

Lossy compression of scientific images of aurora

Peter Rydesäter¹, B. Gustavsson², U. Brändström³, C-F. Enell³, and Å. Steen⁴

¹ITM, Mid Sweden University, 831 25 Östersund, Sweden

²National Institute of Polar Research, Tokyo, Japan

³Swedish Institute of Space Physics, Kiruna, Sweden

⁴RemSpace Group, Linköping, Sweden

Received: 22.10.2001 – Accepted: 20.6.2002

Abstract. Digital storage and transmission of scientific images can be demanding due to the large amounts of data. Lossless compression of data often gives insufficient compression, especially when the data consists of a considerable amount of noise. In general lossy compression of scientific data is avoided, mainly because it is poorly known how it impacts results of further analysis of the data. In this work, images of aurora observed with the Aurora Large Imaging System (ALIS) are used to study how lossy compression with JPEG impacts the data quality. The error from JPEG compression is compared to measurement noise, and we found that it is similar and of an acceptable level for many analysis applications of the data.

1 Introduction

The basic idea of lossy data compression is to keep the information of important features and drop information of less important features and in this way achieve significantly better compression than can be obtained with lossless compression. Lossy image compression with JPEG drops information in the image that is of less importance to the visible impression to the eye.

In this work we will investigate how JPEG impact on the absolute value of image pixels for aurora images and compare this to already existing noise from the measurement process.

For this work we will use images from Aurora Large Imaging System (ALIS) (Steen, 1989; Steen et al., 1990; Steen and Brändström, 1992, 1993; Brändström and Steen, 1994; Steen et al., 1997; Gustavsson, 2000). The images and the measurement noise characteristics that we use origin from the ALIS CCD cameras. Maximum resolution for the ALIS camera is 1024x1024 pixels where each pixel value is quantified in 16 *bits*. It has an Instrument noise level of 5 – 10 *counts*, that is smaller than typical random fluctuation

from photon counting that has normal distributed characteristics (Gustavsson, 1998), $p(I) = N(I, \sqrt{I})$ where I is the expected pixel intensity and the variance is \sqrt{I} . The total noise in an image pixel with intensity I is thus $N(0, \sqrt{I+10})$.

2 Used images for evaluation

In this work the images in figure 1 of aurora emission at 5577 Å are used for illustration. Image-A (256 × 256 *pixels*, 16 *bits/pixel*) is far from saturation. It has been exposed with maximum intensity at about 2% of saturation. Image-B (256 × 256 *pixels*, 16 *bits/pixel*) with maximum intensity close to saturation.

The images are preprocessed by replacing “hot” pixels and removing the dark current level determined from the covered extra pixels at each line of the CCD.

Signal level is about 50 times higher for image-B than image-A. This gives a signal-to-noise ratio at about $\sqrt{50}$ times better for image-B because the noise level is proportional to square root of intensity.

3 Lossy compression with JPEG applied to the aurora images

In this study JPEG (JPEG; Wallace, 1991) is used as the lossy compression algorithm and the CJPEG and DJPEG tools (Lane, 1998) are used for performing the compression. As the ALIS images are quantified in 16 *bits/pixel* and JPEG only supports 8 *bits/pixel*, the images must first be transformed to 8 *bits/pixel*. Typically it is found that a few pixels are extremely bright, caused by: cosmic rays, stars and noise, and these peak values are in most cases of little interest. To reduce truncation errors, the range of interest is limited to not include the intensity of the 0.1% pixels with the highest intensity.

The processing steps are as follows:

