NAME

hcompress, hdecompress, fcompress, fdecompress – compress and decompress images

SYNOPSIS

hcomp -v -i raw|net|fits|hhh -s *scale* **-r** *rows* **-c** *columns* [*datafile* [*headerfile*]]

hdecomp -v -s -o raw|net|fits|hhh

hcompress -s *scale* **-k -r** *files*... [*options*] *files*...

fcompress -s *scale* **-k -r** *files*... [*options*] *files*...

hdecompress -s -u -k -r -o raw|net|fits|hhh *files*... [*options*] *files*...

fdecompress -s -u -k -r -o raw|net|fits|hhh *files*... [*options*] *files*...

DESCRIPTION

hcompress is an image compression package written to compress the Space Telescope Science Institute (STScI) Digitized Sky Survey images. It is well-suited to the compression of astronomical images, which generally consist of flat sky sprinkled with point sources (stars) and the occasional extended source (galaxies and nebulae). Briefly, the method used is:

(1) a wavelet transform called the H-transform (a Haar transform generalized to two dimensions), followed by

(2) quantization that discards noise in the image while retaining the signal on all scales, followed by

(3) quadtree coding of the quantized coefficients.

The technique gives very good compression for astronomical images and is fast, requiring about 4 seconds for compression or decompression of a 512x512 image on a Sun SPARCstation 1. The calculations are carried out using integer arithmetic and are entirely reversible. Consequently, the program can be used for either lossy or lossless compression, with no special approach needed for the lossless case (e.g. there is no need for a file of residuals.)

These programs can compress 2-byte integer images in a small variety of input formats. The resulting compressed image file is machine-independent and may be transferred between computers with binary FTP.

The executable programs **hcomp** and **hdecomp** are the basic tools for compression and decompression of images. Users with images in FITS format will generally find it easier to use the **fcompress** and **fdecompress** scripts, while users with images in the IRAF and STSDAS hhh/hhd file formats will prefer the **hcompress** and **hdecompress** scripts. (FITS is the Flexible Image Transport System, which is the universal standard for astronomical data. IRAF is a widely used astronomical data analysis system; STSDAS is a package of IRAF tasks designed for Hubble Space Telescope data.) Those who wish to use this compression software for non-astronomical images will probably find these scripts inadequate and are encouraged to create their own scripts using these as a model. It also would be rather easy to modify the source code so that 8-bit images can be used as input.

In Unix these programs and scripts can be installed in some directory found in your path using the **make install** command. In VMS it is necessary to define symbols pointing to the directory and the programs by executing the @**hcompinit** command file.

To compress a 2-byte integer image in FITS format,

> **fcompress -s** *scale* image.fits

produces an output file named image.fits.H (or image.fits_H in VMS) with the compressed image. The input file image.fits is not removed unless the **-r** option is specified. To decompress this image,

fdecompress image.fits.H

will create a FITS file called image.fits.

The **-s** *scale* option determines the amount of compression. *Scale* = 0 or 1 leads to lossless compression, i.e. the decompressed image has exactly the same pixel values as the original image. If *scale* > 1 then the compression is lossy: the decompressed image will *not* be exactly the same as the original. For astronomical images, lossless compression is generally rather ineffective because the images have a good deal of noise, which is inherently incompressible. However, if some of this noise is discarded then the images compress very well. The *scale* factor determines how much of the noise is discarded. We have found that setting *scale* to 2 times sigma, the RMS noise in the image, usually results in compression by about a factor of 10 (i.e. the compressed image requires about 1.5 bits/pixel), while producing a decompressed image that is nearly indistinguishable from the original. In fact, the RMS difference between the decompressed image and the original image will be only about 1/2 sigma. Our experiments indicate that this level of loss has no noticeable effect on either the visual appearance of the image or on quantitative analysis of the image (e.g. measurements of positions and brightnesses of stars are not adversely affected.)

Using a larger value for *scale* results in higher compression at the cost of larger differences between the compressed and original images. A rough rule of thumb is that if *scale* equals N sigma, then the image will compress to about 3/N bits/pixel, and the RMS difference between the original and the compressed image will be about N/4 sigma. This crude relationship is inaccurate both for very high compression ratios and for lossless compression, but it does at least give an indication of what to expect of the compressed images.

For images in which the noise varies from pixel to pixel (e.g. CCD images, where the noise is larger for brighter pixels), the appropriate value for *scale* is determined by the RMS noise level in the sky regions of the image. For images that are essentially noiseless, any lossy compression is noticeable under sufficiently close inspection of the image, but some loss is nonetheless acceptable for typical applications. Note that the quantization scheme used in **hcompress** is *not* designed to give images that appear as much like the original as possible to the human eye, but rather is designed to produce images that are as similar as possible to the original under quantitative analysis. Thus, our emphasis is on discarding noise without affecting the signal rather than on discarding components of the image that are not very noticeable to the eye (as may be done, for example, by JPEG compression.) The resulting compression scheme is not ideal for typical terrestrial images (though it is still a reasonably good method for those images), but we believe it is close to optimal for astronomical images.

OPTIONS

hcomp -v -i raw|net|fits|hhh -s *scale* **-r** *rows* **-c** *columns* [*datafile* [*headerfile*]]

hcomp reads an image from the standard input (or from *datafile* and *headerfile* if specified) and writes a compressed image to the standard output.

-v Verbose. A description of the input file and the compression achieved is written to the standard error output.

