

## A Scheme for Compressing Floating-Point Images

Richard L. White, Perry Greenfield

*Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218*

**Abstract.** While many techniques have been used to compress integer data, compressing floating-point data presents a number of additional problems. We have implemented a scheme for compressing floating-point images that is fast, robust, and automatic, that allows random access to pixels without decompressing the whole image, and that generally has a scientifically negligible effect on the noise present in the image. The compressed data are stored in an FITS binary table. Most astronomical images can be compressed by approximately a factor of 3, using conservative settings for the permitted level of changes in the data.

We intend to work with NOAO to incorporate this compression method into the IRAF image kernel, so that FITS images compressed using this scheme can be accessed transparently from IRAF applications without any explicit decompression steps. The scheme is simple, and it should be possible to include it in other FITS libraries as well.

### 1. Introduction

Our goal is to develop a compressed image format that is suitable for use as a working format — images will be stored and used directly in the compressed format. This is in contrast to most compressed file formats (e.g., gzip), where the entire file must be uncompressed before the data can be used. We also want the compressed data to be of archival quality, so that access to the original uncompressed data is not necessary for any scientific purposes. With these goals in mind, there are a number of desirable characteristics for the compression method:

***Fast in execution time.*** The time to read and decompress the data and should be comparable to (or faster than) the time to read the uncompressed data.

***Effective in compressing the data.*** It should produce compression factors competitive with the best available methods.

***Lossless (exactly reversible) for integer data.*** Lossy compression methods can produce substantially higher compression ratios. However, for most modern astronomical detectors the noise in the raw data is sufficiently low that it must be compressed losslessly for archival purposes.

***Small memory requirements.*** Any memory needed for the compression or decompression should be much smaller than the full image size.

***Random access to pixels.*** It must be possible to read a small image section located anywhere within the image without having to decompress the entire image.

For nearly all floating point astronomical images, the least significant bits of the float mantissa are essentially random and so are incompressible. Zeroing out some of these random bits corresponds to quantizing the floating point values. This leads to some additional considerations:

**Automatic setting of quantization parameters.** The appropriate quantization parameters should be determined directly from the data.

**Nearly lossless quantization.** The quantization must not significantly increase the noise in the data for scientific purposes.

**Unbiased noise properties.** Features smaller than the quantization should be recoverable from the mean of many data sets.

**Easily understandable effect on noise.** The quantization scheme should not introduce correlations in the noise in adjacent pixels. It must therefore operate on each pixel independently, which eliminates, e.g., wavelet-transform based methods like `hcompress` (White, Postman, & Lattanzi 1992; White & Percival 1994).

## 2. Compression Overview

We quantize and compress the data one row at a time. This allows random access to different parts of the image; no more than one row need be decompressed to access any pixel, and when a pixel value changes, it affects only a single row in the compressed image. It also helps avoid bias and dynamic range problems, which can result from a global quantization of the image. Finally, it makes the quantization more spatially adaptive so that it can respond to noise characteristics that change across the image.

For each row, there are 3 basic steps in the algorithm: (1) Estimate the noise level in the row. (2) Convert float pixel values to integers by linearly quantizing them at levels that are a specified fraction of the noise. (3) Losslessly compress the integer pixel values using the Rice algorithm. (Note that for integer images, the first 2 steps are not necessary.) The steps are described in more detail below.

**Noise Estimation** The rms noise  $\sigma$  in each row is robustly estimated as 1.483 times the median of the absolute value of the differences of successive pixels. If the median absolute difference is zero, then we fall back on a direct computation of the rms using a sigma-clipping algorithm to reject large outliers.

**Quantization** The separation of quantization levels is  $2^{-B}\sigma$ , where  $B$  is the user-specified number of bits of accuracy in the quantization of the noise. By specifying the quantization level in terms of  $\sigma$ , we ensure that a given value of  $B$  will produce comparable quality images regardless of the noise level.  $B$  is directly related to the compressed file size: increasing  $B$  by 1 will increase the file size by 1 bit/pixel.

**Rice Coding** We use the Rice algorithm (Rice, Yeh, & Miller 1993) to compress the integer data values. The Rice algorithm is simple and very fast, compressing or decompressing  $\sim 5 \times 10^6$  pixels/sec on modern workstations. It requires only enough memory to hold a single block of 16 or 32 pixels at a time. It codes the pixels in small blocks and so is able to adapt very quickly to changes in the input image statistics (e.g., Rice has no problem handling cosmic rays, bright stars, saturated pixels, etc.).

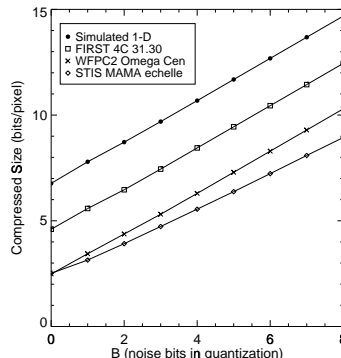


Figure 1. Compression as a function of quantization parameter  $B$  for floating point images.

For lossless compression schemes, the compressed file size depends on the nature of the image. For typical images the compressed file size is about 4 to 8 bits/pixel for  $B = 1$ ; one can thus keep 4–6 bits in the noise (which is plenty, as we show below) and still achieve a compression factor of 3 from the original 32 bit/pixel floating point image.

