actually require the stdio library: only the default source/destination managers and error handler need it. You can use the library in a stdio-less environment if you replace those modules and use jmemnobs.c (or another memory manager of your own devising). More info about the minimum system library requirements may be found in jinclude.h.

ADVANCED FEATURES

Compression parameter selection

This section describes all the optional parameters you can set for JPEG compression, as well as the "helper" routines provided to assist in this task. Proper setting of some parameters requires detailed understanding of the JPEG standard; if you don't know what a parameter is for, it's best not to mess with it! See REFERENCES in the README file for pointers to more info about JPEG.

It's a good idea to call jpeg_set_defaults() first, even if you plan to set all the parameters; that way your code is more likely to work with future JPEG libraries that have additional parameters. For the same reason, we recommend you use a helper routine where one is provided, in preference to twiddling cinfo fields directly.

The helper routines are:

jpeg_set_defaults (j_compress_ptr cinfo)
    This routine sets all JPEG parameters to reasonable defaults, using only the input image's color space (field in_color_space, which must already be set in cinfo). Many applications will only need to use this routine and perhaps jpeg_set_quality().

jpeg_set_colorspace (j_compress_ptr cinfo, J_COLOR_SPACE colorspace)
    Sets the JPEG file's colorspace (field jpeg_color_space) as specified,

```
	and sets other color-space-dependent parameters appropriately.  See
	"Special color spaces", below, before using this.  A large number of
	parameters, including all per-component parameters, are set by this
	routine; if you want to twiddle individual parameters you should call
	jpeg_set_colorspace() before rather than after.

jpeg_default_colorspace (j_compress_ptr cinfo)
	Selects an appropriate JPEG colorspace based on cinfo->in_color_space,
	and calls jpeg_set_colorspace().  This is actually a subroutine of
	jpeg_set_defaults().  It's broken out in case you want to change
	just the colorspace-dependent JPEG parameters.

jpeg_set_quality (j_compress_ptr cinfo, int quality, boolean force_baseline)
	Constructs JPEG quantization tables appropriate for the indicated
	quality setting.  The quality value is expressed on the 0..100 scale
	recommended by IJG (cjpeg's "-quality" switch uses this routine).
	Note that the exact mapping from quality values to tables may change
	in future IJG releases as more is learned about DCT quantization.
	If the force_baseline parameter is TRUE, then the quantization table
	entries are constrained to the range 1..255 for full JPEG baseline
	compatibility.  In the current implementation, this only makes a
	difference for quality settings below 25, and it effectively prevents
	very small/low quality files from being generated.  The IJG decoder
	is capable of reading the non-baseline files generated at low quality
	settings when force_baseline is FALSE, but other decoders may not be.

jpeg_set_linear_quality (j_compress_ptr cinfo, int scale_factor,
			 boolean force_baseline)
	Same as jpeg_set_quality() except that the generated tables are the
	sample tables given in the JPEC spec section K.1, multiplied by the
	specified scale factor (which is expressed as a percentage; thus
	scale_factor = 100 reproduces the spec's tables).  Note that larger
	scale factors give lower quality.  This entry point is useful for
	conforming to the Adobe PostScript DCT conventions, but we do not
	recommend linear scaling as a user-visible quality scale otherwise.
	force_baseline again constrains the computed table entries to 1..255.

int jpeg_quality_scaling (int quality)
	Converts a value on the IJG-recommended quality scale to a linear
	scaling percentage.  Note that this routine may change or go away
	in future releases --- IJG may choose to adopt a scaling method that
	can't be expressed as a simple scalar multiplier, in which case the
	premise of this routine collapses.  Caveat user.

jpeg_add_quant_table (j_compress_ptr cinfo, int which_tbl,
		      const unsigned int *basic_table,
		      int scale_factor, boolean force_baseline)
	Allows an arbitrary quantization table to be created.  which_tbl
	indicates which table slot to fill.  basic_table points to an array
	of 64 unsigned ints given in normal array order.  These values are
	multiplied by scale_factor/100 and then clamped to the range 1..65535
	(or to 1..255 if force_baseline is TRUE).
	CAUTION: prior to library version 6a, jpeg_add_quant_table expected
	the basic table to be given in JPEG zigzag order.  If you need to
	write code that works with either older or newer versions of this
	routine, you must check the library version number.  Something like
	"#if JPEG_LIB_VERSION >= 61" is the right test.
```

jpeg_simple_progression (j_compress_ptr cinfo)
	Generates a default scan script for writing a progressive-JPEG file.
	This is the recommended method of creating a progressive file,
	unless you want to make a custom scan sequence. You must ensure that
	the JPEG color space is set correctly before calling this routine.

Compression parameters (cinfo fields) include:

J_DCT_METHOD dct_method
	Selects the algorithm used for the DCT step. Choices are:
		JDCT_ISLOW: slow but accurate integer algorithm
		JDCT_IFAST: faster, less accurate integer method
		JDCT_FLOAT: floating-point method
		JDCT_DEFAULT: default method (normally JDCT_ISLOW)
		JDCT_FASTEST: fastest method (normally JDCT_IFAST)
	The FLOAT method is very slightly more accurate than the ISLOW method,
	but may give different results on different machines due to varying
	roundoff behavior. The integer methods should give the same results
	on all machines. On machines with sufficiently fast FP hardware, the
	floating-point method may also be the fastest. The IFAST method is
	considerably less accurate than the other two; its use is not
	recommended if high quality is a concern. JDCT_DEFAULT and
	JDCT_FASTEST are macros configurable by each installation.

J_COLOR_SPACE jpeg_color_space
int num_components
	The JPEG color space and corresponding number of components; see
	"Special color spaces", below, for more info. We recommend using
	jpeg_set_color_space() if you want to change these.

boolean optimize_coding
	TRUE causes the compressor to compute optimal Huffman coding tables
	for the image. This requires an extra pass over the data and
	therefore costs a good deal of space and time. The default is
	FALSE, which tells the compressor to use the supplied or default
	Huffman tables. In most cases optimal tables save only a few percent
	of file size compared to the default tables. Note that when this is
	TRUE, you need not supply Huffman tables at all, and any you do
	supply will be overwritten.

unsigned int restart_interval
int restart_in_rows
	To emit restart markers in the JPEG file, set one of these nonzero.
	Set restart_interval to specify the exact interval in MCU blocks.
	Set restart_in_rows to specify the interval in MCU rows. (If
	restart_in_rows is not 0, then restart_interval is set after the
	image width in MCUs is computed.) Defaults are zero (no restarts).

const jpeg_scan_info * scan_info
int num_scans
	By default, scan_info is NULL; this causes the compressor to write a
	single-scan sequential JPEG file. If not NULL, scan_info points to
	an array of scan definition records of length num_scans. The
	compressor will then write a JPEG file having one scan for each scan
	definition record. This is used to generate noninterleaved or

progressive JPEG files. The library checks that the scan array
defines a valid JPEG scan sequence. (jpeg_simple_progression creates
a suitable scan definition array for progressive JPEG.) This is
discussed further under "Progressive JPEG support".

int smoothing_factor
	If non-zero, the input image is smoothed; the value should be 1 for
	minimal smoothing to 100 for maximum smoothing. Consult jcsample.c
	for details of the smoothing algorithm. The default is zero.

boolean write_JFIF_header
	If TRUE, a JFIF APP0 marker is emitted. jpeg_set_defaults() and
	jpeg_set_colorspace() set this TRUE if a JFIF-legal JPEG color space
	(ie, YCbCr or grayscale) is selected, otherwise FALSE.

UINT8 density_unit
UINT16 X_density
UINT16 Y_density
	The resolution information to be written into the JFIF marker;
	not used otherwise. density_unit may be 0 for unknown,
	1 for dots/inch, or 2 for dots/cm. The default values are 0,1,1
	indicating square pixels of unknown size.

boolean write_Adobe_marker
	If TRUE, an Adobe APP14 marker is emitted. jpeg_set_defaults() and
	jpeg_set_colorspace() set this TRUE if JPEG color space RGB, CMYK,
	or YCCK is selected, otherwise FALSE. It is generally a bad idea
	to set both write_JFIF_header and write_Adobe_marker. In fact,
	you probably shouldn't change the default settings at all --- the
	default behavior ensures that the JPEG file's color space can be
	recognized by the decoder.

JQUANT_TBL * quant_tbl_ptrs[NUM_QUANT_TBLS]
	Pointers to coefficient quantization tables, one per table slot,
	or NULL if no table is defined for a slot. Usually these should
	be set via one of the above helper routines; jpeg_add_quant_table()
	is general enough to define any quantization table. The other
	routines will set up table slot 0 for luminance quality and table
	slot 1 for chrominance.

JHUFF_TBL * dc_huff_tbl_ptrs[NUM_HUFF_TBLS]
JHUFF_TBL * ac_huff_tbl_ptrs[NUM_HUFF_TBLS]
	Pointers to Huffman coding tables, one per table slot, or NULL if
	no table is defined for a slot. Slots 0 and 1 are filled with the
	JPEG sample tables by jpeg_set_defaults(). If you need to allocate
	more table structures, jpeg_alloc_huff_table() may be used.
	Note that optimal Huffman tables can be computed for an image
	by setting optimize_coding, as discussed above; there's seldom
	any need to mess with providing your own Huffman tables.

There are some additional cinfo fields which are not documented here
because you currently can't change them; for example, you can't set
arith_code TRUE because arithmetic coding is unsupported.

Per-component parameters are stored in the struct cinfo.comp_info[i] for
component number i. Note that components here refer to components of the

JPEG color space, *not* the source image color space. A suitably large comp_info[] array is allocated by jpeg_set_defaults(); if you choose not to use that routine, it's up to you to allocate the array.

int component_id
    The one-byte identifier code to be recorded in the JPEG file for this component. For the standard color spaces, we recommend you leave the default values alone.

int h_samp_factor
int v_samp_factor
    Horizontal and vertical sampling factors for the component; must be 1..4 according to the JPEG standard. Note that larger sampling factors indicate a higher-resolution component; many people find this behavior quite unintuitive. The default values are 2,2 for luminance components and 1,1 for chrominance components, except for grayscale where 1,1 is used.

int quant_tbl_no
    Quantization table number for component. The default value is 0 for luminance components and 1 for chrominance components.

int dc_tbl_no
int ac_tbl_no
    DC and AC entropy coding table numbers. The default values are 0 for luminance components and 1 for chrominance components.

int component_index
    Must equal the component's index in comp_info[]. (Beginning in release v6, the compressor library will fill this in automatically; you don't have to.)

Decompression parameter selection

Decompression parameter selection is somewhat simpler than compression parameter selection, since all of the JPEG internal parameters are recorded in the source file and need not be supplied by the application. (Unless you are working with abbreviated files, in which case see "Abbreviated datastreams", below.) Decompression parameters control the postprocessing done on the image to deliver it in a format suitable for the application's use. Many of the parameters control speed/quality tradeoffs, in which faster decompression may be obtained at the price of a poorer-quality image. The defaults select the highest quality (slowest) processing.

The following fields in the JPEG object are set by jpeg_read_header() and may be useful to the application in choosing decompression parameters:

```
JDIMENSION image_width              Width and height of image
JDIMENSION image_height
int num_components                  Number of color components
J_COLOR_SPACE jpeg_color_space            Colorspace of image
boolean saw_JFIF_marker             TRUE if a JFIF APP0 marker was seen
  UINT8 density_unit                Resolution data from JFIF marker
  UINT16 X_density
```

```
  UINT16 Y_density
boolean saw_Adobe_marker            TRUE if an Adobe APP14 marker was seen
  UINT8 Adobe_transform             Color transform code from Adobe marker
```

The JPEG color space, unfortunately, is something of a guess since the JPEG standard proper does not provide a way to record it. In practice most files adhere to the JFIF or Adobe conventions, and the decoder will recognize these correctly. See "Special color spaces", below, for more info.

The decompression parameters that determine the basic properties of the returned image are:

J_COLOR_SPACE out_color_space

Output color space. jpeg_read_header() sets an appropriate default based on jpeg_color_space; typically it will be RGB or grayscale. The application can change this field to request output in a different colorspace. For example, set it to JCS_GRAYSCALE to get grayscale output from a color file. (This is useful for previewing: grayscale output is faster than full color since the color components need not be processed.) Note that not all possible color space transforms are currently implemented; you may need to extend jdcolor.c if you want an unusual conversion.

unsigned int scale_num, scale_denom

Scale the image by the fraction scale_num/scale_denom. Default is 1/1, or no scaling. Currently, the only supported scaling ratios are 1/1, 1/2, 1/4, and 1/8. (The library design allows for arbitrary scaling ratios but this is not likely to be implemented any time soon.) Smaller scaling ratios permit significantly faster decoding since fewer pixels need be processed and a simpler IDCT method can be used.

boolean quantize_colors

If set TRUE, colormapped output will be delivered. Default is FALSE, meaning that full-color output will be delivered.

The next three parameters are relevant only if quantize_colors is TRUE.

int desired_number_of_colors

Maximum number of colors to use in generating a library-supplied color map (the actual number of colors is returned in a different field). Default 256. Ignored when the application supplies its own color map.

boolean two_pass_quantize

If TRUE, an extra pass over the image is made to select a custom color map for the image. This usually looks a lot better than the one-size-fits-all colormap that is used otherwise. Default is TRUE. Ignored when the application supplies its own color map.

J_DITHER_MODE dither_mode

Selects color dithering method. Supported values are:

```
JDITHER_NONE      no dithering: fast, very low quality
JDITHER_ORDERED   ordered dither: moderate speed and quality
JDITHER_FS  Floyd-Steinberg dither: slow, high quality
```

Default is JDITHER_FS. (At present, ordered dither is implemented only in the single-pass, standard-colormap case. If you ask for ordered dither when two_pass_quantize is TRUE or when you supply an external color map, you'll get F-S dithering.)

When quantize_colors is TRUE, the target color map is described by the next two fields. colormap is set to NULL by jpeg_read_header(). The application can supply a color map by setting colormap non-NULL and setting actual_number_of_colors to the map size. Otherwise, jpeg_start_decompress() selects a suitable color map and sets these two fields itself. [Implementation restriction: at present, an externally supplied colormap is only accepted for 3-component output color spaces.]

JSAMPARRAY colormap
    The color map, represented as a 2-D pixel array of out_color_components rows and actual_number_of_colors columns. Ignored if not quantizing. CAUTION: if the JPEG library creates its own colormap, the storage pointed to by this field is released by jpeg_finish_decompress(). Copy the colormap somewhere else first, if you want to save it.

int actual_number_of_colors
    The number of colors in the color map.

Additional decompression parameters that the application may set include:

J_DCT_METHOD dct_method
    Selects the algorithm used for the DCT step. Choices are the same as described above for compression.

boolean do_fancy_upsampling
    If TRUE, do careful upsampling of chroma components. If FALSE, a faster but sloppier method is used. Default is TRUE. The visual impact of the sloppier method is often very small.

boolean do_block_smoothing
    If TRUE, interblock smoothing is applied in early stages of decoding progressive JPEG files; if FALSE, not. Default is TRUE. Early progression stages look "fuzzy" with smoothing, "blocky" without. In any case, block smoothing ceases to be applied after the first few AC coefficients are known to full accuracy, so it is relevant only when using buffered-image mode for progressive images.

boolean enable_1pass_quant
boolean enable_external_quant
boolean enable_2pass_quant
    These are significant only in buffered-image mode, which is described in its own section below.

The output image dimensions are given by the following fields. These are computed from the source image dimensions and the decompression parameters by jpeg_start_decompress(). You can also call jpeg_calc_output_dimensions() to obtain the values that will result from the current parameter settings. This can be useful if you are trying to pick a scaling ratio that will get close to a desired target size. It's also important if you are using the JPEG library's memory manager to allocate output buffer space, because you are supposed to request such buffers *before* jpeg_start_decompress().