1. To avoid waste of precision in the “to 8-bit transformation” due to stretching of the intensity range to cover

Correspondence to: Peter.Rydesater@mh.se

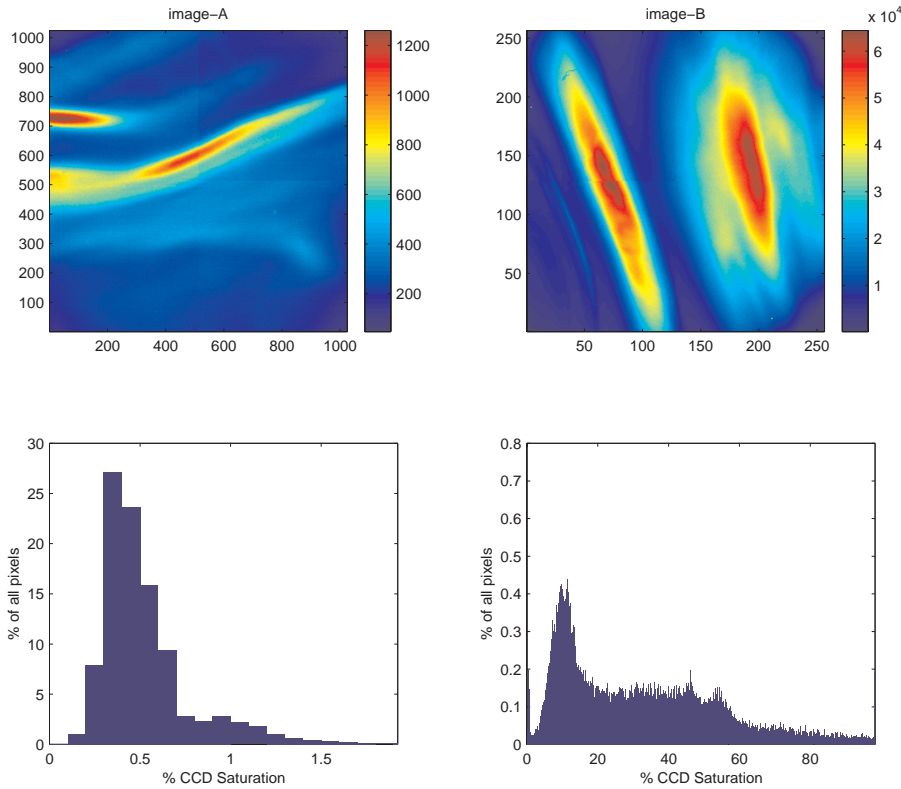


Fig. 1. Used images viewed pseudo colored. Below are the corresponding histograms. 1% saturation corresponds to 655 counts (left), 50% corresponds to 32767 counts (right).

some few bright star pixels, we limit the intensity range to cover all intensities but the 0.1% brightest pixels.

2. Calculate and store the scaling factor from the detected range to 8 bits/pixel. Then scale the 16 bits/pixel image to 8 bits/pixel.
3. Use CJPEG to lossy compress the image. The options for floating-point DCT method and optimization of entropy encoding are used to achieve the best quality/size-of-file ratio possible.
4. Compare the file size of the lossy compressed image with the original image and calculate compression ratio as compressed-file-size/16-bits-data-file-size.
5. Use DJPEG to decompress the image with the option for floating-point DCT method turned on.
6. Multiply with the scaling factors to restore original intensity range. Now the image intensities are truncated to 256 fixed levels in that range.
7. Calculate the difference from this image and the original image.
8. Calculate an error histogram over the difference. (See next section)

4 Error from JPEG-compression compared to measurement noise

To analyze the errors from JPEG-compression, error histograms in figure 2 are created as follows:

1. The absolute value of the error for each pixel is divided by the truncation-level (top of used intensity range) to get relative errors.
2. The range 0-5% are separated in 100 bins and the absolute relative errors are counted into their relative bins. All errors greater than 5% are counted in the last bin.
3. All bins are divided by the total number of pixels to get the relative number of pixels corresponding to each bin.
4. The truncation to 8 bits/pixel causes a nearly equally distributed error between about 0 – 0.2% (8-bit + gzip). For comparison, lossless compression of this with gzip (Gailly and Adler, 1999) gives 33.7% file size compared to the original 16 bits representation.
5. The black lines are a histogram of error caused by the measurement noise and the stochastic process of discrete photon counting. As the noise is impossible to separate from the images, noise images are artificially generated from the noise characteristic equation of the cameras.