**Data Format** We store the compressed data in FITS binary tables. Each row of the input image is associated with a row of the table. The compressed byte stream is stored as a variable length array in the table heap. The quantization scale factors are stored as additional columns.

### 3. Performance on Float Images

For floating point images, one can effectively choose the compression factor for the data by varying the target number of quantization bits  $B$ . In this case the compressed image has slightly different pixel values than the original image, so we are concerned both with the compression factor and the degradation of the data. We assess the data quality by applying realistic data analysis procedures to the data before and after compression.

**Compression Performance** The size of the compressed data (in bits/pixel) is linearly related to the number of noise bits  $B$  used in quantization. The compression as a function of  $B$  for various images is shown in Fig. 1.

**WFPC2 Astrometry & Photometry Test** We tested the effects of quantization on astrometry and photometry using Omega Cen observations and the WFPC2 photometry script (Whitmore & Heyer 1995). The effect of quantization is much smaller than the noise (Fig. 2).

**Bias Test** To test for bias, we constructed 1-D simulated data with a weak signal buried in the noise. The “truth” spectrum includes a strong source at pixel 75 and a weak source at pixel 25. Each simulated realization of the spectrum has Gaussian noise of rms 1 count added (Fig. 3a). We averaged 1000 such spectra to recover the faint signal; the mean spectrum is shown in Fig. 3b, and the mean of coarsely quantized ( $B = 1$ ) spectra in Fig. 3c. The signal is smaller than the quantization but is recovered very accurately. The quantization method is quite unbiased and introduces almost no additional noise into the data.

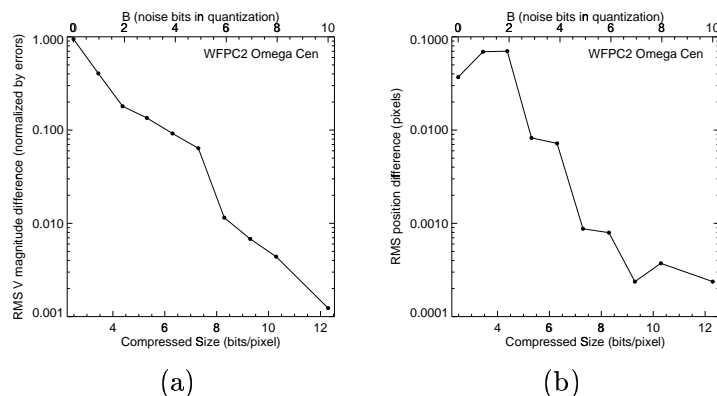


Figure 2. (a) RMS difference in magnitudes (normalized by uncertainty in the magnitudes) as a function of quantization scale factor  $B$  (*top axis*) and the compressed size (*bottom axis*). (b) RMS difference in positions as a function of quantization scale factor  $B$ .

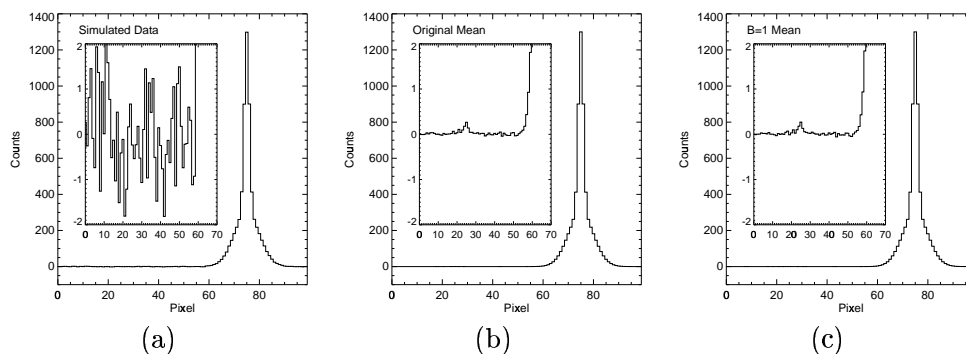


Figure 3. (a) Sample realization of simulated data. (b) Mean spectrum. (c) Mean of  $B = 1$  quantized spectra.

**Acknowledgments.** Thanks to Eric Wyckoff for his assistance in testing.

## References

- Rice, R. F., Yeh, P.-S., & Miller, W. H. 1993, in Proc. of the 9th AIAA Computing in Aerospace Conf., (AIAA-93-4541-CP), American Institute of Aeronautics and Astronautics
- White, R. L. & Percival, J. W. 1994, in SPIE Proc., Vol. 2199, Advanced Technology Optical Telescopes V, ed. L. M. Stepp, (Bellingham: SPIE), 703
- White, R. L., Postman, M., & Lattanzi, M. G. 1992, in Digitised Optical Sky Surveys, ed. H. T. MacGillivray & E. B. Thomson (Amsterdam: Kluwer), 167
- Whitmore, B. & Heyer, I. 1995, "A Demonstration Analysis Script for Performing Aperture Photometry, WFPC2 ISR 95-04," [http://www.stsci.edu/instrument-news/isr/wfpc2/9504/9504\\_1.html](http://www.stsci.edu/instrument-news/isr/wfpc2/9504/9504_1.html)