JDIMENSION output_width   Actual dimensions of output image.
JDIMENSION output_height

```
int out_color_components      Number of color components in out_color_space.
int output_components         Number of color components returned.
int rec_outbuf_height         Recommended height of scanline buffer.
```

When quantizing colors, output_components is 1, indicating a single color map index per pixel. Otherwise it equals out_color_components. The output arrays are required to be output_width * output_components JSAMPLEs wide.

rec_outbuf_height is the recommended minimum height (in scanlines) of the buffer passed to jpeg_read_scanlines(). If the buffer is smaller, the library will still work, but time will be wasted due to unnecessary data copying. In high-quality modes, rec_outbuf_height is always 1, but some faster, lower-quality modes set it to larger values (typically 2 to 4). If you are going to ask for a high-speed processing mode, you may as well go to the trouble of honoring rec_outbuf_height so as to avoid data copying.

Special color spaces

The JPEG standard itself is "color blind" and doesn't specify any particular color space. It is customary to convert color data to a luminance/chrominance color space before compressing, since this permits greater compression. The existing de-facto JPEG file format standards specify YCbCr or grayscale data (JFIF), or grayscale, RGB, YCbCr, CMYK, or YCCK (Adobe). For special applications such as multispectral images, other color spaces can be used, but it must be understood that such files will be unportable.

The JPEG library can handle the most common colorspace conversions (namely RGB <=> YCbCr and CMYK <=> YCCK). It can also deal with data of an unknown color space, passing it through without conversion. If you deal extensively with an unusual color space, you can easily extend the library to understand additional color spaces and perform appropriate conversions.

For compression, the source data's color space is specified by field in_color_space. This is transformed to the JPEG file's color space given by jpeg_color_space. jpeg_set_defaults() chooses a reasonable JPEG color space depending on in_color_space, but you can override this by calling jpeg_set_colorspace(). Of course you must select a supported transformation. jccolor.c currently supports the following transformations:

```
      RGB => YCbCr
      RGB => GRAYSCALE
      YCbCr => GRAYSCALE
      CMYK => YCCK
```

plus the null transforms: GRAYSCALE => GRAYSCALE, RGB => RGB, YCbCr => YCbCr, CMYK => CMYK, YCCK => YCCK, and UNKNOWN => UNKNOWN.

The de-facto file format standards (JFIF and Adobe) specify APPn markers that indicate the color space of the JPEG file. It is important to ensure that these are written correctly, or omitted if the JPEG file's color space is not one of the ones supported by the de-facto standards. jpeg_set_colorspace() will set the compression parameters to include or omit the APPn markers properly, so long as it is told the truth about the JPEG color space. For example, if you are writing some random 3-component color space without conversion, don't try to fake out the library by setting in_color_space and jpeg_color_space to JCS_YCbCr; use JCS_UNKNOWN. You may want to write an APPn marker of your own devising to identify the colorspace --- see "Special markers", below.

When told that the color space is UNKNOWN, the library will default to using luminance-quality compression parameters for all color components. You may well want to change these parameters. See the source code for jpeg_set_colorspace(), in jcparam.c, for details.

For decompression, the JPEG file's color space is given in jpeg_color_space, and this is transformed to the output color space out_color_space. jpeg_read_header's setting of jpeg_color_space can be relied on if the file conforms to JFIF or Adobe conventions, but otherwise it is no better than a guess. If you know the JPEG file's color space for certain, you can override jpeg_read_header's guess by setting jpeg_color_space. jpeg_read_header also selects a default output color space based on (its guess of) jpeg_color_space; set out_color_space to override this. Again, you must select a supported transformation. jdcolor.c currently supports

```
YCbCr => GRAYSCALE
YCbCr => RGB
YCCK => CMYK
```

as well as the null transforms.

The two-pass color quantizer, jquant2.c, is specialized to handle RGB data (it weights distances appropriately for RGB colors). You'll need to modify the code if you want to use it for non-RGB output color spaces. Note that jquant2.c is used to map to an application-supplied colormap as well as for the normal two-pass colormap selection process.

CAUTION: it appears that Adobe Photoshop writes inverted data in CMYK JPEG files: 0 represents 100% ink coverage, rather than 0% ink as you'd expect. This is arguably a bug in Photoshop, but if you need to work with Photoshop CMYK files, you will have to deal with it in your application. We cannot "fix" this in the library by inverting the data during the CMYK<=>YCCK transform, because that would break other applications, notably Ghostscript. Photoshop versions prior to 3.0 write EPS files containing JPEG-encoded CMYK data in the same inverted-YCCK representation used in bare JPEG files, but the surrounding PostScript code performs an inversion using the PS image operator. I am told that Photoshop 3.0 will write uninverted YCCK in EPS/JPEG files, and will omit the PS-level inversion. (But the data polarity used in bare JPEG files will not change in 3.0.) In either case, the JPEG library must not invert the data itself, or else Ghostscript would read these EPS files incorrectly.

Error handling

When the default error handler is used, any error detected inside the JPEG routines will cause a message to be printed on stderr, followed by exit(). You can supply your own error handling routines to override this behavior and to control the treatment of nonfatal warnings and trace/debug messages. The file example.c illustrates the most common case, which is to have the application regain control after an error rather than exiting.

The JPEG library never writes any message directly; it always goes through the error handling routines. Three classes of messages are recognized:

* Fatal errors: the library cannot continue.
* Warnings: the library can continue, but the data is corrupt, and a

damaged output image is likely to result.
* Trace/informational messages. These come with a trace level indicating the importance of the message; you can control the verbosity of the program by adjusting the maximum trace level that will be displayed.

You may, if you wish, simply replace the entire JPEG error handling module (jerror.c) with your own code. However, you can avoid code duplication by only replacing some of the routines depending on the behavior you need. This is accomplished by calling jpeg_std_error() as usual, but then overriding some of the method pointers in the jpeg_error_mgr struct, as illustrated by example.c.

All of the error handling routines will receive a pointer to the JPEG object (a j_common_ptr which points to either a jpeg_compress_struct or a jpeg_decompress_struct; if you need to tell which, test the is_decompressor field). This struct includes a pointer to the error manager struct in its "err" field. Frequently, custom error handler routines will need to access additional data which is not known to the JPEG library or the standard error handler. The most convenient way to do this is to embed either the JPEG object or the jpeg_error_mgr struct in a larger structure that contains additional fields; then casting the passed pointer provides access to the additional fields. Again, see example.c for one way to do it.

The individual methods that you might wish to override are:

error_exit (j_common_ptr cinfo)

Receives control for a fatal error. Information sufficient to generate the error message has been stored in cinfo->err; call output_message to display it. Control must NOT return to the caller; generally this routine will exit() or longjmp() somewhere. Typically you would override this routine to get rid of the exit() default behavior. Note that if you continue processing, you should clean up the JPEG object with jpeg_abort() or jpeg_destroy().

output_message (j_common_ptr cinfo)

Actual output of any JPEG message. Override this to send messages somewhere other than stderr. Note that this method does not know how to generate a message, only where to send it.

format_message (j_common_ptr cinfo, char * buffer)

Constructs a readable error message string based on the error info stored in cinfo->err. This method is called by output_message. Few applications should need to override this method. One possible reason for doing so is to implement dynamic switching of error message language.

emit_message (j_common_ptr cinfo, int msg_level)

Decide whether or not to emit a warning or trace message; if so, calls output_message. The main reason for overriding this method would be to abort on warnings. msg_level is -1 for warnings, 0 and up for trace messages.

Only error_exit() and emit_message() are called from the rest of the JPEG library; the other two are internal to the error handler.

The actual message texts are stored in an array of strings which is pointed to by the field err->jpeg_message_table. The messages are numbered from 0 to err->last_jpeg_message, and it is these code numbers that are used in the JPEG library code. You could replace the message texts (for instance, with messages in French or German) by changing the message table pointer. See jerror.h for the default texts. CAUTION: this table will almost certainly change or grow from one library version to the next.

It may be useful for an application to add its own message texts that are handled by the same mechanism. The error handler supports a second "add-on" message table for this purpose. To define an addon table, set the pointer err->addon_message_table and the message numbers err->first_addon_message and err->last_addon_message. If you number the addon messages beginning at 1000 or so, you won't have to worry about conflicts with the library's built-in messages. See the sample applications cjpeg/djpeg for an example of using addon messages (the addon messages are defined in cderror.h).

Actual invocation of the error handler is done via macros defined in jerror.h:

```
	ERREXITn(...)	for fatal errors
	WARNMSn(...)	for corrupt-data warnings
	TRACEMSn(...)	for trace and informational messages.
```

These macros store the message code and any additional parameters into the error handler struct, then invoke the error_exit() or emit_message() method. The variants of each macro are for varying numbers of additional parameters. The additional parameters are inserted into the generated message using standard printf() format codes.

See jerror.h and jerror.c for further details.

Compressed data handling (source and destination managers)

The JPEG compression library sends its compressed data to a "destination manager" module. The default destination manager just writes the data to a stdio stream, but you can provide your own manager to do something else. Similarly, the decompression library calls a "source manager" to obtain the compressed data; you can provide your own source manager if you want the data to come from somewhere other than a stdio stream.

In both cases, compressed data is processed a bufferload at a time: the destination or source manager provides a work buffer, and the library invokes the manager only when the buffer is filled or emptied. (You could define a one-character buffer to force the manager to be invoked for each byte, but that would be rather inefficient.) The buffer's size and location are controlled by the manager, not by the library. For example, if you desired to decompress a JPEG datastream that was all in memory, you could just make the buffer pointer and length point to the original data in memory. Then the buffer-reload procedure would be invoked only if the decompressor ran off the end of the datastream, which would indicate an erroneous datastream.

The work buffer is defined as an array of datatype JOCTET, which is generally "char" or "unsigned char". On a machine where char is not exactly 8 bits wide, you must define JOCTET as a wider data type and then modify the data source and destination modules to transcribe the work arrays into 8-bit units on external storage.

A data destination manager struct contains a pointer and count defining the next byte to write in the work buffer and the remaining free space:

```
	JOCTET * next_output_byte;	/* => next byte to write in buffer */
	size_t free_in_buffer;		/* # of byte spaces remaining in buffer */
```

The library increments the pointer and decrements the count until the buffer is filled. The manager's empty_output_buffer method must reset the pointer and count. The manager is expected to remember the buffer's starting address and total size in private fields not visible to the library.

A data destination manager provides three methods:

init_destination (j_compress_ptr cinfo)
	Initialize destination. This is called by jpeg_start_compress() before any data is actually written. It must initialize next_output_byte and free_in_buffer. free_in_buffer must be initialized to a positive value.

empty_output_buffer (j_compress_ptr cinfo)
	This is called whenever the buffer has filled (free_in_buffer reaches zero). In typical applications, it should write out the *entire* buffer (use the saved start address and buffer length; ignore the current state of next_output_byte and free_in_buffer). Then reset the pointer & count to the start of the buffer, and return TRUE indicating that the buffer has been dumped. free_in_buffer must be set to a positive value when TRUE is returned. A FALSE return should only be used when I/O suspension is desired (this operating mode is discussed in the next section).

term_destination (j_compress_ptr cinfo)
	Terminate destination --- called by jpeg_finish_compress() after all data has been written. In most applications, this must flush any data remaining in the buffer. Use either next_output_byte or free_in_buffer to determine how much data is in the buffer.

term_destination() is NOT called by jpeg_abort() or jpeg_destroy(). If you want the destination manager to be cleaned up during an abort, you must do it yourself.

You will also need code to create a jpeg_destination_mgr struct, fill in its method pointers, and insert a pointer to the struct into the "dest" field of the JPEG compression object. This can be done in-line in your setup code if you like, but it's probably cleaner to provide a separate routine similar to the jpeg_stdio_dest() routine of the supplied destination manager.

Decompression source managers follow a parallel design, but with some additional frammishes. The source manager struct contains a pointer and count defining the next byte to read from the work buffer and the number of bytes remaining:

```
	const JOCTET * next_input_byte;	/* => next byte to read from buffer */
	size_t bytes_in_buffer;		/* # of bytes remaining in buffer */
```

The library increments the pointer and decrements the count until the buffer is emptied. The manager's fill_input_buffer method must reset the pointer and count. In most applications, the manager must remember the buffer's starting address and total size in private fields not visible to the library.

A data source manager provides five methods:

init_source (j_decompress_ptr cinfo)
      Initialize source. This is called by jpeg_read_header() before any
      data is actually read. Unlike init_destination(), it may leave
      bytes_in_buffer set to 0 (in which case a fill_input_buffer() call
      will occur immediately).

fill_input_buffer (j_decompress_ptr cinfo)
      This is called whenever bytes_in_buffer has reached zero and more
      data is wanted. In typical applications, it should read fresh data
      into the buffer (ignoring the current state of next_input_byte and
      bytes_in_buffer), reset the pointer & count to the start of the
      buffer, and return TRUE indicating that the buffer has been reloaded.
      It is not necessary to fill the buffer entirely, only to obtain at
      least one more byte. bytes_in_buffer MUST be set to a positive value
      if TRUE is returned. A FALSE return should only be used when I/O
      suspension is desired (this mode is discussed in the next section).

skip_input_data (j_decompress_ptr cinfo, long num_bytes)
      Skip num_bytes worth of data. The buffer pointer and count should
      be advanced over num_bytes input bytes, refilling the buffer as
      needed. This is used to skip over a potentially large amount of
      uninteresting data (such as an APPn marker). In some applications
      it may be possible to optimize away the reading of the skipped data,
      but it's not clear that being smart is worth much trouble; large
      skips are uncommon. bytes_in_buffer may be zero on return.
      A zero or negative skip count should be treated as a no-op.

resync_to_restart (j_decompress_ptr cinfo, int desired)
      This routine is called only when the decompressor has failed to find
      a restart (RSTn) marker where one is expected. Its mission is to
      find a suitable point for resuming decompression. For most
      applications, we recommend that you just use the default resync
      procedure, jpeg_resync_to_restart(). However, if you are able to back
      up in the input data stream, or if you have a-priori knowledge about
      the likely location of restart markers, you may be able to do better.
      Read the read_restart_marker() and jpeg_resync_to_restart() routines
      in jdmarker.c if you think you'd like to implement your own resync
      procedure.

term_source (j_decompress_ptr cinfo)
      Terminate source --- called by jpeg_finish_decompress() after all
      data has been read. Often a no-op.

For both fill_input_buffer() and skip_input_data(), there is no such thing
as an EOF return. If the end of the file has been reached, the routine has
a choice of exiting via ERREXIT() or inserting fake data into the buffer.
In most cases, generating a warning message and inserting a fake EOI marker
is the best course of action --- this will allow the decompressor to output
however much of the image is there. In pathological cases, the decompressor
may swallow the EOI and again demand data ... just keep feeding it fake EOIs.
jdatasrc.c illustrates the recommended error recovery behavior.

term_source() is NOT called by jpeg_abort() or jpeg_destroy(). If you want
the source manager to be cleaned up during an abort, you must do it yourself.

You will also need code to create a jpeg_source_mgr struct, fill in its method pointers, and insert a pointer to the struct into the "src" field of the JPEG decompression object. This can be done in-line in your setup code if you like, but it's probably cleaner to provide a separate routine similar to the jpeg_stdio_src() routine of the supplied source manager.

For more information, consult the stdio source and destination managers in jdatasrc.c and jdatadst.c.

I/O suspension

Some applications need to use the JPEG library as an incremental memory-to-memory filter: when the compressed data buffer is filled or emptied, they want control to return to the outer loop, rather than expecting that the buffer can be emptied or reloaded within the data source/destination manager subroutine. The library supports this need by providing an "I/O suspension" mode, which we describe in this section.

The I/O suspension mode is not a panacea: nothing is guaranteed about the maximum amount of time spent in any one call to the library, so it will not eliminate response-time problems in single-threaded applications. If you need guaranteed response time, we suggest you "bite the bullet" and implement a real multi-tasking capability.