-i raw|net|fits|hhh
Input format. Choices are:

raw Pixel values have 2 bytes in machine-dependent order, i.e. no byte swapping is done on input. This is the default format if none is specified.

net Pixel values have 2 bytes in "network" order: high byte first, then low byte for each pixel. Byte-swapping is done on input if needed. Note that this is the same as raw format on some machines (e.g. Suns) but is different on others (e.g. VAXes). Files in net format can be transferred from one machine to another without modification, but files in raw format cannot.

fits FITS (Flexible Image Transport System) format image. Header gives image size. Bytes are in network order. FITS format is the universal standard for transport of astronomical images.

hhh FITS-type header (but without padding lines making header multiple of 2880 bytes) in a file named *rootname.xxh*, with the pixel values (bytes in machine-dependent order) in a different file named *rootname.xxd*. No byte-swapping is done on input.

-s *scale* Scale factor. Non-negative integer that determines the amount of compression. Default is *scale* = 1024, which could be wildly inappropriate for your images. Use *scale* = 0 for lossless compression. See the discussion above for more information.

-r Pixels/row. This is the quickly varying dimension in the file, i.e. the order of pixels in the file is all pixels from first row, all pixels from second row, etc. Note that the size of the image can be specified this way only for **raw** and **net** format images; **fits** and **hhh** format images include a header giving the size of the image. Default value is 256.

-c Pixels/column. This is the slowly varying dimension in the file. Default value is 256.

datafile Input data file. If not specified the input is read from stdin.

headerfile
Input header file. This is only specified for format **hhh**, in which the header and data are in separate files. It is also possible to concatenate the header and data files into stdin for **hhh** format images.

hdecomp -v -s -o raw|net|fits|hhh

hdecomp reads a compressed image from the standard input and writes a decompressed image to the standard output.

-v Verbose. A description of the output file is written to the standard error output.

-s Smoothing flag. At high compressions the decompressed image begins to appear blocky because of the way information is discarded. This blockiness is greatly reduced, producing more pleasing images, if the **-s** flag is specified during decompression. Default is no smoothing.

-o raw|net|fits|hhh
Output format. Default is **fits** for input files that were in FITS format and is **raw** for other input file formats. Note that this means that an image that was originally in **net** format is not, by default, decompressed to the same format. Other input formats are properly decompressed to the same format on output.

hcompress -s *scale* **-k -r** *files*... [*options*] *files*...

hcompress compresses one or more images in **hhh** input format.

-s *scale* Scale factor. Default in script is 666. Note that this is different than the default in **hcomp**. You can modify the script to insert a typical value for your images as the default.

-k Keep original files after compression (default).

-r Remove original files after compression. Only the data file (hhd) is removed, the header file is retained in any case.

files... One or more hhh/hhd format images to compress. Wild cards may be used. The names of the data files (not the header files) should be specified. Note that one can mix options and file names; options apply only to files that follow them on the command line. For example, the removal of the original images may be turned on and off using the **-r** and **-k** flags. If *infile.xxd* is the name of an original image data file, the resulting compressed image is in a file named *infile.xxd***.H** (or *infile.xxd*_**H** in VMS.)

fcompress -s *scale* **-k -r** *files*. . . [*options*] *files*. . .

fcompress compresses one or more images in **fits** input format.

-s *scale* Scale factor. As for **hcompress**, the script default value is 666.

-k Keep original files after compression (default).

-r Remove original files after compression.

files. . . One or more FITS format images to compress. If *infile* is the name of an input FITS file, then the resulting compressed image is in a file named *infile***.H** (or *infile*_**H** in VMS.)

hdecompress -u -s -k -r -o raw|**net**|**fits**|**hhh** *files*. . . [*options*] *files*. . .

fdecompress -u -s -k -r -o raw|**net**|**fits**|**hhh** *files*. . . [*options*] *files*. . .

These programs decompress one or more compressed image files. **hdecompress** and **fdecompress** are actually the same program; the two different names are provided for consistency with the two compression programs, **hcompress** and **fcompress**.

-u Disable smoothing (default).

-s Enable smoothing.

-k Keep compressed image file after decompression (default).

-r Remove compressed image file after decompression.

-o raw|**net**|**fits**|**hhh**

Output format. Default is **fits** for files originally in FITS format, and **raw** for other files. Note that **raw** format output is correct for **hhh** format data files.

files. . . Names of files to decompress. The **.H** extension is added to the end of the filename if it is not explicitly specified. (In VMS the filenames must already have _**H** on the end of the file extension.) For an input compressed image in *infile***.H** (or *infile*_**H** in VMS) the output decompressed image is in *infile*.

BUGS

Current input formats include only 16 bit, 2-dimensional images, though the extension to other formats and to higher dimensions would be straightforward.

FITS extensions such as multiple groups and group parameters are not accepted.

FITS header lines are not compressed when they are included in the compressed image file. This can significantly affect the compression achieved for small images with large headers. It does have the advantage that the header can be easily examined to determined the contents of the file.

If a highly inappropriate value is specified for the scale factor, the image is vastly overcompressed and much information is lost. The compression program ought to estimate the range of plausible scale factors from the image statistics and issue a warning of the user-supplied value does not appear reasonable. It would also be useful to have the program make a good guess for the scale factor if it is not specified.

If the wrong size is given (using the **-r** and **-c** parameters) for a **raw** or **net** format image, the compression program may fail to read the entire image. No error message is issued if the image is actually bigger than the specified size.

On VAXes running VMS, the decompressed image file may not have the same record format as the original image. This can cause problems if the software reading the image requires it to have special characteristics (e.g. fixed record lengths.) The **modify** program, included with the VMS command files, can be used to change the decompressed image files so that they have fixed record lengths, but it uses an unsupported VMS system routine, SYS$MODIFY, to make the modification so it is at best a makeshift solution.

AUTHOR

Richard L. White, Space Telescope Science Institute

Problems, suggestions, gripes, and praise to rlw@stsci.edu