Maximizing Science Return: Data Compression and Onboard Science Processing

Aaron Kiely, Roberto Manduchi, and Matthew Klimesh

Introduction

In the pursuit of deep space exploration we are constantly trying to increase the science data returned from each mission. As we strive to squeeze every dB out of the deep space communication link, it's important to keep in mind that maximizing the science return also requires both efficient data compression and intelligent data selection. That is, we must exploit onboard processing power to optimize both *how* data are transmitted, as well as *what* data are transmitted.

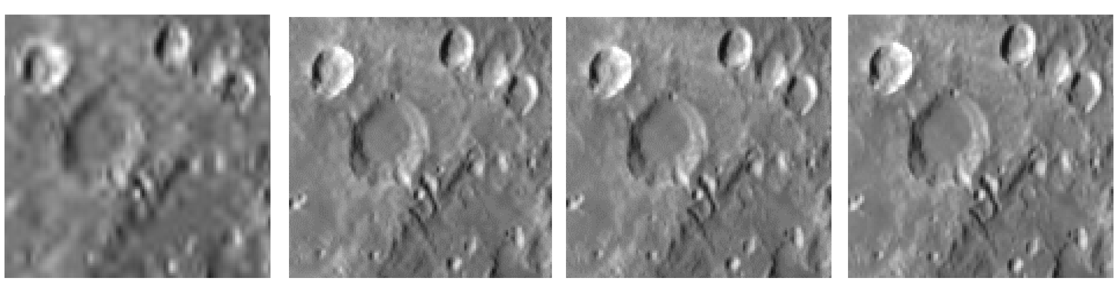
In the first part of this article, we'll get a glimpse of modern spacecraft data compression techniques. The second part of the article discusses how future missions will benefit even further when data compression is combined with intelligent onboard science processing methods. Our work thus integrates more conventional compression techniques with a science processing and buffer management component to yield a system that increases the science return of a mission.

Spacecraft Data Compression

It is clear that missions benefit when data compression techniques are used, but it may not be obvious *how much* of an improvement can be offered. Any gain obtained by employing data compression can be measured in dB—for example, data compression that provides a 2:1 compression ratio achieves the same benefit as a 3 dB increase in transmitter power. Figure 1 shows the savings in dB when the ICER image compression algorithm is used to compress a sample image.

With the recent emphasis on faster, better, cheaper missions, the use of commercial off-the-shelf (COTS) compression techniques might seem like a good idea. However, there are several reasons why the compression algorithm that works well on your web browser might not be so successful on a spacecraft:

- *Complexity*: Spacecraft processors are typically doing a lot more than just data compression, so low-complexity compression algorithms are essential. For typical Internet applications, the goal is fast decompression, while compression speed is not so important. On a spacecraft, the opposite is true.
- *Error Containment*: Without error containment techniques designed for the deep space link, a single bit error can corrupt large segments of data. Error containment schemes are not frequently used in common compression algorithms because terrestrial networks are able to use simple retransmission protocols to accommodate packet losses. Also, the packetization scheme used in deep space missions is different than that used on the Internet, and error containment strategies tailored to this structure can be expected to give better performance.



15.5 dB savings ***9.0 dB savings*** ***6.0 dB savings*** ***2.3 dB savings (lossless compression)***

Figure 1. This detail from a larger image shows an example of the effective link performance improvement offered by the ICER progressive wavelet image compression algorithm at various quality levels

• *User Requirements:* Generic data compression algorithms may not be suitable for scientific applications. The type of data, required quality, and computational resources available for spacecraft applications may be quite different from commercial scenarios. The compression algorithm that seems to do a good job compressing pictures of Aunt Bessie may not be so suitable for compressing pictures of the surface of Mars.

For example, the emerging JPEG-2000 image compression standard incorporates region-of-interest coding and accommodates low-memory implementations; however, it is not a low-complexity algorithm and it is not suitable for data types such as star field images or one-dimensional data.