To use I/O suspension, cooperation is needed between the calling application and the data source or destination manager; you will always need a custom source/destination manager. (Please read the previous section if you haven't already.) The basic idea is that the empty_output_buffer() or fill_input_buffer() routine is a no-op, merely returning FALSE to indicate that it has done nothing. Upon seeing this, the JPEG library suspends operation and returns to its caller. The surrounding application is responsible for emptying or refilling the work buffer before calling the JPEG library again.

Compression suspension:

For compression suspension, use an empty_output_buffer() routine that returns FALSE; typically it will not do anything else. This will cause the compressor to return to the caller of jpeg_write_scanlines(), with the return value indicating that not all the supplied scanlines have been accepted. The application must make more room in the output buffer, adjust the output buffer pointer/count appropriately, and then call jpeg_write_scanlines() again, pointing to the first unconsumed scanline.

When forced to suspend, the compressor will backtrack to a convenient stopping point (usually the start of the current MCU); it will regenerate some output data when restarted. Therefore, although empty_output_buffer() is only called when the buffer is filled, you should NOT write out the entire buffer after a suspension. Write only the data up to the current position of next_output_byte/free_in_buffer. The data beyond that point will be regenerated after resumption.

Because of the backtracking behavior, a good-size output buffer is essential for efficiency; you don't want the compressor to suspend often. (In fact, an overly small buffer could lead to infinite looping, if a single MCU required more data than would fit in the buffer.) We recommend a buffer of at least several Kbytes. You may want to insert explicit code to ensure that you don't call jpeg_write_scanlines() unless there is a reasonable amount of space in the output buffer; in other words, flush the buffer before trying to compress more data.

The compressor does not allow suspension while it is trying to write JPEG markers at the beginning and end of the file. This means that:

* At the beginning of a compression operation, there must be enough free space in the output buffer to hold the header markers (typically 600 or so bytes). The recommended buffer size is bigger than this anyway, so this is not a problem as long as you start with an empty buffer. However, this restriction might catch you if you insert large special markers, such as a JFIF thumbnail image, without flushing the buffer afterwards.
* When you call jpeg_finish_compress(), there must be enough space in the output buffer to emit any buffered data and the final EOI marker. In the current implementation, half a dozen bytes should suffice for this, but for safety's sake we recommend ensuring that at least 100 bytes are free before calling jpeg_finish_compress().

A more significant restriction is that jpeg_finish_compress() cannot suspend. This means you cannot use suspension with multi-pass operating modes, namely Huffman code optimization and multiple-scan output. Those modes write the whole file during jpeg_finish_compress(), which will certainly result in buffer overrun. (Note that this restriction applies only to compression, not decompression. The decompressor supports input suspension in all of its operating modes.)

Decompression suspension:

For decompression suspension, use a fill_input_buffer() routine that simply returns FALSE (except perhaps during error recovery, as discussed below). This will cause the decompressor to return to its caller with an indication that suspension has occurred. This can happen at four places:

* jpeg_read_header(): will return JPEG_SUSPENDED.
* jpeg_start_decompress(): will return FALSE, rather than its usual TRUE.
* jpeg_read_scanlines(): will return the number of scanlines already completed (possibly 0).
* jpeg_finish_decompress(): will return FALSE, rather than its usual TRUE.

The surrounding application must recognize these cases, load more data into the input buffer, and repeat the call. In the case of jpeg_read_scanlines(), increment the passed pointers past any scanlines successfully read.

Just as with compression, the decompressor will typically backtrack to a convenient restart point before suspending. When fill_input_buffer() is called, next_input_byte/bytes_in_buffer point to the current restart point, which is where the decompressor will backtrack to if FALSE is returned. The data beyond that position must NOT be discarded if you suspend; it needs to be re-read upon resumption. In most implementations, you'll need to shift this data down to the start of your work buffer and then load more data after it. Again, this behavior means that a several-Kbyte work buffer is essential for decent performance; furthermore, you should load a reasonable amount of new data before resuming decompression. (If you loaded, say, only one new byte each time around, you could waste a LOT of cycles.)

The skip_input_data() source manager routine requires special care in a suspension scenario. This routine is NOT granted the ability to suspend the decompressor; it can decrement bytes_in_buffer to zero, but no more. If the requested skip distance exceeds the amount of data currently in the input buffer, then skip_input_data() must set bytes_in_buffer to zero and record the additional skip distance somewhere else. The decompressor will immediately call fill_input_buffer(), which should return FALSE, which will cause a suspension return. The surrounding application must then arrange to discard the recorded number of bytes before it resumes loading the input buffer. (Yes, this design is rather baroque, but it avoids complexity in the far more common case where a non-suspending source manager is used.)

If the input data has been exhausted, we recommend that you emit a warning and insert dummy EOI markers just as a non-suspending data source manager would do. This can be handled either in the surrounding application logic or within fill_input_buffer(); the latter is probably more efficient. If fill_input_buffer() knows that no more data is available, it can set the pointer/count to point to a dummy EOI marker and then return TRUE just as though it had read more data in a non-suspending situation.

The decompressor does not attempt to suspend within any JPEG marker; it will backtrack to the start of the marker. Hence the input buffer must be large enough to hold the longest marker in the file. We recommend at least a 2K buffer. The buffer would need to be 64K to allow for arbitrary COM or APPn markers, but the decompressor does not actually try to read these; it just skips them by calling skip_input_data(). If you provide a special marker handling routine that does look at such markers, coping with buffer overflow is your problem. Ordinary JPEG markers should normally not exceed a few hundred bytes each (DHT tables are typically the longest). For robustness against damaged marker length counts, you may wish to insert a test in your application for the case that the input buffer is completely full and yet the decoder has suspended without consuming any data --- otherwise, if this situation did occur, it would lead to an endless loop.

Multiple-buffer management:

In some applications it is desirable to store the compressed data in a linked list of buffer areas, so as to avoid data copying. This can be handled by having empty_output_buffer() or fill_input_buffer() set the pointer and count to reference the next available buffer; FALSE is returned only if no more buffers are available. Although seemingly straightforward, there is a pitfall in this approach: the backtrack that occurs when FALSE is returned could back up into an earlier buffer. For example, when fill_input_buffer() is called, the current pointer & count indicate the backtrack restart point. Since fill_input_buffer() will set the pointer and count to refer to a new buffer, the restart position must be saved somewhere else. Suppose a second call to fill_input_buffer() occurs in the same library call, and no additional input data is available, so fill_input_buffer must return FALSE. If the JPEG library has not moved the pointer/count forward in the current buffer, then *the correct restart point is the saved position in the prior buffer*. Prior buffers may be discarded only after the library establishes a restart point within a later buffer. Similar remarks apply for output into a chain of buffers.

The library will never attempt to backtrack over a skip_input_data() call, so any skipped data can be permanently discarded. You still have to deal with the case of skipping not-yet-received data, however.

It's much simpler to use only a single buffer; when fill_input_buffer() is called, move any unconsumed data (beyond the current pointer/count) down to the beginning of this buffer and then load new data into the remaining buffer space. This approach requires a little more data copying but is far easier to get right.

Progressive JPEG support

Progressive JPEG rearranges the stored data into a series of scans of increasing quality. In situations where a JPEG file is transmitted across a slow communications link, a decoder can generate a low-quality image very quickly from the first scan, then gradually improve the displayed quality as more scans are received. The final image after all scans are complete is identical to that of a regular (sequential) JPEG file of the same quality setting. Progressive JPEG files are often slightly smaller than equivalent sequential JPEG files, but the possibility of incremental display is the main reason for using progressive JPEG.

The IJG encoder library generates progressive JPEG files when given a suitable "scan script" defining how to divide the data into scans. Creation of progressive JPEG files is otherwise transparent to the encoder. Progressive JPEG files can also be read transparently by the decoder library. If the decoding application simply uses the library as defined above, it will receive a final decoded image without any indication that the file was progressive. Of course, this approach does not allow incremental display. To perform incremental display, an application needs to use the decoder library's "buffered-image" mode, in which it receives a decoded image multiple times.

Each displayed scan requires about as much work to decode as a full JPEG image of the same size, so the decoder must be fairly fast in relation to the data transmission rate in order to make incremental display useful. However, it is possible to skip displaying the image and simply add the incoming bits to the decoder's coefficient buffer. This is fast because only Huffman decoding need be done, not IDCT, upsampling, colorspace conversion, etc. The IJG decoder library allows the application to switch dynamically between displaying the image and simply absorbing the incoming bits. A properly coded application can automatically adapt the number of display passes to suit the time available as the image is received. Also, a final higher-quality display cycle can be performed from the buffered data after the end of the file is reached.

Progressive compression:

To create a progressive JPEG file (or a multiple-scan sequential JPEG file), set the scan_info cinfo field to point to an array of scan descriptors, and perform compression as usual. Instead of constructing your own scan list, you can call the jpeg_simple_progression() helper routine to create a recommended progression sequence; this method should be used by all applications that don't want to get involved in the nitty-gritty of progressive scan sequence design. (If you want to provide user control of scan sequences, you may wish to borrow the scan script reading code found in rdswitch.c, so that you can read scan script files just like cjpeg's.) When scan_info is not NULL, the compression library will store DCT'd data into a buffer array as jpeg_write_scanlines() is called, and will emit all the requested scans during jpeg_finish_compress(). This implies that multiple-scan output cannot be created with a suspending data destination manager, since jpeg_finish_compress() does not support suspension. We should also note that the compressor currently forces Huffman optimization mode when creating a progressive JPEG file, because the default Huffman tables are unsuitable for progressive files.

Progressive decompression:

When buffered-image mode is not used, the decoder library will read all of a multi-scan file during jpeg_start_decompress(), so that it can provide a final decoded image. (Here "multi-scan" means either progressive or multi-scan sequential.) This makes multi-scan files transparent to the decoding application. However, existing applications that used suspending input with version 5 of the IJG library will need to be modified to check for a suspension return from jpeg_start_decompress().

To perform incremental display, an application must use the library's buffered-image mode. This is described in the next section.

Buffered-image mode

In buffered-image mode, the library stores the partially decoded image in a coefficient buffer, from which it can be read out as many times as desired. This mode is typically used for incremental display of progressive JPEG files, but it can be used with any JPEG file. Each scan of a progressive JPEG file adds more data (more detail) to the buffered image. The application can display in lockstep with the source file (one display pass per input scan), or it can allow input processing to outrun display processing. By making input and display processing run independently, it is possible for the application to adapt progressive display to a wide range of data transmission rates.

The basic control flow for buffered-image decoding is

```
jpeg_create_decompress()
set data source
jpeg_read_header()
set overall decompression parameters
cinfo.buffered_image = TRUE;  /* select buffered-image mode */
jpeg_start_decompress()
for (each output pass) {
    adjust output decompression parameters if required
    jpeg_start_output()       /* start a new output pass */
    for (all scanlines in image) {
        jpeg_read_scanlines()
        display scanlines
    }
    jpeg_finish_output()      /* terminate output pass */
}
jpeg_finish_decompress()
jpeg_destroy_decompress()
```

This differs from ordinary unbuffered decoding in that there is an additional level of looping. The application can choose how many output passes to make and how to display each pass.

The simplest approach to displaying progressive images is to do one display pass for each scan appearing in the input file. In this case the outer loop condition is typically

```
      while (! jpeg_input_complete(&cinfo))
```

and the start-output call should read

```
      jpeg_start_output(&cinfo, cinfo.input_scan_number);
```

The second parameter to jpeg_start_output() indicates which scan of the input file is to be displayed; the scans are numbered starting at 1 for this purpose. (You can use a loop counter starting at 1 if you like, but using the library's input scan counter is easier.) The library automatically reads data as necessary to complete each requested scan, and jpeg_finish_output() advances to the next scan or end-of-image marker (hence input_scan_number will be incremented by the time control arrives back at jpeg_start_output()). With this technique, data is read from the input file only as needed, and input and output processing run in lockstep.

After reading the final scan and reaching the end of the input file, the buffered image remains available; it can be read additional times by repeating the jpeg_start_output()/jpeg_read_scanlines()/jpeg_finish_output() sequence. For example, a useful technique is to use fast one-pass color quantization for display passes made while the image is arriving, followed by a final display pass using two-pass quantization for highest quality. This is done by changing the library parameters before the final output pass. Changing parameters between passes is discussed in detail below.

In general the last scan of a progressive file cannot be recognized as such until after it is read, so a post-input display pass is the best approach if you want special processing in the final pass.

When done with the image, be sure to call jpeg_finish_decompress() to release the buffered image (or just use jpeg_destroy_decompress()).

If input data arrives faster than it can be displayed, the application can cause the library to decode input data in advance of what's needed to produce output. This is done by calling the routine jpeg_consume_input(). The return value is one of the following:

```
      JPEG_REACHED_SOS:    reached an SOS marker (the start of a new scan)
      JPEG_REACHED_EOI:    reached the EOI marker (end of image)
      JPEG_ROW_COMPLETED:  completed reading one MCU row of compressed data
      JPEG_SCAN_COMPLETED: completed reading last MCU row of current scan
      JPEG_SUSPENDED:      suspended before completing any of the above
```

(JPEG_SUSPENDED can occur only if a suspending data source is used.) This routine can be called at any time after initializing the JPEG object. It reads some additional data and returns when one of the indicated significant events occurs. (If called after the EOI marker is reached, it will immediately return JPEG_REACHED_EOI without attempting to read more data.)

The library's output processing will automatically call jpeg_consume_input() whenever the output processing overtakes the input; thus, simple lockstep display requires no direct calls to jpeg_consume_input(). But by adding calls to jpeg_consume_input(), you can absorb data in advance of what is being displayed. This has two benefits:

* You can limit buildup of unprocessed data in your input buffer.
* You can eliminate extra display passes by paying attention to the state of the library's input processing.

The first of these benefits only requires interspersing calls to jpeg_consume_input() with your display operations and any other processing you may be doing. To avoid wasting cycles due to backtracking, it's best to call jpeg_consume_input() only after a hundred or so new bytes have arrived. This is discussed further under "I/O suspension", above. (Note: the JPEG library currently is not thread-safe. You must not call jpeg_consume_input() from one thread of control if a different library routine is working on the same JPEG object in another thread.)

When input arrives fast enough that more than one new scan is available before you start a new output pass, you may as well skip the output pass corresponding to the completed scan. This occurs for free if you pass cinfo.input_scan_number as the target scan number to jpeg_start_output(). The input_scan_number field is simply the index of the scan currently being consumed by the input processor. You can ensure that this is up-to-date by emptying the input buffer just before calling jpeg_start_output(): call jpeg_consume_input() repeatedly until it returns JPEG_SUSPENDED or JPEG_REACHED_EOI.

The target scan number passed to jpeg_start_output() is saved in the cinfo.output_scan_number field. The library's output processing calls jpeg_consume_input() whenever the current input scan number and row within that scan is less than or equal to the current output scan number and row. Thus, input processing can "get ahead" of the output processing but is not allowed to "fall behind". You can achieve several different effects by manipulating this interlock rule. For example, if you pass a target scan number greater than the current input scan number, the output processor will wait until that scan starts to arrive before producing any output. (To avoid an infinite loop, the target scan number is automatically reset to the last scan number when the end of image is reached. Thus, if you specify a large target scan number, the library will just absorb the entire input file and then perform an output pass. This is effectively the same as what jpeg_start_decompress() does when you don't select buffered-image mode.) When you pass a target scan number equal to the current input scan number, the image is displayed no faster than the current input scan arrives. The final possibility is to pass a target scan number less than the current input scan number; this disables the input/output interlock and causes the output processor to simply display whatever it finds in the image buffer, without waiting for input. (However, the library will not accept a target scan number less than one, so you can't avoid waiting for the first scan.)

When data is arriving faster than the output display processing can advance through the image, jpeg_consume_input() will store data into the buffered image beyond the point at which the output processing is reading data out again. If the input arrives fast enough, it may "wrap around" the buffer to the point where the input is more than one whole scan ahead of the output. If the output processing simply proceeds through its display pass without paying attention to the input, the effect seen on-screen is that the lower part of the image is one or more scans better in quality than the upper part. Then, when the next output scan is started, you have a choice of what target scan number to use. The recommended choice is to use the current input scan number at that time, which implies that you've skipped the output scans corresponding to the input scans that were completed while you processed the previous output scan. In this way, the decoder automatically adapts its speed to the arriving data, by skipping output scans as necessary to keep up with the arriving data.