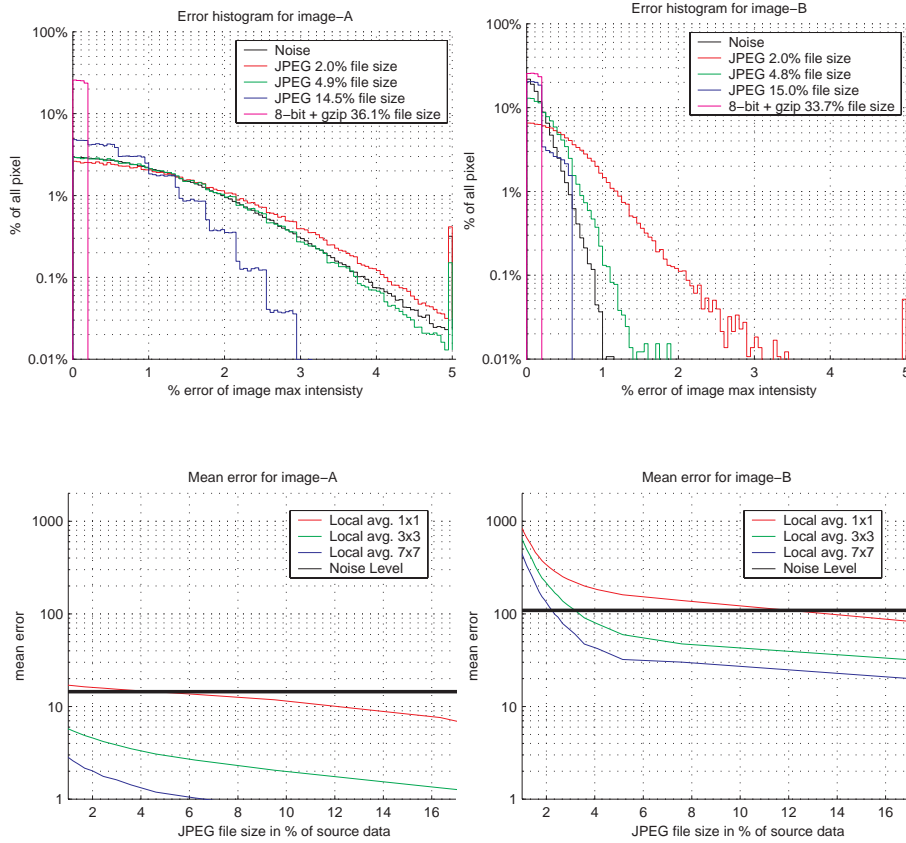


Fig. 2. Error from truncation of the range-of-interest in 16 *bits/pixel* representation to 8 *bits/pixel* representation causes a nearly equal distributed error between about 0 – 0.2% as illustrated. Lossy JPEG compression of the 8 *bits/pixel* representation to about 2%, 5% and 15% of the original 16 *bits/pixel* non-compressed image size gives the illustrated error histograms.

Fig. 3. Mean error for JPEG compressed images as function of file size. Red line is mean error for individual pixels, green line is mean error for a 3×3 *pixels* local average, and blue is a 7×7 *pixels* local average. The black line indicates the error level due to measurement noise.

5 Mean error as function of file size

In figure 3 error values are calculated for each compression ratio and image. Compression ratio is calculated as compressed-file-size/16-bits-data-file-size. The error values are calculated as the mean of absolute value of error for all pixels and are expressed as

$$e_t = \frac{\sum_{1 \leq n \leq N_I} |\hat{I}_n - I_n|}{N_I} \quad (1)$$

where I_n is pixel values before lossy compression, \hat{I}_n is decompressed data and N_I are the number of pixels. The noise level is calculated similar to this but with $|\hat{I}_n - I_n|$ replaced by absolute value of noise distribution function, $|N(0, \sqrt{T+10})|$.

The absolute noise level in image-A is less than image-B, but relative to the signal level, image-A has larger noise.

How errors from lossy compression are related to an already existing error from measurement noise (The red curve and black line). Image-A has about equal error level from measurement noise and lossy compression at 5% compression ratio, image-B at about 12% compression ratio. How the error will decrease if we look at local average intensities over 3×3 or 7×7 *pixels* (green and blue curves).

6 Summary and discussion

For future use of measurement data files, it is important to use simple and well known data formats in which to store and document the measurement results. JPEG is not simple but it is well known, widely used and supported which in all makes it a good choice.