One of the relatively recent trends in image compression is the use of progressive compression techniques. In a progressive algorithm, when any fraction of the compressed data has been received, we can reconstruct the entire image at nearly the best quality and resolution possible for that amount of data. Under the same circumstances, a non-progressive algorithm would only be able to reconstruct a fraction of the image. The ICER algorithm (a wavelet-based image compression algorithm developed in the Science Processing and Information Management work area) is an example of a progressive compression algorithm for deep space missions.

The use of progressive compression techniques allows for simpler and more effective use of downlink resources. Using non-progressive methods, users must adjust algorithm parameters to attempt to compress an image to the desired size. With a progressive algorithm a user can allocate space for an image and be assured that the resulting image quality is nearly optimum, subject to the allocation.

The ICER algorithm also incorporates an effective error containment scheme that exploits the progressive nature of the algorithm. For error containment, a compression algorithm must partition the data into segments and compress each segment independently of the others so that loss of data during the transmission of one segment does not affect the other segments. One method of doing this for image compression would be to treat an image as a collection of several smaller independent images that are compressed separately. Instead, the ICER algorithm incorporates a more sophisticated technique that does not segment image data until after wavelet transforms have been performed. The error containment technique used also trades rate between segments, so that regions of the image that are more difficult to compress are allocated more bits, resulting in improved compression and a uniform overall image. This also eliminates edge artifacts that would be seen under the more conventional approach—in the absence of channel errors, the error containment segmentation will be imperceptible to the viewer. The benefits of progressive transmission can also be seen under this error containment strategy. Figure 2 shows how an image might look after three segments have been affected by packet losses. Because the technique is progressive, even in the segments where much data is lost, low or medium quality versions of these segments can be reproduced.

A modified version of the ICER algorithm is also being produced that allows an onboard processing algorithm to specify regions of the image that should have higher priority, and therefore be produced with higher fidelity.

Science Processing and Buffer Management

While data compression plays a significant role in enabling deep space missions to maximize science data return over the constrained downlink channel, even the most sophisticated data compression methods become inadequate when the ability to collect sensor data on a spacecraft greatly exceeds the data transmission capability. Conversely, onboard data storage and processing capabilities are becoming increasingly affordable.

When a mission can collect much more data than can be transmitted to Earth, we must ask what part of the data should be downlinked, and what part should be discarded? We attempt to answer this question by:

- Defining a metric for the relative importance of gathered information; this is implemented by a suitable onboard science processing module;
- Making the best use of limited onboard memory resources via a prioritized buffer management mechanism which ensures that the more important data segments are transmitted to Earth with higher quality.

We can extend the idea of progressive transmission by incorporating semantic value to the “importance” attributes of encoded data. In particular, we define a simple measure of *science return* based on information-theoretic considerations as well as on the estimated *scientific value* of the transmitted data. The task of determining the relative scientific importance of segments of