When using this strategy, you'll want to be sure that you perform a final output pass after receiving all the data; otherwise your last display may not be full quality across the whole screen. So the right outer loop logic is something like this:

```
      do {
          absorb any waiting input by calling jpeg_consume_input()
          final_pass = jpeg_input_complete(&cinfo);
          adjust output decompression parameters if required
          jpeg_start_output(&cinfo, cinfo.input_scan_number);
          ...
          jpeg_finish_output()
      } while (! final_pass);
```

rather than quitting as soon as jpeg_input_complete() returns TRUE. This arrangement makes it simple to use higher-quality decoding parameters for the final pass. But if you don't want to use special parameters for the final pass, the right loop logic is like this:

```
      for (;;) {
          absorb any waiting input by calling jpeg_consume_input()
          jpeg_start_output(&cinfo, cinfo.input_scan_number);
          ...
          jpeg_finish_output()
          if (jpeg_input_complete(&cinfo) &&
              cinfo.input_scan_number == cinfo.output_scan_number)
            break;
      }
```

In this case you don't need to know in advance whether an output pass is to be the last one, so it's not necessary to have reached EOF before starting the final output pass; rather, what you want to test is whether the output pass was performed in sync with the final input scan. This form of the loop will avoid an extra output pass whenever the decoder is able (or nearly able) to keep up with the incoming data.

When the data transmission speed is high, you might begin a display pass, then find that much or all of the file has arrived before you can complete the pass. (You can detect this by noting the JPEG_REACHED_EOI return code from jpeg_consume_input(), or equivalently by testing jpeg_input_complete().) In this situation you may wish to abort the current display pass and start a new one using the newly arrived information. To do so, just call jpeg_finish_output() and then start a new pass with jpeg_start_output().

A variant strategy is to abort and restart display if more than one complete scan arrives during an output pass; this can be detected by noting JPEG_REACHED_SOS returns and/or examining cinfo.input_scan_number. This idea should be employed with caution, however, since the display process might never get to the bottom of the image before being aborted, resulting in the lower part of the screen being several passes worse than the upper. In most cases it's probably best to abort an output pass only if the whole file has arrived and you want to begin the final output pass immediately.

When receiving data across a communication link, we recommend always using the current input scan number for the output target scan number; if a higher-quality final pass is to be done, it should be started (aborting any incomplete output pass) as soon as the end of file is received. However, many other strategies are possible. For example, the application can examine the parameters of the current input scan and decide whether to display it or not. If the scan contains only chroma data, one might choose not to use it as the target scan, expecting that the scan will be small and will arrive quickly. To skip to the next scan, call jpeg_consume_input() until it returns JPEG_REACHED_SOS or JPEG_REACHED_EOI. Or just use the next higher number as the target scan for jpeg_start_output(); but that method doesn't let you inspect the next scan's parameters before deciding to display it.

In buffered-image mode, jpeg_start_decompress() never performs input and thus never suspends. An application that uses input suspension with buffered-image mode must be prepared for suspension returns from these routines:

* jpeg_start_output() performs input only if you request 2-pass quantization and the target scan isn't fully read yet. (This is discussed below.)
* jpeg_read_scanlines(), as always, returns the number of scanlines that it was able to produce before suspending.
* jpeg_finish_output() will read any markers following the target scan, up to the end of the file or the SOS marker that begins another scan. (But it reads no input if jpeg_consume_input() has already reached the end of the file or a SOS marker beyond the target output scan.)
* jpeg_finish_decompress() will read until the end of file, and thus can suspend if the end hasn't already been reached (as can be tested by calling jpeg_input_complete()).

jpeg_start_output(), jpeg_finish_output(), and jpeg_finish_decompress() all return TRUE if they completed their tasks, FALSE if they had to suspend. In the event of a FALSE return, the application must load more input data and repeat the call. Applications that use non-suspending data sources need not check the return values of these three routines.

It is possible to change decoding parameters between output passes in the buffered-image mode. The decoder library currently supports only very limited changes of parameters. ONLY THE FOLLOWING parameter changes are allowed after jpeg_start_decompress() is called:

* dct_method can be changed before each call to jpeg_start_output(). For example, one could use a fast DCT method for early scans, changing to a higher quality method for the final scan.
* dither_mode can be changed before each call to jpeg_start_output(); of course this has no impact if not using color quantization. Typically one would use ordered dither for initial passes, then switch to Floyd-Steinberg dither for the final pass. Caution: changing dither mode can cause more memory to be allocated by the library. Although the amount of memory involved is not large (a scanline or so), it may cause the initial max_memory_to_use specification to be exceeded, which in the worst case would result in an out-of-memory failure.
* do_block_smoothing can be changed before each call to jpeg_start_output(). This setting is relevant only when decoding a progressive JPEG image. During the first DC-only scan, block smoothing provides a very "fuzzy" look instead of the very "blocky" look seen without it; which is better seems a matter of personal taste. But block smoothing is nearly always a win during later stages, especially when decoding a successive-approximation image: smoothing helps to hide the slight blockiness that otherwise shows up on smooth gradients until the lowest coefficient bits are sent.
* Color quantization mode can be changed under the rules described below. You *cannot* change between full-color and quantized output (because that would alter the required I/O buffer sizes), but you can change which quantization method is used.

When generating color-quantized output, changing quantization method is a

very useful way of switching between high-speed and high-quality display.
The library allows you to change among its three quantization methods:
1. Single-pass quantization to a fixed color cube.
   Selected by cinfo.two_pass_quantize = FALSE and cinfo.colormap = NULL.
2. Single-pass quantization to an application-supplied colormap.
   Selected by setting cinfo.colormap to point to the colormap (the value of two_pass_quantize is ignored); also set cinfo.actual_number_of_colors.
3. Two-pass quantization to a colormap chosen specifically for the image.
   Selected by cinfo.two_pass_quantize = TRUE and cinfo.colormap = NULL.
   (This is the default setting selected by jpeg_read_header, but it is probably NOT what you want for the first pass of progressive display!)

These methods offer successively better quality and lesser speed. However, only the first method is available for quantizing in non-RGB color spaces.

IMPORTANT: because the different quantizer methods have very different working-storage requirements, the library requires you to indicate which one(s) you intend to use before you call jpeg_start_decompress(). (If we did not require this, the max_memory_to_use setting would be a complete fiction.) You do this by setting one or more of these three cinfo fields to TRUE:

```
	enable_1pass_quant		Fixed color cube colormap
	enable_external_quant		Externally-supplied colormap
	enable_2pass_quant		Two-pass custom colormap
```

All three are initialized FALSE by jpeg_read_header(). But jpeg_start_decompress() automatically sets TRUE the one selected by the current two_pass_quantize and colormap settings, so you only need to set the enable flags for any other quantization methods you plan to change to later.

After setting the enable flags correctly at jpeg_start_decompress() time, you can change to any enabled quantization method by setting two_pass_quantize and colormap properly just before calling jpeg_start_output(). The following special rules apply:
1. You must explicitly set cinfo.colormap to NULL when switching to 1-pass or 2-pass mode from a different mode, or when you want the 2-pass quantizer to be re-run to generate a new colormap.
2. To switch to an external colormap, or to change to a different external colormap than was used on the prior pass, you must call jpeg_new_colormap() after setting cinfo.colormap.

NOTE: if you want to use the same colormap as was used in the prior pass, you should not do either of these things. This will save some nontrivial switchover costs.
(These requirements exist because cinfo.colormap will always be non-NULL after completing a prior output pass, since both the 1-pass and 2-pass quantizers set it to point to their output colormaps. Thus you have to do one of these two things to notify the library that something has changed. Yup, it's a bit klugy, but it's necessary to do it this way for backwards compatibility.)

Note that in buffered-image mode, the library generates any requested colormap during jpeg_start_output(), not during jpeg_start_decompress().

When using two-pass quantization, jpeg_start_output() makes a pass over the buffered image to determine the optimum color map; it therefore may take a significant amount of time, whereas ordinarily it does little work. The progress monitor hook is called during this pass, if defined. It is also important to realize that if the specified target scan number is greater than or equal to the current input scan number, jpeg_start_output() will attempt to consume input as it makes this pass. If you use a suspending data source, you need to check for a FALSE return from jpeg_start_output() under these conditions. The combination of 2-pass quantization and a not-yet-fully-read target scan is the only case in which jpeg_start_output() will consume input.

Application authors who support buffered-image mode may be tempted to use it for all JPEG images, even single-scan ones. This will work, but it is inefficient: there is no need to create an image-sized coefficient buffer for single-scan images. Requesting buffered-image mode for such an image wastes memory. Worse, it can cost time on large images, since the buffered data has to be swapped out or written to a temporary file. If you are concerned about maximum performance on baseline JPEG files, you should use buffered-image mode only when the incoming file actually has multiple scans. This can be tested by calling jpeg_has_multiple_scans(), which will return a correct result at any time after jpeg_read_header() completes.

It is also worth noting that when you use jpeg_consume_input() to let input processing get ahead of output processing, the resulting pattern of access to the coefficient buffer is quite nonsequential. It's best to use the memory manager jmemnobs.c if you can (ie, if you have enough real or virtual main memory). If not, at least make sure that max_memory_to_use is set as high as possible. If the JPEG memory manager has to use a temporary file, you will probably see a lot of disk traffic and poor performance. (This could be improved with additional work on the memory manager, but we haven't gotten around to it yet.)

In some applications it may be convenient to use jpeg_consume_input() for all input processing, including reading the initial markers; that is, you may wish to call jpeg_consume_input() instead of jpeg_read_header() during startup. This works, but note that you must check for JPEG_REACHED_SOS and JPEG_REACHED_EOI return codes as the equivalent of jpeg_read_header's codes. Once the first SOS marker has been reached, you must call jpeg_start_decompress() before jpeg_consume_input() will consume more input; it'll just keep returning JPEG_REACHED_SOS until you do. If you read a tables-only file this way, jpeg_consume_input() will return JPEG_REACHED_EOI without ever returning JPEG_REACHED_SOS; be sure to check for this case. If this happens, the decompressor will not read any more input until you call jpeg_abort() to reset it. It is OK to call jpeg_consume_input() even when not using buffered-image mode, but in that case it's basically a no-op after the initial markers have been read: it will just return JPEG_SUSPENDED.

Abbreviated datastreams and multiple images

A JPEG compression or decompression object can be reused to process multiple images. This saves a small amount of time per image by eliminating the "create" and "destroy" operations, but that isn't the real purpose of the feature. Rather, reuse of an object provides support for abbreviated JPEG datastreams. Object reuse can also simplify processing a series of images in a single input or output file. This section explains these features.

A JPEG file normally contains several hundred bytes worth of quantization and Huffman tables. In a situation where many images will be stored or transmitted with identical tables, this may represent an annoying overhead. The JPEG standard therefore permits tables to be omitted. The standard defines three classes of JPEG datastreams:

* "Interchange" datastreams contain an image and all tables needed to decode the image. These are the usual kind of JPEG file.
* "Abbreviated image" datastreams contain an image, but are missing some or all of the tables needed to decode that image.
* "Abbreviated table specification" (henceforth "tables-only") datastreams contain only table specifications.

To decode an abbreviated image, it is necessary to load the missing table(s) into the decoder beforehand. This can be accomplished by reading a separate tables-only file. A variant scheme uses a series of images in which the first image is an interchange (complete) datastream, while subsequent ones are abbreviated and rely on the tables loaded by the first image. It is assumed that once the decoder has read a table, it will remember that table until a new definition for the same table number is encountered.

It is the application designer's responsibility to figure out how to associate the correct tables with an abbreviated image. While abbreviated datastreams can be useful in a closed environment, their use is strongly discouraged in any situation where data exchange with other applications might be needed. Caveat designer.

The JPEG library provides support for reading and writing any combination of tables-only datastreams and abbreviated images. In both compression and decompression objects, a quantization or Huffman table will be retained for the lifetime of the object, unless it is overwritten by a new table definition.

To create abbreviated image datastreams, it is only necessary to tell the compressor not to emit some or all of the tables it is using. Each quantization and Huffman table struct contains a boolean field "sent_table", which normally is initialized to FALSE. For each table used by the image, the header-writing process emits the table and sets sent_table = TRUE unless it is already TRUE. (In normal usage, this prevents outputting the same table definition multiple times, as would otherwise occur because the chroma components typically share tables.) Thus, setting this field to TRUE before calling jpeg_start_compress() will prevent the table from being written at all.

If you want to create a "pure" abbreviated image file containing no tables, just call "jpeg_suppress_tables(&cinfo, TRUE)" after constructing all the tables. If you want to emit some but not all tables, you'll need to set the individual sent_table fields directly.

To create an abbreviated image, you must also call jpeg_start_compress() with a second parameter of FALSE, not TRUE. Otherwise jpeg_start_compress() will force all the sent_table fields to FALSE. (This is a safety feature to prevent abbreviated images from being created accidentally.)

To create a tables-only file, perform the same parameter setup that you normally would, but instead of calling jpeg_start_compress() and so on, call jpeg_write_tables(&cinfo). This will write an abbreviated datastream containing only SOI, DQT and/or DHT markers, and EOI. All the quantization and Huffman tables that are currently defined in the compression object will be emitted unless their sent_tables flag is already TRUE, and then all the sent_tables flags will be set TRUE.

A sure-fire way to create matching tables-only and abbreviated image files is to proceed as follows:

```
create JPEG compression object
set JPEG parameters
set destination to tables-only file
jpeg_write_tables(&cinfo);
set destination to image file
jpeg_start_compress(&cinfo, FALSE);
write data...
jpeg_finish_compress(&cinfo);
```

Since the JPEG parameters are not altered between writing the table file and the abbreviated image file, the same tables are sure to be used. Of course, you can repeat the jpeg_start_compress() ... jpeg_finish_compress() sequence many times to produce many abbreviated image files matching the table file.

You cannot suppress output of the computed Huffman tables when Huffman optimization is selected. (If you could, there'd be no way to decode the image...) Generally, you don't want to set optimize_coding = TRUE when you are trying to produce abbreviated files.

In some cases you might want to compress an image using tables which are not stored in the application, but are defined in an interchange or tables-only file readable by the application. This can be done by setting up a JPEG decompression object to read the specification file, then copying the tables into your compression object. See jpeg_copy_critical_parameters() for an example of copying quantization tables.

To read abbreviated image files, you simply need to load the proper tables into the decompression object before trying to read the abbreviated image. If the proper tables are stored in the application program, you can just allocate the table structs and fill in their contents directly. More commonly you'd want to read the tables from a tables-only file. The jpeg_read_header() call is sufficient to read a tables-only file. You must pass a second parameter of FALSE to indicate that you do not require an image to be present. Thus, the typical scenario is

```
create JPEG decompression object
set source to tables-only file
jpeg_read_header(&cinfo, FALSE);
set source to abbreviated image file
jpeg_read_header(&cinfo, TRUE);
set decompression parameters
jpeg_start_decompress(&cinfo);
read data...
jpeg_finish_decompress(&cinfo);
```

In some cases, you may want to read a file without knowing whether it contains an image or just tables. In that case, pass FALSE and check the return value from jpeg_read_header(): it will be JPEG_HEADER_OK if an image was found, JPEG_HEADER_TABLES_ONLY if only tables were found. (A third return value, JPEG_SUSPENDED, is possible when using a suspending data source manager.) Note that jpeg_read_header() will not complain if you read an abbreviated image for which you haven't loaded the missing tables; the missing-table check occurs later, in jpeg_start_decompress().