JPEG 2000 (JPEG2000) is a new lossy compression standard with improved compression based on wavelet transform. It also has direct support for 16 *bits/pixel* data. For future work, it would be interesting to evaluate how well suited it is for the aurora images. Unfortunately today the support for JPEG 2000 is poor and its future is still uncertain.

With lossy compression such as JPEG the information is first analysed, and from some criteria is then chosen how to truncate and thereby save as much important information as possible. In the measurement process we always lose some information due to limited resolution and measurement noise. This means that the measurement process itself by nature is “lossy”.

Now when we look upon the measurement process and the data compression from the same point, we show in this work that errors from lossy compression can be kept at an “acceptable” level, as well as errors from measurement noise and limited resolution can be kept at acceptable levels. Particularly if its impact on the data is known, and weak in relation to errors from measurement noise.

From this work with JPEG-compression, the following conclusions can be made:

- Its possible to lossy compress down to $\sim 5\%$ – $\sim 12\%$ of original file size, with equal or less errors than already exist in the data from measurements.
- The local average of intensity over 3×3 or 7×7 pixels is significantly less affected by lossy compression. This implies that if analysis of the data uses local average values the error caused by lossy compression will be of less significance than if individual pixel values is used.
- Image-A, with low saturation and low signal-to-noise ratio, has larger errors from compression than image-B. This can be explained by: Low signal-to-noise ratio in image-A gives a high noise level in the 8 bits/pixel representation which causes JPEG compression to truncate the noisy pixel values to reach down to the specified compression ratio. This truncation then ends up as a higher error level.
- We come to the conclusion that it is possible to use lossy compression of ALIS measurement data which induces errors (noise) that are small compared to the measurement noise, provided that we know the intensity and characteristic of the measurement noise.

If this is true for all ALIS images is not verified. But the representative images chosen for this study cover the wide intensity range encountered with ALIS imaging and should give the correct qualitative relations between compression and induced errors.

References

- Brändström, U. and Steen, Å., Alis – a new ground-based facility for auroral imaging in northern scandinavia, in *11th ESA Symposium on 'European Rocket and Balloon Programmes and Related Research'*, Montreux, Switzerland, 1994.
- Gailly, J.-L. and Adler, M., gzip version 1.2.4, Free Software Foundation (GNU), 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA, file compression tool common in UNIX operating systems, 1999.
- Gustavsson, B., Tomographic inversion for alis noise and resolution, *J. Geophys. Res.*, pp. 26,621–626,632, 1998.
- Gustavsson, B., *Three dimensional imaging of aurora and airglow*, Ph.D. thesis, Swedish Institute for Space Physics, Kiruna, 2000.
- JPEG, Fcd 14495, lossless and near-lossless coding of continuous tone still images (jpeg-ls), ISO/IEC JTC1/SC29 WG1 (JPEG/JBIG), 1997.
- JPEG2000, Jpeg 2000, part i final committee draft version 1.0, ISO/IEC JTC1/SC29 WG1, JPEG 2000 Editor Martin Boliek, Co-editors Charilaos Chrisopoulos, and Eric Majani, 1997.
- Lane, T. G., Independent jpeg group's cjpeg, version 6b, Free software, 1998.
- Steen, Å., Alis - an auroral large imaging system in northern scandinavia, in *Ninth ESA/PAC Symposium on 'European Rocket and Balloon Programmes and Related Research'*, edited by Lahnstein, vol. ESA SP-291, 1989.
- Steen, Å. and Brändström, U., Requirements on ccd-detectors for ground-based auroral imaging, in *Proceedings of ab ESA Symposium on Photon Detectors for Space Instrumentation*, 1992.
- Steen, Å. and Brändström, U., Alis a multi-station ground-based imaging system at high latitudes, *STEP International Newsletter*, 1993.
- Steen, Å., Brändström, U., and Kaila, K., A scientific and technical description of alis, in *NSSR Annual Meeting, Bolkesjö, Norway*, 1990.
- Steen, Å., Brändström, U., and Gustavsson, B., Alis – a multi-station imaging system at high latitude with multi-disciplinary scientific objectives, in *13th ESA Symposium on 'European Rocket and Balloon Programmes and Related Research'*, Öland, Sweden, 1997.
- Wallace, G. K., The jpeg still picture compression standard, *Communications of the ACM*, 34, 30–44, 1991.