CONTINUED ON NEXT PAGE

CONTINUED FROM PAGE 13

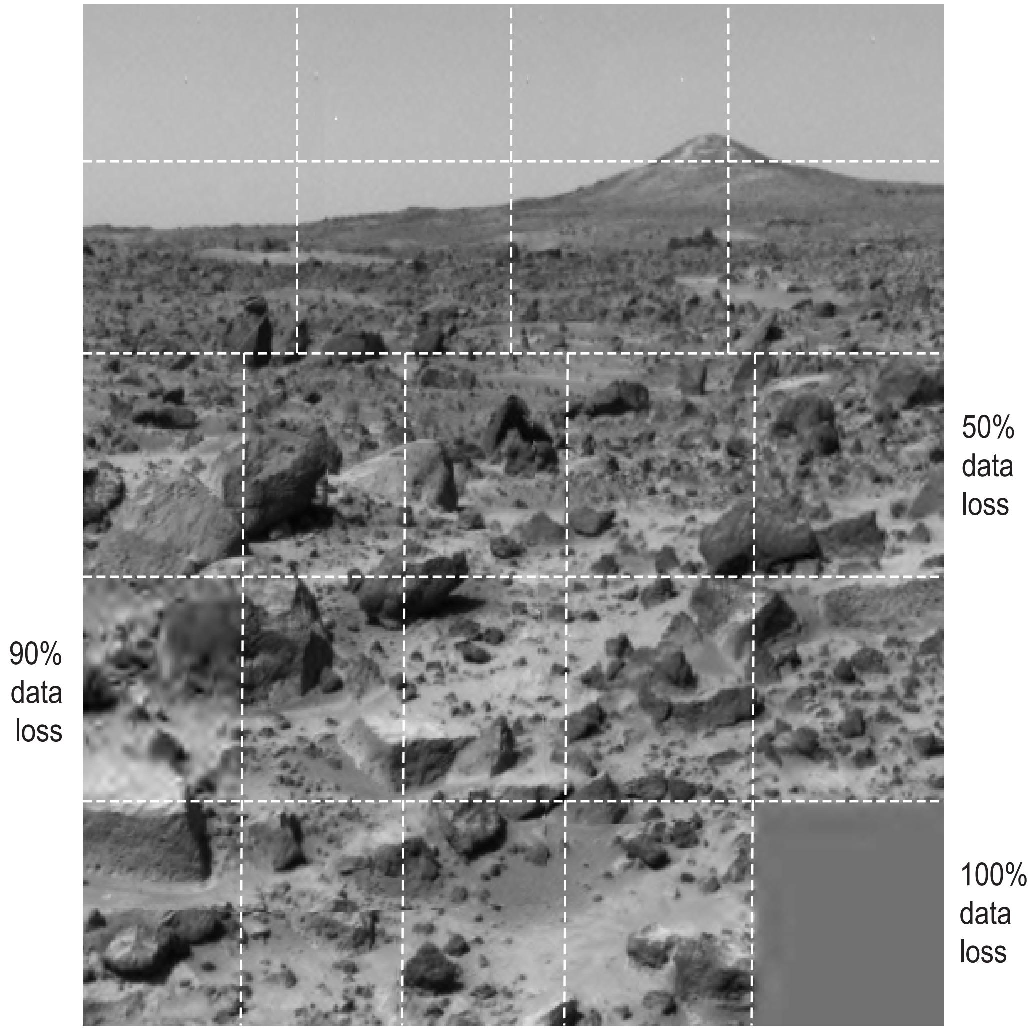


Figure 2: A 512 x 512 image segmented into 23 regions (shown by dashed lines) for error containment purposes and compressed to 2 bits/pixel. Three simulated packet losses produce varying amounts of loss in three regions of the image

data is carried out by an *onboard science processing* module, designed according to guidelines provided by the remote user (the science community). This module pre-processes the image and provides input to the progressive encoder in the form of a suitable classification map. By combining the semantic characterization produced by the science processing module with the content-blind data organization criteria of a traditional progressive encoder, we obtain a new measure of the importance of each segment of data. Algorithms that best utilize the available resources can thus be designed, and their performances assessed, in terms of science return.

The function of the science processing module is thus to assign priority values to the different parts of the acquired data. For example, in Figure 3 we show a sequence of images (first row) acquired at distinct times by a rover in a hypothetical experiment (the images were actually taken at the JPL Mars Yard). For this simulation, we assume a very simple set of science priority rules: “rocks” are more important than “soil”; and within the class “rocks” there are three level of priority: “basalt”, “obsidian”, and “shadow” (where there is very little textural or photometric information). It is understood that this is just one example of the possible prioritization rules that can be chosen—it will be up to the final user (the scientists) to decide on the most suitable science processing mechanisms and prioritization. In the second row of Figure 3 we show the results of image segmentation into the different selected classes operated by a statistical color-based classifier. The parts of the image labeled with black (basalt) have highest priority; those labeled with light gray (soil) have the least priority. The resulting image segmentation is used to drive a suitable region-of-interest (ROI) progressive compression algorithm such as our ROI extension of the ICER algorithm.