It is possible to read a series of images from a single source file by repeating the jpeg_read_header() ... jpeg_finish_decompress() sequence, without releasing/recreating the JPEG object or the data source module. (If you did reinitialize, any partial bufferload left in the data source buffer at the end of one image would be discarded, causing you to lose the start of the next image.) When you use this method, stored tables are automatically carried forward, so some of the images can be abbreviated images that depend on tables from earlier images.

If you intend to write a series of images into a single destination file, you might want to make a specialized data destination module that doesn't flush the output buffer at term_destination() time. This would speed things up by some trifling amount. Of course, you'd need to remember to flush the buffer after the last image. You can make the later images be abbreviated ones by passing FALSE to jpeg_start_compress().

Special markers

Some applications may need to insert or extract special data in the JPEG datastream. The JPEG standard provides marker types "COM" (comment) and "APP0" through "APP15" (application) to hold application-specific data. Unfortunately, the use of these markers is not specified by the standard. COM markers are fairly widely used to hold user-supplied text. The JFIF file format spec uses APP0 markers with specified initial strings to hold certain data. Adobe applications use APP14 markers beginning with the string "Adobe" for miscellaneous data. Other APPn markers are rarely seen, but might contain almost anything.

If you wish to store user-supplied text, we recommend you use COM markers and place readable 7-bit ASCII text in them. Newline conventions are not standardized --- expect to find LF (Unix style), CR/LF (DOS style), or CR (Mac style). A robust COM reader should be able to cope with random binary garbage, including nulls, since some applications generate COM markers containing non-ASCII junk. (But yours should not be one of them.)

For program-supplied data, use an APPn marker, and be sure to begin it with an identifying string so that you can tell whether the marker is actually yours. It's probably best to avoid using APP0 or APP14 for any private markers. (NOTE: the upcoming SPIFF standard will use APP8 markers; we recommend you not use APP8 markers for any private purposes, either.)

Keep in mind that at most 65533 bytes can be put into one marker, but you can have as many markers as you like.

By default, the IJG compression library will write a JFIF APP0 marker if the selected JPEG colorspace is grayscale or YCbCr, or an Adobe APP14 marker if the selected colorspace is RGB, CMYK, or YCCK. You can disable this, but we don't recommend it. The decompression library will recognize JFIF and Adobe markers and will set the JPEG colorspace properly when one is found.

You can write special markers immediately following the datastream header by calling jpeg_write_marker() after jpeg_start_compress() and before the first call to jpeg_write_scanlines(). When you do this, the markers appear after the SOI and the JFIF APP0 and Adobe APP14 markers (if written), but before all else. Specify the marker type parameter as "JPEG_COM" for COM or

"JPEG_APP0 + n" for APPn. (Actually, jpeg_write_marker will let you write any marker type, but we don't recommend writing any other kinds of marker.) For example, to write a user comment string pointed to by comment_text:

```
jpeg_write_marker(cinfo, JPEG_COM, comment_text, strlen(comment_text));
```

Or if you prefer to synthesize the marker byte sequence yourself, you can just cram it straight into the data destination module.

For decompression, you can supply your own routine to process COM or APPn markers by calling jpeg_set_marker_processor(). Usually you'd call this after creating a decompression object and before calling jpeg_read_header(), because the markers of interest will normally be scanned by jpeg_read_header. Once you've supplied a routine, it will be used for the life of that decompression object. A separate routine may be registered for COM and for each APPn marker code.

A marker processor routine must have the signature

```
boolean jpeg_marker_parser_method (j_decompress_ptr cinfo)
```

Although the marker code is not explicitly passed, the routine can find it in cinfo->unread_marker. At the time of call, the marker proper has been read from the data source module. The processor routine is responsible for reading the marker length word and the remaining parameter bytes, if any. Return TRUE to indicate success. (FALSE should be returned only if you are using a suspending data source and it tells you to suspend. See the standard marker processors in jdmarker.c for appropriate coding methods if you need to use a suspending data source.)

If you override the default APP0 or APP14 processors, it is up to you to recognize JFIF and Adobe markers if you want colorspace recognition to occur properly. We recommend copying and extending the default processors if you want to do that.

A simple example of an external COM processor can be found in djpeg.c.

Raw (downsampled) image data

Some applications need to supply already-downsampled image data to the JPEG compressor, or to receive raw downsampled data from the decompressor. The library supports this requirement by allowing the application to write or read raw data, bypassing the normal preprocessing or postprocessing steps. The interface is different from the standard one and is somewhat harder to use. If your interest is merely in bypassing color conversion, we recommend that you use the standard interface and simply set jpeg_color_space = in_color_space (or jpeg_color_space = out_color_space for decompression). The mechanism described in this section is necessary only to supply or receive downsampled image data, in which not all components have the same dimensions.

To compress raw data, you must supply the data in the colorspace to be used in the JPEG file (please read the earlier section on Special color spaces) and downsampled to the sampling factors specified in the JPEG parameters. You must supply the data in the format used internally by the JPEG library, namely a JSAMPIMAGE array. This is an array of pointers to two-dimensional arrays, each of type JSAMPARRAY. Each 2-D array holds the values for one color component. This structure is necessary since the components are of different sizes. If the image dimensions are not a multiple of the MCU size, you must also pad the data correctly (usually, this is done by replicating the last column and/or row). The data must be padded to a multiple of a DCT block in each component: that is, each downsampled row must contain a multiple of 8 valid samples, and there must be a multiple of 8 sample rows for each component. (For applications such as conversion of digital TV images, the standard image size is usually a multiple of the DCT block size, so that no padding need actually be done.)

The procedure for compression of raw data is basically the same as normal compression, except that you call jpeg_write_raw_data() in place of jpeg_write_scanlines(). Before calling jpeg_start_compress(), you must do the following:

* Set cinfo->raw_data_in to TRUE. (It is set FALSE by jpeg_set_defaults().) This notifies the library that you will be supplying raw data.
* Ensure jpeg_color_space is correct --- an explicit jpeg_set_colorspace() call is a good idea. Note that since color conversion is bypassed, in_color_space is ignored, except that jpeg_set_defaults() uses it to choose the default jpeg_color_space setting.
* Ensure the sampling factors, cinfo->comp_info[i].h_samp_factor and cinfo->comp_info[i].v_samp_factor, are correct. Since these indicate the dimensions of the data you are supplying, it's wise to set them explicitly, rather than assuming the library's defaults are what you want.

To pass raw data to the library, call jpeg_write_raw_data() in place of jpeg_write_scanlines(). The two routines work similarly except that jpeg_write_raw_data takes a JSAMPIMAGE data array rather than JSAMPARRAY. The scanlines count passed to and returned from jpeg_write_raw_data is measured in terms of the component with the largest v_samp_factor.

jpeg_write_raw_data() processes one MCU row per call, which is to say v_samp_factor*DCTSIZE sample rows of each component. The passed num_lines value must be at least max_v_samp_factor*DCTSIZE, and the return value will be exactly that amount (or possibly some multiple of that amount, in future library versions). This is true even on the last call at the bottom of the image; don't forget to pad your data as necessary.

The required dimensions of the supplied data can be computed for each component as

```
      cinfo->comp_info[i].width_in_blocks*DCTSIZE  samples per row
      cinfo->comp_info[i].height_in_blocks*DCTSIZE rows in image
```

after jpeg_start_compress() has initialized those fields. If the valid data is smaller than this, it must be padded appropriately. For some sampling factors and image sizes, additional dummy DCT blocks are inserted to make the image a multiple of the MCU dimensions. The library creates such dummy blocks itself; it does not read them from your supplied data. Therefore you need never pad by more than DCTSIZE samples. An example may help here. Assume 2h2v downsampling of YCbCr data, that is

```
      cinfo->comp_info[0].h_samp_factor = 2           for Y
      cinfo->comp_info[0].v_samp_factor = 2
      cinfo->comp_info[1].h_samp_factor = 1           for Cb
      cinfo->comp_info[1].v_samp_factor = 1
      cinfo->comp_info[2].h_samp_factor = 1           for Cr
      cinfo->comp_info[2].v_samp_factor = 1
```

and suppose that the nominal image dimensions (cinfo->image_width and cinfo->image_height) are 101x101 pixels. Then jpeg_start_compress() will compute downsampled_width = 101 and width_in_blocks = 13 for Y, downsampled_width = 51 and width_in_blocks = 7 for Cb and Cr (and the same for the height fields). You must pad the Y data to at least 13*8 = 104 columns and rows, the Cb/Cr data to at least 7*8 = 56 columns and rows. The MCU height is max_v_samp_factor = 2 DCT rows so you must pass at least 16 scanlines on each call to jpeg_write_raw_data(), which is to say 16 actual sample rows of Y and 8 each of Cb and Cr. A total of 7 MCU rows are needed, so you must pass a total of 7*16 = 112 "scanlines". The last DCT block row of Y data is dummy, so it doesn't matter what you pass for it in the data arrays, but the scanlines count must total up to 112 so that all of the Cb and Cr data gets passed.

Output suspension is supported with raw-data compression: if the data destination module suspends, jpeg_write_raw_data() will return 0. In this case the same data rows must be passed again on the next call.

Decompression with raw data output implies bypassing all postprocessing: you cannot ask for rescaling or color quantization, for instance. More seriously, you must deal with the color space and sampling factors present in the incoming file. If your application only handles, say, 2h1v YCbCr data, you must check for and fail on other color spaces or other sampling factors. The library will not convert to a different color space for you.

To obtain raw data output, set cinfo->raw_data_out = TRUE before jpeg_start_decompress() (it is set FALSE by jpeg_read_header()). Be sure to verify that the color space and sampling factors are ones you can handle. Then call jpeg_read_raw_data() in place of jpeg_read_scanlines(). The decompression process is otherwise the same as usual.

jpeg_read_raw_data() returns one MCU row per call, and thus you must pass a buffer of at least max_v_samp_factor*DCTSIZE scanlines (scanline counting is the same as for raw-data compression). The buffer you pass must be large enough to hold the actual data plus padding to DCT-block boundaries. As with compression, any entirely dummy DCT blocks are not processed so you need not allocate space for them, but the total scanline count includes them. The above example of computing buffer dimensions for raw-data compression is equally valid for decompression.

Input suspension is supported with raw-data decompression: if the data source module suspends, jpeg_read_raw_data() will return 0. You can also use buffered-image mode to read raw data in multiple passes.

Really raw data: DCT coefficients

It is possible to read or write the contents of a JPEG file as raw DCT coefficients. This facility is mainly intended for use in lossless transcoding between different JPEG file formats. Other possible applications include lossless cropping of a JPEG image, lossless reassembly of a multi-strip or multi-tile TIFF/JPEG file into a single JPEG datastream, etc.

To read the contents of a JPEG file as DCT coefficients, open the file and do jpeg_read_header() as usual. But instead of calling jpeg_start_decompress() and jpeg_read_scanlines(), call jpeg_read_coefficients(). This will read the entire image into a set of virtual coefficient-block arrays, one array per component. The return value is a pointer to an array of virtual-array descriptors. Each virtual array can be accessed directly using the JPEG memory manager's access_virt_barray method (see Memory management, below, and also read structure.doc's discussion of virtual array handling). Or, for simple transcoding to a different JPEG file format, the array list can just be handed directly to jpeg_write_coefficients().

When you are done using the virtual arrays, call jpeg_finish_decompress() to release the array storage and return the decompression object to an idle state; or just call jpeg_destroy() if you don't need to reuse the object.

If you use a suspending data source, jpeg_read_coefficients() will return NULL if it is forced to suspend; a non-NULL return value indicates successful completion. You need not test for a NULL return value when using a non-suspending data source.

Each block in the block arrays contains quantized coefficient values in normal array order (not JPEG zigzag order). The block arrays contain only DCT blocks containing real data; any entirely-dummy blocks added to fill out interleaved MCUs at the right or bottom edges of the image are discarded during reading and are not stored in the block arrays. (The size of each block array can be determined from the width_in_blocks and height_in_blocks fields of the component's comp_info entry.) This is also the data format expected by jpeg_write_coefficients().

To write the contents of a JPEG file as DCT coefficients, you must provide the DCT coefficients stored in virtual block arrays. You can either pass block arrays read from an input JPEG file by jpeg_read_coefficients(), or allocate virtual arrays from the JPEG compression object and fill them yourself. In either case, jpeg_write_coefficients() is substituted for jpeg_start_compress() and jpeg_write_scanlines(). Thus the sequence is

* Create compression object
* Set all compression parameters as necessary
* Request virtual arrays if needed
* jpeg_write_coefficients()
* jpeg_finish_compress()
* Destroy or re-use compression object

jpeg_write_coefficients() is passed a pointer to an array of virtual block array descriptors; the number of arrays is equal to cinfo.num_components.

The virtual arrays need only have been requested, not realized, before jpeg_write_coefficients() is called. A side-effect of jpeg_write_coefficients() is to realize any virtual arrays that have been requested from the compression object's memory manager. Thus, when obtaining the virtual arrays from the compression object, you should fill the arrays after calling jpeg_write_coefficients(). The data is actually written out when you call jpeg_finish_compress(); jpeg_write_coefficients() only writes the file header.

When writing raw DCT coefficients, it is crucial that the JPEG quantization tables and sampling factors match the way the data was encoded, or the resulting file will be invalid. For transcoding from an existing JPEG file, we recommend using jpeg_copy_critical_parameters(). This routine initializes all the compression parameters to default values (like jpeg_set_defaults()), then copies the critical information from a source decompression object. The decompression object should have just been used to read the entire JPEG input file --- that is, it should be awaiting jpeg_finish_decompress().

jpeg_write_coefficients() marks all tables stored in the compression object as needing to be written to the output file (thus, it acts like jpeg_start_compress(cinfo, TRUE)). This is for safety's sake, to avoid emitting abbreviated JPEG files by accident. If you really want to emit an abbreviated JPEG file, call jpeg_suppress_tables(), or set the tables' individual sent_table flags, between calling jpeg_write_coefficients() and jpeg_finish_compress().

Progress monitoring

Some applications may need to regain control from the JPEG library every so often. The typical use of this feature is to produce a percent-done bar or other progress display. (For a simple example, see cjpeg.c or djpeg.c.) Although you do get control back frequently during the data-transferring pass (the jpeg_read_scanlines or jpeg_write_scanlines loop), any additional passes will occur inside jpeg_finish_compress or jpeg_start_decompress; those routines may take a long time to execute, and you don't get control back until they are done.

You can define a progress-monitor routine which will be called periodically by the library. No guarantees are made about how often this call will occur, so we don't recommend you use it for mouse tracking or anything like that. At present, a call will occur once per MCU row, scanline, or sample row group, whichever unit is convenient for the current processing mode; so the wider the image, the longer the time between calls. During the data transferring pass, only one call occurs per call of jpeg_read_scanlines or jpeg_write_scanlines, so don't pass a large number of scanlines at once if you want fine resolution in the progress count. (If you really need to use the callback mechanism for time-critical tasks like mouse tracking, you could insert additional calls inside some of the library's inner loops.)

To establish a progress-monitor callback, create a struct jpeg_progress_mgr, fill in its progress_monitor field with a pointer to your callback routine, and set cinfo->progress to point to the struct. The callback will be called whenever cinfo->progress is non-NULL. (This pointer is set to NULL by jpeg_create_compress or jpeg_create_decompress; the library will not change it thereafter. So if you allocate dynamic storage for the progress struct, make sure it will live as long as the JPEG object does. Allocating from the JPEG memory manager with lifetime JPOOL_PERMANENT will work nicely.) You can use the same callback routine for both compression and decompression.

The jpeg_progress_mgr struct contains four fields which are set by the library:

```
	long pass_counter;	/* work units completed in this pass */
	long pass_limit;	/* total number of work units in this pass */
	int completed_passes;	/* passes completed so far */
	int total_passes;	/* total number of passes expected */
```

During any one pass, pass_counter increases from 0 up to (not including) pass_limit; the step size is usually but not necessarily 1. The pass_limit value may change from one pass to another. The expected total number of passes is in total_passes, and the number of passes already completed is in completed_passes. Thus the fraction of work completed may be estimated as

```
		completed_passes + (pass_counter/pass_limit)
		--------------------------------------------
				total_passes
```

ignoring the fact that the passes may not be equal amounts of work.

When decompressing, pass_limit can even change within a pass, because it depends on the number of scans in the JPEG file, which isn't always known in advance. The computed fraction-of-work-done may jump suddenly (if the library discovers it has overestimated the number of scans) or even decrease (in the opposite case). It is not wise to put great faith in the work estimate.

When using the decompressor's buffered-image mode, the progress monitor work estimate is likely to be completely unhelpful, because the library has no way to know how many output passes will be demanded of it. Currently, the library sets total_passes based on the assumption that there will be one more output pass if the input file end hasn't yet been read (jpeg_input_complete() isn't TRUE), but no more output passes if the file end has been reached when the output pass is started. This means that total_passes will rise as additional output passes are requested. If you have a way of determining the input file size, estimating progress based on the fraction of the file that's been read will probably be more useful than using the library's value.

Memory management

This section covers some key facts about the JPEG library's built-in memory manager. For more info, please read structure.doc's section about the memory manager, and consult the source code if necessary.

All memory and temporary file allocation within the library is done via the memory manager. If necessary, you can replace the "back end" of the memory manager to control allocation yourself (for example, if you don't want the library to use malloc() and free() for some reason).

Some data is allocated "permanently" and will not be freed until the JPEG object is destroyed. Most data is allocated "per image" and is freed by jpeg_finish_compress, jpeg_finish_decompress, or jpeg_abort. You can call the memory manager yourself to allocate structures that will automatically be freed at these times. Typical code for this is

```
  ptr = (*cinfo->mem->alloc_small) ((j_common_ptr) cinfo, JPOOL_IMAGE, size);
```

Use JPOOL_PERMANENT to get storage that lasts as long as the JPEG object. Use alloc_large instead of alloc_small for anything bigger than a few Kbytes. There are also alloc_sarray and alloc_barray routines that automatically build 2-D sample or block arrays.

The library's minimum space requirements to process an image depend on the image's width, but not on its height, because the library ordinarily works with "strip" buffers that are as wide as the image but just a few rows high. Some operating modes (eg, two-pass color quantization) require full-image buffers. Such buffers are treated as "virtual arrays": only the current strip need be in memory, and the rest can be swapped out to a temporary file.

If you use the simplest memory manager back end (jmemnobs.c), then no temporary files are used; virtual arrays are simply malloc()'d. Images bigger than memory can be processed only if your system supports virtual memory. The other memory manager back ends support temporary files of various flavors and thus work in machines without virtual memory. They may also be useful on Unix machines if you need to process images that exceed available swap space.

When using temporary files, the library will make the in-memory buffers for its virtual arrays just big enough to stay within a "maximum memory" setting. Your application can set this limit by setting cinfo->mem->max_memory_to_use after creating the JPEG object. (Of course, there is still a minimum size for the buffers, so the max-memory setting is effective only if it is bigger than the minimum space needed.) If you allocate any large structures yourself, you must allocate them before jpeg_start_compress() or jpeg_start_decompress() in order to have them counted against the max memory limit. Also keep in mind that space allocated with alloc_small() is ignored, on the assumption that it's too small to be worth worrying about; so a reasonable safety margin should be left when setting max_memory_to_use.

If you use the jmemname.c or jmemdos.c memory manager back end, it is important to clean up the JPEG object properly to ensure that the temporary files get deleted. (This is especially crucial with jmemdos.c, where the "temporary files" may be extended-memory segments; if they are not freed, DOS will require a reboot to recover the memory.) Thus, with these memory managers, it's a good idea to provide a signal handler that will trap any early exit from your program. The handler should call either jpeg_abort() or jpeg_destroy() for any active JPEG objects. A handler is not needed with jmemnobs.c, and shouldn't be necessary with jmemansi.c or jmemmac.c either, since the C library is supposed to take care of deleting files made with tmpfile().

Library compile-time options

A number of compile-time options are available by modifying jmorecfg.h.

The JPEG standard provides for both the baseline 8-bit DCT process and a 12-bit DCT process. 12-bit lossy JPEG is supported if you define BITS_IN_JSAMPLE as 12 rather than 8. Note that this causes JSAMPLE to be larger than a char, so it affects the surrounding application's image data. The sample applications cjpeg and djpeg can support 12-bit mode only for PPM and GIF file formats; you must disable the other file formats to compile a 12-bit cjpeg or djpeg. (install.doc has more information about that.) At present, a 12-bit library can handle *only* 12-bit images, not both precisions. (If you need to include both 8- and 12-bit libraries in a single application, you could probably do it by defining NEED_SHORT_EXTERNAL_NAMES for just one of the copies. You'd have to access the 8-bit and 12-bit copies from separate application source files. This is untested ... if you try it, we'd like to hear whether it works!)

Note that a 12-bit library always compresses in Huffman optimization mode, in order to generate valid Huffman tables. This is necessary because our default Huffman tables only cover 8-bit data. If you need to output 12-bit files in one pass, you'll have to supply suitable default Huffman tables.

The maximum number of components (color channels) in the image is determined by MAX_COMPONENTS. The JPEG standard allows up to 255 components, but we expect that few applications will need more than four or so.

On machines with unusual data type sizes, you may be able to improve performance or reduce memory space by tweaking the various typedefs in jmorecfg.h. In particular, on some RISC CPUs, access to arrays of "short"s is quite slow; consider trading memory for speed by making JCOEF, INT16, and UINT16 be "int" or "unsigned int". UINT8 is also a candidate to become int.

You probably don't want to make JSAMPLE be int unless you have lots of memory to burn.

You can reduce the size of the library by compiling out various optional functions. To do this, undefine xxx_SUPPORTED symbols as necessary.

Portability considerations

The JPEG library has been written to be extremely portable; the sample applications cjpeg and djpeg are slightly less so. This section summarizes the design goals in this area. (If you encounter any bugs that cause the library to be less portable than is claimed here, we'd appreciate hearing about them.)

The code works fine on both ANSI and pre-ANSI C compilers, using any of the popular system include file setups, and some not-so-popular ones too. See install.doc for configuration procedures.

The code is not dependent on the exact sizes of the C data types. As distributed, we make the assumptions that

```
      char  is at least 8 bits wide
      short is at least 16 bits wide
      int   is at least 16 bits wide
      long  is at least 32 bits wide
```

(These are the minimum requirements of the ANSI C standard.) Wider types will work fine, although memory may be used inefficiently if char is much larger than 8 bits or short is much bigger than 16 bits. The code should work equally well with 16- or 32-bit ints.

In a system where these assumptions are not met, you may be able to make the code work by modifying the typedefs in jmorecfg.h. However, you will probably have difficulty if int is less than 16 bits wide, since references to plain int abound in the code.

char can be either signed or unsigned, although the code runs faster if an unsigned char type is available. If char is wider than 8 bits, you will need to redefine JOCTET and/or provide custom data source/destination managers so that JOCTET represents exactly 8 bits of data on external storage.

The JPEG library proper does not assume ASCII representation of characters. But some of the image file I/O modules in cjpeg/djpeg do have ASCII dependencies in file-header manipulation; so does cjpeg's select_file_type() routine.

The JPEG library does not rely heavily on the C library. In particular, C stdio is used only by the data source/destination modules and the error handler, all of which are application-replaceable. (cjpeg/djpeg are more heavily dependent on stdio.) malloc and free are called only from the memory manager "back end" module, so you can use a different memory allocator by replacing that one file.

The code generally assumes that C names must be unique in the first 15 characters. However, global function names can be made unique in the first 6 characters by defining NEED_SHORT_EXTERNAL_NAMES.

More info about porting the code may be gleaned by reading jconfig.doc,
jmorecfg.h, and jinclude.h.

Notes for MS-DOS implementors

The IJG code is designed to work efficiently in 80x86 "small" or "medium" memory models (i.e., data pointers are 16 bits unless explicitly declared "far"; code pointers can be either size). You may be able to use small model to compile cjpeg or djpeg by itself, but you will probably have to use medium model for any larger application. This won't make much difference in performance. You *will* take a noticeable performance hit if you use a large-data memory model (perhaps 10%-25%), and you should avoid "huge" model if at all possible.

The JPEG library typically needs 2Kb-3Kb of stack space. It will also malloc about 20K-30K of near heap space while executing (and lots of far heap, but that doesn't count in this calculation). This figure will vary depending on selected operating mode, and to a lesser extent on image size. There is also about 5Kb-6Kb of constant data which will be allocated in the near data segment (about 4Kb of this is the error message table).
Thus you have perhaps 20K available for other modules' static data and near heap space before you need to go to a larger memory model. The C library's static data will account for several K of this, but that still leaves a good deal for your needs. (If you are tight on space, you could reduce the sizes of the I/O buffers allocated by jdatasrc.c and jdatadst.c, say from 4K to 1K. Another possibility is to move the error message table to far memory; this should be doable with only localized hacking on jerror.c.)

About 2K of the near heap space is "permanent" memory that will not be released until you destroy the JPEG object. This is only an issue if you save a JPEG object between compression or decompression operations.

Far data space may also be a tight resource when you are dealing with large images. The most memory-intensive case is decompression with two-pass color quantization, or single-pass quantization to an externally supplied color map. This requires a 128Kb color lookup table plus strip buffers amounting to about 50 bytes per column for typical sampling ratios (eg, about 32000 bytes for a 640-pixel-wide image). You may not be able to process wide images if you have large data structures of your own.

Of course, all of these concerns vanish if you use a 32-bit flat-memory-model compiler, such as DJGPP or Watcom C. We highly recommend flat model if you can use it; the JPEG library is significantly faster in flat model.

----- end libjpeg.txt inclusion -----
----- begin structure.txt inclusion -----

IJG JPEG LIBRARY: SYSTEM ARCHITECTURE

This file provides an overview of the architecture of the IJG JPEG software;

that is, the functions of the various modules in the system and the interfaces between modules. For more precise details about any data structure or calling convention, see the include files and comments in the source code.

We assume that the reader is already somewhat familiar with the JPEG standard. The README file includes references for learning about JPEG. The file libjpeg.doc describes the library from the viewpoint of an application programmer using the library; it's best to read that file before this one. Also, the file coderules.doc describes the coding style conventions we use.

In this document, JPEG-specific terminology follows the JPEG standard:

  A "component" means a color channel, e.g., Red or Luminance.
  A "sample" is a single component value (i.e., one number in the image data).
  A "coefficient" is a frequency coefficient (a DCT transform output number).
  A "block" is an 8x8 group of samples or coefficients.
  An "MCU" (minimum coded unit) is an interleaved set of blocks of size determined by the sampling factors, or a single block in a noninterleaved scan.

We do not use the terms "pixel" and "sample" interchangeably. When we say pixel, we mean an element of the full-size image, while a sample is an element of the downsampled image. Thus the number of samples may vary across components while the number of pixels does not. (This terminology is not used rigorously throughout the code, but it is used in places where confusion would otherwise result.)

*** System features ***

The IJG distribution contains two parts:

* A subroutine library for JPEG compression and decompression.
* cjpeg/djpeg, two sample applications that use the library to transform JFIF JPEG files to and from several other image formats.

cjpeg/djpeg are of no great intellectual complexity: they merely add a simple command-line user interface and I/O routines for several uncompressed image formats. This document concentrates on the library itself.

We desire the library to be capable of supporting all JPEG baseline, extended sequential, and progressive DCT processes. Hierarchical processes are not supported.

The library does not support the lossless (spatial) JPEG process. Lossless JPEG shares little or no code with lossy JPEG, and would normally be used without the extensive pre- and post-processing provided by this library. We feel that lossless JPEG is better handled by a separate library.

Within these limits, any set of compression parameters allowed by the JPEG spec should be readable for decompression. (We can be more restrictive about what formats we can generate.) Although the system design allows for all parameter values, some uncommon settings are not yet implemented and may never be; nonintegral sampling ratios are the prime example. Furthermore, we treat 8-bit vs. 12-bit data precision as a compile-time switch, not a run-time option, because most machines can store 8-bit pixels much more compactly than 12-bit.

For legal reasons, JPEG arithmetic coding is not currently supported, but extending the library to include it would be straightforward.

By itself, the library handles only interchange JPEG datastreams --- in particular the widely used JFIF file format. The library can be used by surrounding code to process interchange or abbreviated JPEG datastreams that are embedded in more complex file formats. (For example, libtiff uses this library to implement JPEG compression within the TIFF file format.)

The library includes a substantial amount of code that is not covered by the JPEG standard but is necessary for typical applications of JPEG. These functions preprocess the image before JPEG compression or postprocess it after decompression. They include colorspace conversion, downsampling/upsampling, and color quantization. This code can be omitted if not needed.

A wide range of quality vs. speed tradeoffs are possible in JPEG processing, and even more so in decompression postprocessing. The decompression library provides multiple implementations that cover most of the useful tradeoffs, ranging from very-high-quality down to fast-preview operation. On the compression side we have generally not provided low-quality choices, since compression is normally less time-critical. It should be understood that the low-quality modes may not meet the JPEG standard's accuracy requirements; nonetheless, they are useful for viewers.

*** Portability issues ***

Portability is an essential requirement for the library. The key portability issues that show up at the level of system architecture are:

1. Memory usage. We want the code to be able to run on PC-class machines with limited memory. Images should therefore be processed sequentially (in strips), to avoid holding the whole image in memory at once. Where a full-image buffer is necessary, we should be able to use either virtual memory or temporary files.

2. Near/far pointer distinction. To run efficiently on 80x86 machines, the code should distinguish "small" objects (kept in near data space) from "large" ones (kept in far data space). This is an annoying restriction, but fortunately it does not impact code quality for less brain-damaged machines, and the source code clutter turns out to be minimal with sufficient use of pointer typedefs.

3. Data precision. We assume that "char" is at least 8 bits, "short" and "int" at least 16, "long" at least 32. The code will work fine with larger data sizes, although memory may be used inefficiently in some cases. However, the JPEG compressed datastream must ultimately appear on external storage as a sequence of 8-bit bytes if it is to conform to the standard. This may pose a problem on machines where char is wider than 8 bits. The library represents compressed data as an array of values of typedef JOCTET. If no data type exactly 8 bits wide is available, custom data source and data destination modules must be written to unpack and pack the chosen JOCTET datatype into 8-bit external representation.

*** System overview ***

The compressor and decompressor are each divided into two main sections: the JPEG compressor or decompressor proper, and the preprocessing or postprocessing functions. The interface between these two sections is the image data that the official JPEG spec regards as its input or output: this data is in the colorspace to be used for compression, and it is downsampled to the sampling factors to be used. The preprocessing and postprocessing steps are responsible for converting a normal image representation to or from this form. (Those few applications that want to deal with YCbCr downsampled data can skip the preprocessing or postprocessing step.)

Looking more closely, the compressor library contains the following main elements:

Preprocessing:
  * Color space conversion (e.g., RGB to YCbCr).
  * Edge expansion and downsampling. Optionally, this step can do simple smoothing --- this is often helpful for low-quality source data.

JPEG proper:
  * MCU assembly, DCT, quantization.
  * Entropy coding (sequential or progressive, Huffman or arithmetic).

In addition to these modules we need overall control, marker generation, and support code (memory management & error handling). There is also a module responsible for physically writing the output data --- typically this is just an interface to fwrite(), but some applications may need to do something else with the data.

The decompressor library contains the following main elements:

JPEG proper:
  * Entropy decoding (sequential or progressive, Huffman or arithmetic).
  * Dequantization, inverse DCT, MCU disassembly.

Postprocessing:
  * Upsampling. Optionally, this step may be able to do more general rescaling of the image.
  * Color space conversion (e.g., YCbCr to RGB). This step may also provide gamma adjustment [currently it does not].
  * Optional color quantization (e.g., reduction to 256 colors).
  * Optional color precision reduction (e.g., 24-bit to 15-bit color). [This feature is not currently implemented.]