The priority information is used in our intelligent buffer management scheme. Our strategy is to allow the spacecraft/rover to take as many images as possible, and to transmit as much important information as allowed by the onboard and communication resources. In other words, we assume that the onboard buffer is constantly full and overflowing (any data packet which overflows is lost forever). Given the priority information produced by the science processing module, our buffer management scheme ensures that the high priority data packets in the buffer will be transmitted before the low priority ones, and that only the lowest priority data packets are discarded due to overflow. This strategy automatically adapts to the dynamics of the acquisition process, optimally exploiting the onboard resources. For example, in Figure 3 (third row) we show the received decoded data assuming that the downlink data rate was five times less than the acquisition data rate. The "important" image areas (rocks) are transmitted with much higher fidelity than the "uninteresting" parts (soil). In the fourth row we show the received data using a system *without* science prioritization. The amount of data received and the encoding mechanism are the same; in this case, however, a large number of bits is used to describe the "uninteresting" soil, and a much smaller number of bits are left to describe the "important" parts (rocks).

In our science processing scenario, scientists have less low-level control over the choice and quality of images returned, so it is important to verify that the automatic decisions can be made with sufficient accuracy. Related concerns apply when compression is used without science processing, unless the compression is lossless. With this idea in mind, a skeptical scientist once asked whether the serendipitous discovery of an erupting volcano on Io in Voyager images would have occurred if lossy compression had been in use. Compressed images showed that even at a ratio of 70:1, the volcano was clearly visible, which seemed like good news to the scientist—even at this high compression ratio, it was still possible to make significant discoveries. But, in fact, the news is even better—with a 70:1 compression we can transmit 70 times more images, maning scientists are 70 times *more* likely to make such discoveries!

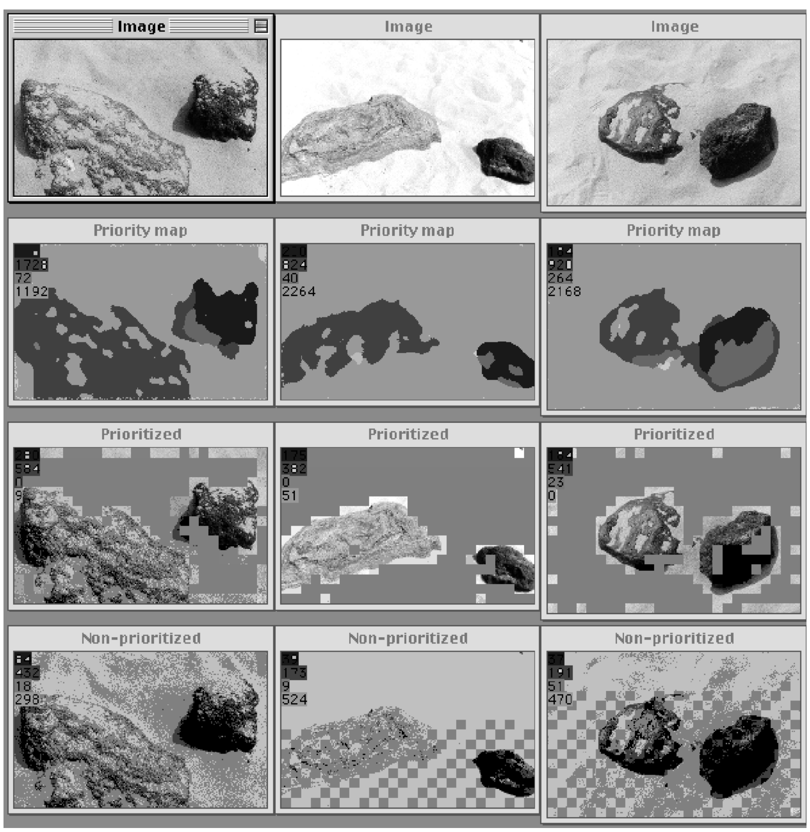


Figure 3: A simulation of science-prioritized data transmission: First row, original acquired data; second row, science-classified data; third row, data transmitted* with *our science-directed prioritization mechanism; fourth row, data transmitted* without *our science prioritization. The amount of data transmitted is the same in the prioritized and non-prioritized cases