We also need overall control, marker parsing, and a data source module. The support code (memory management & error handling) can be shared with the compression half of the library.

There may be several implementations of each of these elements, particularly in the decompressor, where a wide range of speed/quality tradeoffs is very useful. It must be understood that some of the best speedups involve merging adjacent steps in the pipeline. For example, upsampling, color space conversion, and color quantization might all be done at once when using a low-quality ordered-dither technique. The system architecture is designed to allow such merging where appropriate.

Note: it is convenient to regard edge expansion (padding to block boundaries) as a preprocessing/postprocessing function, even though the JPEG spec includes it in compression/decompression. We do this because downsampling/upsampling can be simplified a little if they work on padded data: it's not necessary to have special cases at the right and bottom edges. Therefore the interface buffer is always an integral number of blocks wide and high, and we expect compression preprocessing to pad the source data properly. Padding will occur only to the next block (8-sample) boundary. In an interleaved-scan situation, additional dummy blocks may be used to fill out MCUs, but the MCU assembly and disassembly logic will create or discard these blocks internally. (This is advantageous for speed reasons, since we avoid DCTing the dummy blocks. It also permits a small reduction in file size, because the compressor can choose dummy block contents so as to minimize their size in compressed form. Finally, it makes the interface buffer specification independent of whether the file is actually interleaved or not.) Applications that wish to deal directly with the downsampled data must provide similar buffering and padding for odd-sized images.

*** Poor man's object-oriented programming ***

It should be clear by now that we have a lot of quasi-independent processing steps, many of which have several possible behaviors. To avoid cluttering the code with lots of switch statements, we use a simple form of object-style programming to separate out the different possibilities.

For example, two different color quantization algorithms could be implemented as two separate modules that present the same external interface; at runtime, the calling code will access the proper module indirectly through an "object".

We can get the limited features we need while staying within portable C. The basic tool is a function pointer. An "object" is just a struct containing one or more function pointer fields, each of which corresponds to a method name in real object-oriented languages. During initialization we fill in the function pointers with references to whichever module we have determined we need to use in this run. Then invocation of the module is done by indirecting through a function pointer; on most machines this is no more expensive than a switch statement, which would be the only other way of making the required run-time choice. The really significant benefit, of course, is keeping the source code clean and well structured.

We can also arrange to have private storage that varies between different implementations of the same kind of object. We do this by making all the module-specific object structs be separately allocated entities, which will be accessed via pointers in the master compression or decompression struct. The "public" fields or methods for a given kind of object are specified by a commonly known struct. But a module's initialization code can allocate a larger struct that contains the common struct as its first member, plus additional private fields. With appropriate pointer casting, the module's internal functions can access these private fields. (For a simple example, see jdatadst.c, which implements the external interface specified by struct jpeg_destination_mgr, but adds extra fields.)

(Of course this would all be a lot easier if we were using C++, but we are not yet prepared to assume that everyone has a C++ compiler.)

An important benefit of this scheme is that it is easy to provide multiple versions of any method, each tuned to a particular case. While a lot of precalculation might be done to select an optimal implementation of a method, the cost per invocation is constant. For example, the upsampling step might have a "generic" method, plus one or more "hardwired" methods for the most popular sampling factors; the hardwired methods would be faster because they'd use straight-line code instead of for-loops. The cost to determine which method to use is paid only once, at startup, and the selection criteria are hidden from the callers of the method.

This plan differs a little bit from usual object-oriented structures, in that only one instance of each object class will exist during execution. The reason for having the class structure is that on different runs we may create different instances (choose to execute different modules). You can think of the term "method" as denoting the common interface presented by a particular set of interchangeable functions, and "object" as denoting a group of related methods, or the total shared interface behavior of a group of modules.

*** Overall control structure ***

We previously mentioned the need for overall control logic in the compression and decompression libraries. In IJG implementations prior to v5, overall control was mostly provided by "pipeline control" modules, which proved to be large, unwieldy, and hard to understand. To improve the situation, the control logic has been subdivided into multiple modules. The control modules consist of:

1. Master control for module selection and initialization. This has two responsibilities:

   1A. Startup initialization at the beginning of image processing. The individual processing modules to be used in this run are selected and given initialization calls.

   1B. Per-pass control. This determines how many passes will be performed and calls each active processing module to configure itself appropriately at the beginning of each pass. End-of-pass processing, where necessary, is also invoked from the master control module.

   Method selection is partially distributed, in that a particular processing module may contain several possible implementations of a particular method, which it will select among when given its initialization call. The master control code need only be concerned with decisions that affect more than one module.

2. Data buffering control. A separate control module exists for each inter-processing-step data buffer. This module is responsible for invoking the processing steps that write or read that data buffer.

Each buffer controller sees the world as follows:

```
input data => processing step A => buffer => processing step B => output data
                      |              |               |
              ------------------ controller ------------------
```

The controller knows the dataflow requirements of steps A and B: how much data they want to accept in one chunk and how much they output in one chunk. Its function is to manage its buffer and call A and B at the proper times.

A data buffer control module may itself be viewed as a processing step by a higher-level control module; thus the control modules form a binary tree with elementary processing steps at the leaves of the tree.

The control modules are objects. A considerable amount of flexibility can be had by replacing implementations of a control module. For example:

* Merging of adjacent steps in the pipeline is done by replacing a control module and its pair of processing-step modules with a single processing-step module. (Hence the possible merges are determined by the tree of control modules.)
* In some processing modes, a given interstep buffer need only be a "strip" buffer large enough to accommodate the desired data chunk sizes. In other modes, a full-image buffer is needed and several passes are required. The control module determines which kind of buffer is used and manipulates virtual array buffers as needed. One or both processing steps may be unaware of the multi-pass behavior.

In theory, we might be able to make all of the data buffer controllers interchangeable and provide just one set of implementations for all. In practice, each one contains considerable special-case processing for its particular job. The buffer controller concept should be regarded as an overall system structuring principle, not as a complete description of the task performed by any one controller.

*** Compression object structure ***

Here is a sketch of the logical structure of the JPEG compression library:

```
                                                 |-- Colorspace conversion
                  |-- Preprocessing controller --|
                  |                              |-- Downsampling
Main controller --|
                  |                            |-- Forward DCT, quantize
                  |-- Coefficient controller --|
                                               |-- Entropy encoding
```

This sketch also describes the flow of control (subroutine calls) during typical image data processing. Each of the components shown in the diagram is an "object" which may have several different implementations available. One or more source code files contain the actual implementation(s) of each object.

The objects shown above are:

* Main controller: buffer controller for the subsampled-data buffer, which holds the preprocessed input data. This controller invokes preprocessing to fill the subsampled-data buffer, and JPEG compression to empty it. There is usually no need for a full-image buffer here; a strip buffer is adequate.

* Preprocessing controller: buffer controller for the downsampling input data buffer, which lies between colorspace conversion and downsampling. Note that a unified conversion/downsampling module would probably replace this controller entirely.

* Colorspace conversion: converts application image data into the desired JPEG color space; also changes the data from pixel-interleaved layout to separate component planes. Processes one pixel row at a time.

* Downsampling: performs reduction of chroma components as required. Optionally may perform pixel-level smoothing as well. Processes a "row group" at a time, where a row group is defined as Vmax pixel rows of each

component before downsampling, and Vk sample rows afterwards (remember Vk differs across components). Some downsampling or smoothing algorithms may require context rows above and below the current row group; the preprocessing controller is responsible for supplying these rows via proper buffering. The downsampler is responsible for edge expansion at the right edge (i.e., extending each sample row to a multiple of 8 samples); but the preprocessing controller is responsible for vertical edge expansion (i.e., duplicating the bottom sample row as needed to make a multiple of 8 rows).

* Coefficient controller: buffer controller for the DCT-coefficient data. This controller handles MCU assembly, including insertion of dummy DCT blocks when needed at the right or bottom edge. When performing Huffman-code optimization or emitting a multiscan JPEG file, this controller is responsible for buffering the full image. The equivalent of one fully interleaved MCU row of subsampled data is processed per call, even when the JPEG file is noninterleaved.

* Forward DCT and quantization: Perform DCT, quantize, and emit coefficients. Works on one or more DCT blocks at a time. (Note: the coefficients are now emitted in normal array order, which the entropy encoder is expected to convert to zigzag order as necessary. Prior versions of the IJG code did the conversion to zigzag order within the quantization step.)

* Entropy encoding: Perform Huffman or arithmetic entropy coding and emit the coded data to the data destination module. Works on one MCU per call. For progressive JPEG, the same DCT blocks are fed to the entropy coder during each pass, and the coder must emit the appropriate subset of coefficients.

In addition to the above objects, the compression library includes these objects:

* Master control: determines the number of passes required, controls overall and per-pass initialization of the other modules.

* Marker writing: generates JPEG markers (except for RSTn, which is emitted by the entropy encoder when needed).

* Data destination manager: writes the output JPEG datastream to its final destination (e.g., a file). The destination manager supplied with the library knows how to write to a stdio stream; for other behaviors, the surrounding application may provide its own destination manager.

* Memory manager: allocates and releases memory, controls virtual arrays (with backing store management, where required).

* Error handler: performs formatting and output of error and trace messages; determines handling of nonfatal errors. The surrounding application may override some or all of this object's methods to change error handling.

* Progress monitor: supports output of "percent-done" progress reports. This object represents an optional callback to the surrounding application: if wanted, it must be supplied by the application.

The error handler, destination manager, and progress monitor objects are defined as separate objects in order to simplify application-specific customization of the JPEG library. A surrounding application may override individual methods or supply its own all-new implementation of one of these objects. The object interfaces for these objects are therefore treated as part of the application interface of the library, whereas the other objects are internal to the library.

The error handler and memory manager are shared by JPEG compression and decompression; the progress monitor, if used, may be shared as well.

*** Decompression object structure ***

Here is a sketch of the logical structure of the JPEG decompression library:

```
                                                 |-- Entropy decoding
                  |-- Coefficient controller --|
                  |                              |-- Dequantize, Inverse DCT
Main controller --|
                  |                                 |-- Upsampling
                  |-- Postprocessing controller --|   |-- Colorspace conversion
                                                    |-- Color quantization
                                                    |-- Color precision reduction
```

As before, this diagram also represents typical control flow. The objects shown are:

* Main controller: buffer controller for the subsampled-data buffer, which holds the output of JPEG decompression proper. This controller's primary task is to feed the postprocessing procedure. Some upsampling algorithms may require context rows above and below the current row group; when this is true, the main controller is responsible for managing its buffer so as to make context rows available. In the current design, the main buffer is always a strip buffer; a full-image buffer is never required.

* Coefficient controller: buffer controller for the DCT-coefficient data. This controller handles MCU disassembly, including deletion of any dummy DCT blocks at the right or bottom edge. When reading a multiscan JPEG file, this controller is responsible for buffering the full image. (Buffering DCT coefficients, rather than samples, is necessary to support progressive JPEG.) The equivalent of one fully interleaved MCU row of subsampled data is processed per call, even when the source JPEG file is noninterleaved.

* Entropy decoding: Read coded data from the data source module and perform Huffman or arithmetic entropy decoding. Works on one MCU per call. For progressive JPEG decoding, the coefficient controller supplies the prior coefficients of each MCU (initially all zeroes), which the entropy decoder modifies in each scan.

* Dequantization and inverse DCT: like it says. Note that the coefficients buffered by the coefficient controller have NOT been dequantized; we merge dequantization and inverse DCT into a single step for speed reasons. When scaled-down output is asked for, simplified DCT algorithms may be used that emit only 1x1, 2x2, or 4x4 samples per DCT block, not the full 8x8. Works on one DCT block at a time.

* Postprocessing controller: buffer controller for the color quantization input buffer, when quantization is in use. (Without quantization, this

controller just calls the upsampler.) For two-pass quantization, this controller is responsible for buffering the full-image data.

* Upsampling: restores chroma components to full size. (May support more general output rescaling, too. Note that if undersized DCT outputs have been emitted by the DCT module, this module must adjust so that properly sized outputs are created.) Works on one row group at a time. This module also calls the color conversion module, so its top level is effectively a buffer controller for the upsampling->color conversion buffer. However, in all but the highest-quality operating modes, upsampling and color conversion are likely to be merged into a single step.

* Colorspace conversion: convert from JPEG color space to output color space, and change data layout from separate component planes to pixel-interleaved. Works on one pixel row at a time.

* Color quantization: reduce the data to colormapped form, using either an externally specified colormap or an internally generated one. This module is not used for full-color output. Works on one pixel row at a time; may require two passes to generate a color map. Note that the output will always be a single component representing colormap indexes. In the current design, the output values are JSAMPLEs, so an 8-bit compilation cannot quantize to more than 256 colors. This is unlikely to be a problem in practice.

* Color reduction: this module handles color precision reduction, e.g., generating 15-bit color (5 bits/primary) from JPEG's 24-bit output. Not quite clear yet how this should be handled... should we merge it with colorspace conversion???

Note that some high-speed operating modes might condense the entire postprocessing sequence to a single module (upsample, color convert, and quantize in one step).

In addition to the above objects, the decompression library includes these objects:

* Master control: determines the number of passes required, controls overall and per-pass initialization of the other modules. This is subdivided into input and output control: jdinput.c controls only input-side processing, while jdmaster.c handles overall initialization and output-side control.

* Marker reading: decodes JPEG markers (except for RSTn).

* Data source manager: supplies the input JPEG datastream. The source manager supplied with the library knows how to read from a stdio stream; for other behaviors, the surrounding application may provide its own source manager.

* Memory manager: same as for compression library.

* Error handler: same as for compression library.

* Progress monitor: same as for compression library.

As with compression, the data source manager, error handler, and progress monitor are candidates for replacement by a surrounding application.

*** Decompression input and output separation ***

To support efficient incremental display of progressive JPEG files, the decompressor is divided into two sections that can run independently:

1. Data input includes marker parsing, entropy decoding, and input into the coefficient controller's DCT coefficient buffer. Note that this processing is relatively cheap and fast.

2. Data output reads from the DCT coefficient buffer and performs the IDCT and all postprocessing steps.

For a progressive JPEG file, the data input processing is allowed to get arbitrarily far ahead of the data output processing. (This occurs only if the application calls jpeg_consume_input(); otherwise input and output run in lockstep, since the input section is called only when the output section needs more data.) In this way the application can avoid making extra display passes when data is arriving faster than the display pass can run. Furthermore, it is possible to abort an output pass without losing anything, since the coefficient buffer is read-only as far as the output section is concerned. See libjpeg.doc for more detail.

A full-image coefficient array is only created if the JPEG file has multiple scans (or if the application specifies buffered-image mode anyway). When reading a single-scan file, the coefficient controller normally creates only a one-MCU buffer, so input and output processing must run in lockstep in this case. jpeg_consume_input() is effectively a no-op in this situation.

The main impact of dividing the decompressor in this fashion is that we must be very careful with shared variables in the cinfo data structure. Each variable that can change during the course of decompression must be classified as belonging to data input or data output, and each section must look only at its own variables. For example, the data output section may not depend on any of the variables that describe the current scan in the JPEG file, because these may change as the data input section advances into a new scan.

The progress monitor is (somewhat arbitrarily) defined to treat input of the file as one pass when buffered-image mode is not used, and to ignore data input work completely when buffered-image mode is used. Note that the library has no reliable way to predict the number of passes when dealing with a progressive JPEG file, nor can it predict the number of output passes in buffered-image mode. So the work estimate is inherently bogus anyway.

No comparable division is currently made in the compression library, because there isn't any real need for it.

*** Data formats ***

Arrays of pixel sample values use the following data structure:

```
    typedef something JSAMPLE;            a pixel component value, 0..MAXJSAMPLE
    typedef JSAMPLE *JSAMPROW;            ptr to a row of samples
    typedef JSAMPROW *JSAMPARRAY;   ptr to a list of rows
```

```
    typedef JSAMPARRAY *JSAMPIMAGE; ptr to a list of color-component arrays
```

The basic element type JSAMPLE will typically be one of unsigned char, (signed) char, or short. Short will be used if samples wider than 8 bits are to be supported (this is a compile-time option). Otherwise, unsigned char is used if possible. If the compiler only supports signed chars, then it is necessary to mask off the value when reading. Thus, all reads of JSAMPLE values must be coded as "GETJSAMPLE(value)", where the macro will be defined as "((value) & 0xFF)" on signed-char machines and "((int) (value))" elsewhere.

With these conventions, JSAMPLE values can be assumed to be >= 0. This helps simplify correct rounding during downsampling, etc. The JPEG standard's specification that sample values run from -128..127 is accommodated by subtracting 128 just as the sample value is copied into the source array for the DCT step (this will be an array of signed ints). Similarly, during decompression the output of the IDCT step will be immediately shifted back to 0..255. (NB: different values are required when 12-bit samples are in use. The code is written in terms of MAXJSAMPLE and CENTERJSAMPLE, which will be defined as 255 and 128 respectively in an 8-bit implementation, and as 4095 and 2048 in a 12-bit implementation.)

We use a pointer per row, rather than a two-dimensional JSAMPLE array. This choice costs only a small amount of memory and has several benefits:

* Code using the data structure doesn't need to know the allocated width of the rows. This simplifies edge expansion/compression, since we can work in an array that's wider than the logical picture width.
* Indexing doesn't require multiplication; this is a performance win on many machines.
* Arrays with more than 64K total elements can be supported even on machines where malloc() cannot allocate chunks larger than 64K.
* The rows forming a component array may be allocated at different times without extra copying. This trick allows some speedups in smoothing steps that need access to the previous and next rows.

Note that each color component is stored in a separate array; we don't use the traditional layout in which the components of a pixel are stored together. This simplifies coding of modules that work on each component independently, because they don't need to know how many components there are. Furthermore, we can read or write each component to a temporary file independently, which is helpful when dealing with noninterleaved JPEG files.

In general, a specific sample value is accessed by code such as

```
      GETJSAMPLE(image[colorcomponent][row][col])
```

where col is measured from the image left edge, but row is measured from the first sample row currently in memory. Either of the first two indexings can be precomputed by copying the relevant pointer.

Since most image-processing applications prefer to work on images in which the components of a pixel are stored together, the data passed to or from the surrounding application uses the traditional convention: a single pixel is represented by N consecutive JSAMPLE values, and an image row is an array of (# of color components)*(image width) JSAMPLEs. One or more rows of data can be represented by a pointer of type JSAMPARRAY in this scheme. This scheme is converted to component-wise storage inside the JPEG library. (Applications that want to skip JPEG preprocessing or postprocessing will have to contend with component-wise storage.)

Arrays of DCT-coefficient values use the following data structure:

```
    typedef short JCOEF;            a 16-bit signed integer
    typedef JCOEF JBLOCK[DCTSIZE2]; an 8x8 block of coefficients
    typedef JBLOCK *JBLOCKROW;            ptr to one horizontal row of 8x8
blocks
    typedef JBLOCKROW *JBLOCKARRAY; ptr to a list of such rows
    typedef JBLOCKARRAY *JBLOCKIMAGE;     ptr to a list of color component
arrays
```

The underlying type is at least a 16-bit signed integer; while "short" is big enough on all machines of interest, on some machines it is preferable to use "int" for speed reasons, despite the storage cost. Coefficients are grouped into 8x8 blocks (but we always use #defines DCTSIZE and DCTSIZE2 rather than "8" and "64").

The contents of a coefficient block may be in either "natural" or zigzagged order, and may be true values or divided by the quantization coefficients, depending on where the block is in the processing pipeline. In the current library, coefficient blocks are kept in natural order everywhere; the entropy codecs zigzag or dezigzag the data as it is written or read. The blocks contain quantized coefficients everywhere outside the DCT/IDCT subsystems. (This latter decision may need to be revisited to support variable quantization a la JPEG Part 3.)

Notice that the allocation unit is now a row of 8x8 blocks, corresponding to eight rows of samples. Otherwise the structure is much the same as for samples, and for the same reasons.

On machines where malloc() can't handle a request bigger than 64Kb, this data structure limits us to rows of less than 512 JBLOCKs, or a picture width of 4000+ pixels. This seems an acceptable restriction.

On 80x86 machines, the bottom-level pointer types (JSAMPROW and JBLOCKROW) must be declared as "far" pointers, but the upper levels can be "near" (implying that the pointer lists are allocated in the DS segment). We use a #define symbol FAR, which expands to the "far" keyword when compiling on 80x86 machines and to nothing elsewhere.

*** Suspendable processing ***

In some applications it is desirable to use the JPEG library as an incremental, memory-to-memory filter. In this situation the data source or destination may be a limited-size buffer, and we can't rely on being able to empty or refill the buffer at arbitrary times. Instead the application would like to have control return from the library at buffer overflow/underrun, and then resume compression or decompression at a later time.

This scenario is supported for simple cases. (For anything more complex, we recommend that the application "bite the bullet" and develop real multitasking capability.) The libjpeg.doc file goes into more detail about the usage and limitations of this capability; here we address the implications for library structure.

The essence of the problem is that the entropy codec (coder or decoder) must be prepared to stop at arbitrary times. In turn, the controllers that call the entropy codec must be able to stop before having produced or consumed all the data that they normally would handle in one call. That part is reasonably straightforward: we make the controller call interfaces include "progress counters" which indicate the number of data chunks successfully processed, and we require callers to test the counter rather than just assume all of the data was processed.

Rather than trying to restart at an arbitrary point, the current Huffman codecs are designed to restart at the beginning of the current MCU after a suspension due to buffer overflow/underrun. At the start of each call, the codec's internal state is loaded from permanent storage (in the JPEG object structures) into local variables. On successful completion of the MCU, the permanent state is updated. (This copying is not very expensive, and may even lead to *improved* performance if the local variables can be registerized.) If a suspension occurs, the codec simply returns without updating the state, thus effectively reverting to the start of the MCU. Note that this implies leaving some data unprocessed in the source/destination buffer (ie, the compressed partial MCU). The data source/destination module interfaces are specified so as to make this possible. This also implies that the data buffer must be large enough to hold a worst-case compressed MCU; a couple thousand bytes should be enough.

In a successive-approximation AC refinement scan, the progressive Huffman decoder has to be able to undo assignments of newly nonzero coefficients if it suspends before the MCU is complete, since decoding requires distinguishing previously-zero and previously-nonzero coefficients. This is a bit tedious but probably won't have much effect on performance. Other variants of Huffman decoding need not worry about this, since they will just store the same values again if forced to repeat the MCU.

This approach would probably not work for an arithmetic codec, since its modifiable state is quite large and couldn't be copied cheaply. Instead it would have to suspend and resume exactly at the point of the buffer end.

The JPEG marker reader is designed to cope with suspension at an arbitrary point. It does so by backing up to the start of the marker parameter segment, so the data buffer must be big enough to hold the largest marker of interest. Again, a couple KB should be adequate. (A special "skip" convention is used to bypass COM and APPn markers, so these can be larger than the buffer size without causing problems; otherwise a 64K buffer would be needed in the worst case.)

The JPEG marker writer currently does *not* cope with suspension. I feel that this is not necessary; it is much easier simply to require the application to ensure there is enough buffer space before starting. (An empty 2K buffer is more than sufficient for the header markers; and ensuring there are a dozen or two bytes available before calling jpeg_finish_compress() will suffice for the trailer.) This would not work for writing multi-scan JPEG files, but we simply do not intend to support that capability with suspension.

*** Memory manager services ***

The JPEG library's memory manager controls allocation and deallocation of memory, and it manages large "virtual" data arrays on machines where the operating system does not provide virtual memory. Note that the same memory manager serves both compression and decompression operations.

In all cases, allocated objects are tied to a particular compression or decompression master record, and they will be released when that master record is destroyed.

The memory manager does not provide explicit deallocation of objects. Instead, objects are created in "pools" of free storage, and a whole pool can be freed at once. This approach helps prevent storage-leak bugs, and it speeds up operations whenever malloc/free are slow (as they often are). The pools can be regarded as lifetime identifiers for objects. Two pools/lifetimes are defined:

```
  * JPOOL_PERMANENT     lasts until master record is destroyed
  * JPOOL_IMAGE         lasts until done with image (JPEG datastream)
```

Permanent lifetime is used for parameters and tables that should be carried across from one datastream to another; this includes all application-visible parameters. Image lifetime is used for everything else. (A third lifetime, JPOOL_PASS = one processing pass, was originally planned. However it was dropped as not being worthwhile. The actual usage patterns are such that the peak memory usage would be about the same anyway; and having per-pass storage substantially complicates the virtual memory allocation rules --- see below.)

The memory manager deals with three kinds of object:

1. "Small" objects. Typically these require no more than 10K-20K total.
2. "Large" objects. These may require tens to hundreds of K depending on image size. Semantically they behave the same as small objects, but we distinguish them for two reasons:
   * On MS-DOS machines, large objects are referenced by FAR pointers, small objects by NEAR pointers.
   * Pool allocation heuristics may differ for large and small objects.

   Note that individual "large" objects cannot exceed the size allowed by type size_t, which may be 64K or less on some machines.
3. "Virtual" objects. These are large 2-D arrays of JSAMPLEs or JBLOCKs (typically large enough for the entire image being processed). The memory manager provides stripwise access to these arrays. On machines without virtual memory, the rest of the array may be swapped out to a temporary file.

(Note: JSAMPARRAY and JBLOCKARRAY data structures are a combination of large objects for the data proper and small objects for the row pointers. For convenience and speed, the memory manager provides single routines to create these structures. Similarly, virtual arrays include a small control block and a JSAMPARRAY or JBLOCKARRAY working buffer, all created with one call.)

In the present implementation, virtual arrays are only permitted to have image lifespan. (Permanent lifespan would not be reasonable, and pass lifespan is not very useful since a virtual array's raison d'etre is to store data for multiple passes through the image.) We also expect that only "small" objects will be given permanent lifespan, though this restriction is not required by the memory manager.

In a non-virtual-memory machine, some performance benefit can be gained by making the in-memory buffers for virtual arrays be as large as possible. (For small images, the buffers might fit entirely in memory, so blind swapping would be very wasteful.) The memory manager will adjust the height of the buffers to fit within a prespecified maximum memory usage. In order to do this in a reasonably optimal fashion, the manager needs to allocate all of the virtual arrays at once. Therefore, there isn't a one-step allocation routine for virtual arrays; instead, there is a "request" routine that simply allocates the control block, and a "realize" routine (called just once) that determines space allocation and creates all of the actual buffers. The realize routine must allow for space occupied by non-virtual large objects. (We don't bother to factor in the space needed for small objects, on the grounds that it isn't worth the trouble.)

To support all this, we establish the following protocol for doing business with the memory manager:

1. Modules must request virtual arrays (which may have only image lifespan) during the initial setup phase, i.e., in their jinit_xxx routines.
2. All "large" objects (including JSAMPARRAYs and JBLOCKARRAYs) must also be allocated during initial setup.
3. realize_virt_arrays will be called at the completion of initial setup. The above conventions ensure that sufficient information is available for it to choose a good size for virtual array buffers.

Small objects of any lifespan may be allocated at any time. We expect that the total space used for small objects will be small enough to be negligible in the realize_virt_arrays computation.

In a virtual-memory machine, we simply pretend that the available space is infinite, thus causing realize_virt_arrays to decide that it can allocate all the virtual arrays as full-size in-memory buffers. The overhead of the virtual-array access protocol is very small when no swapping occurs.

A virtual array can be specified to be "pre-zeroed"; when this flag is set, never-yet-written sections of the array are set to zero before being made available to the caller. If this flag is not set, never-written sections of the array contain garbage. (This feature exists primarily because the equivalent logic would otherwise be needed in jdcoefct.c for progressive JPEG mode; we may as well make it available for possible other uses.)

The first write pass on a virtual array is required to occur in top-to-bottom order; read passes, as well as any write passes after the first one, may access the array in any order. This restriction exists partly to simplify the virtual array control logic, and partly because some file systems may not support seeking beyond the current end-of-file in a temporary file. The main implication of this restriction is that rearrangement of rows (such as converting top-to-bottom data order to bottom-to-top) must be handled while reading data out of the virtual array, not while putting it in.

*** Memory manager internal structure ***

To isolate system dependencies as much as possible, we have broken the memory manager into two parts. There is a reasonably system-independent "front end" (jmemmgr.c) and a "back end" that contains only the code likely to change across systems. All of the memory management methods outlined above are implemented by the front end. The back end provides the following routines for use by the front end (none of these routines are known to the rest of the JPEG code):

jpeg_mem_init, jpeg_mem_term system-dependent initialization/shutdown

```
jpeg_get_small, jpeg_free_small     interface to malloc and free library
routines
                        (or their equivalents)

jpeg_get_large, jpeg_free_large     interface to FAR malloc/free in MSDOS
machines;
                        else usually the same as
                        jpeg_get_small/jpeg_free_small

jpeg_mem_available            estimate available memory

jpeg_open_backing_store       create a backing-store object

read_backing_store,           manipulate a backing-store object
write_backing_store,
close_backing_store
```

On some systems there will be more than one type of backing-store object (specifically, in MS-DOS a backing store file might be an area of extended memory as well as a disk file). jpeg_open_backing_store is responsible for choosing how to implement a given object. The read/write/close routines are method pointers in the structure that describes a given object; this lets them be different for different object types.

It may be necessary to ensure that backing store objects are explicitly released upon abnormal program termination. For example, MS-DOS won't free extended memory by itself. To support this, we will expect the main program or surrounding application to arrange to call self_destruct (typically via jpeg_destroy) upon abnormal termination. This may require a SIGINT signal handler or equivalent. We don't want to have the back end module install its own signal handler, because that would pre-empt the surrounding application's ability to control signal handling.

The IJG distribution includes several memory manager back end implementations. Usually the same back end should be suitable for all applications on a given system, but it is possible for an application to supply its own back end at need.

*** Implications of DNL marker ***

Some JPEG files may use a DNL marker to postpone definition of the image height (this would be useful for a fax-like scanner's output, for instance). In these files the SOF marker claims the image height is 0, and you only find out the true image height at the end of the first scan.

We could read these files as follows:
1. Upon seeing zero image height, replace it by 65535 (the maximum allowed).
2. When the DNL is found, update the image height in the global image descriptor.

This implies that control modules must avoid making copies of the image height, and must re-test for termination after each MCU row. This would be easy enough to do.

In cases where image-size data structures are allocated, this approach will result in very inefficient use of virtual memory or much-larger-than-necessary temporary files. This seems acceptable for something that probably won't be a mainstream usage. People might have to forgo use of memory-hogging options (such as two-pass color quantization or noninterleaved JPEG files) if they want efficient conversion of such files. (One could improve efficiency by demanding a user-supplied upper bound for the height, less than 65536; in most cases it could be much less.)

The standard also permits the SOF marker to overestimate the image height, with a DNL to give the true, smaller height at the end of the first scan. This would solve the space problems if the overestimate wasn't too great. However, it implies that you don't even know whether DNL will be used.

This leads to a couple of very serious objections:
1. Testing for a DNL marker must occur in the inner loop of the decompressor's Huffman decoder; this implies a speed penalty whether the feature is used or not.
2. There is no way to hide the last-minute change in image height from an application using the decoder. Thus *every* application using the IJG library would suffer a complexity penalty whether it cared about DNL or not.

We currently do not support DNL because of these problems.

A different approach is to insist that DNL-using files be preprocessed by a separate program that reads ahead to the DNL, then goes back and fixes the SOF marker. This is a much simpler solution and is probably far more efficient. Even if one wants piped input, buffering the first scan of the JPEG file needs a lot smaller temp file than is implied by the maximum-height method. For this approach we'd simply treat DNL as a no-op in the decompressor (at most, check that it matches the SOF image height).

We will not worry about making the compressor capable of outputting DNL. Something similar to the first scheme above could be applied if anyone ever wants to make that work.