IN-32
189384
11P

NASA Technical Memorandum 106362
AIAA–93–4670

Transform Coding for Space Applications

Daniel Glover
Lewis Research Center
Cleveland, Ohio

Prepared for the
Computing in Aerospace 9 Conference
sponsored by the American Institute of Aeronautics and Astronautics
San Diego, California, October 19–21, 1993

TRANSFORM CODING FOR SPACE APPLICATIONS

Daniel Glover
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, OH 44135

Abstract

Data compression coding requirements for aerospace applications differ somewhat from the compression requirements for entertainment systems. On the one hand, entertainment applications are bit rate driven with the goal of getting the best quality possible with a given bandwidth. Science applications are quality driven with the goal of getting the lowest bit rate for a given level of reconstruction quality. In the past, the required quality level has been nothing less than perfect allowing only the use of lossless compression methods (if that). With the advent of better, faster, cheaper missions, an opportunity has arisen for lossy data compression methods to find a use in science applications as requirements for perfect quality reconstruction runs into cost constraints.

This paper presents a review of the data compression problem from the space application perspective. Transform coding techniques are described and some simple, integer transforms are presented. The application of these transforms to space-based data compression problems is discussed. Integer transforms have an advantage over conventional transforms in computational complexity. Space applications are different from broadcast or entertainment in that it is desirable to have a simple encoder (in space) and tolerate a more complicated decoder (on the ground) rather than vice versa. Energy compaction with new transforms are compared with the Walsh-Hadamard (WHT), Discrete Cosine (DCT), and Integer Cosine (ICT) transforms.

Introduction

In this new era of better, faster, cheaper projects, scientists are being forced to consider smaller, more focused investigations. Hard choices have to be made concerning what instruments to fly and what data to acquire. Data compression is a tool that can help scientists acquire high priority data. It provides options that allow optimization of limited resources. The use of data compression on scientific missions requires a paradigm shift comparable to that of moving from large, expensive missions that bring back all the data that could possibly be of use to smaller, cheaper missions with highly focused requirements.

Data compression technology is a valuable tool for improving the science data return from future space experiments. Image data is especially voluminous and anything that can be done to cut down on the number of bits needed to represent an image has an impact on the whole communications chain.

Image Compression Approaches

There are two general areas of image data compression: lossy and lossless. In lossless compression the image is reconstructed perfectly with no loss of data. Lossless methods are capable of on the order of 2:1 compression (original:compressed). Lossless coding removes redundant information to get compression. In lossy compression, some data is thrown out to allow higher compression ratios to be obtained. It is very difficult to compress noise, so better compression can be obtained if the noise can be lost. The trick is to lose only the noise and not any important information. Lossy coding is capable of on the order of 50:1 compression. Lossy compression removes redundant and irrelevant information.

Compression ratios are not meaningful by themselves except when discussing lossless compression. For lossy compression, some quality measure is also needed to characterize the coding performance. Quality measures for image data are largely subjective since there is no good model of the human visual system that would permit quantitative measures to be used. Signal-to-noise ratio or mean-square-error are typically used as quantitative measures of image quality. Quality measures for science data need to be developed based on the particular goals of an investigation. It may be that for a particular experiment where a certain type of feature is important, that a good image is not an aesthetically pleasing picture.

Several "standard" compression techniques have been developed in the past few years including JPEG, MPEG, H.261, etc. These techniques are general purpose, designed from the start to provide acceptable compression to the widest possible set of users. The techniques are combinations of data compression methods with some user selectable variables to allow tailoring to a particular application. However, because they are "standard," there is a limit to the flexiblilty possible since the underlying methods are fixed.

One of the most common image compression methods is JPEG which is named after the Joint Photographic Experts Group committee that developed the standard. The method consists of a combination of the 8x8 discrete cosine transform (DCT) with quantization and entropy coding. The method is an intraframe technique (used on individual still images). The user can trade the amount of compression with the quality of the reconstructed image. Example quantization tables are given in the standard, but application specific quantization tables developed by the user give the best results. The standard also suggests possible post-processing for removing blocking artifacts.

The Moving Pictures Experts Group developed a compression method for motion pictures that has also come to be known by the name of the committee that developed the standard: MPEG. This method uses motion compensation to provide greater compression using the interframe correlation between pixels in the same location. The motion compensation operates on 16 x 16 pixel "macroblocks." Some reference frames, called "I frames" for intraframe, are needed to give the motion picture stream a starting point and to recover from errors. MPEG uses the DCT to compress intraframe and prediction error data. MPEG grew out of work by the Experts Group on Video Telephony in the CCITT which produced Recommendation H.261. This compression method is also known as px64 because it uses from 1 to many channels of 64k bits/sec (where p is the number of Integrated Service Digital Network B channels).

A new standard being developed is the JBIG standard for bilevel image data. It turns out that JBIG is also a promising technique for color image data when applied in bit planes. Problems in using the patented "Q coder" in the JBIG standard may soon be overcome.

The current standards are intended for use in entertainment or teleconferencing applications. They do a good job for general purpose viewing, but may not be acceptable for many scientific applications.

Image Data Compression for Space

Space-based data compression systems have peculiar requirements, obstacles, and constraints compared with applications such as entertainment. Current research in data compression is aimed at multimedia, high definition television, videoconferencing, etc. There are many different compression approaches that work quite well on the applications for which they were designed. There are also good general purpose compression methods that do a good job on a wide variety of applications. Space-based compression can take advantage of these solutions, but there is a need for tailoring or developing new, specific methods for space.

Entertainment applications are concerned with obtaining aesthetically pleasing pictures with a high degree of compression. The introduction of artifacts and distortion is not a problem if the noise is not objectionable to the average viewer. Science requirements are generally much different from aesthetics. A scientist may be interested in the motion of an edge and may not care about how the rest of the image looks. Scientists are also generally not tolerant of any introduction of artifacts or distortion to their data. The compression problem is very different depending on the application. Most research today is conducted with color video applications in mind having typical frame rates of 30 frames/sec.

One difference between space-based compression and entertainment approaches is the need for simple algorithms on the transmission side. In broadcast-type applications, the data is compressed on the transmission side without limitations on the complexity or time required to perform the compression. This is because the data need only be processed once then distributed to many receivers. The decoder must be relatively simple to keep costs down and to perform real-time processing. With millions of receivers, the cost of a decoder is significant. Space-based systems have the opposite complexity relationship. The desire is to keep the transmission side simple, reliable, and inexpensive while the decoder can be as complex as needed. The processors available on spacecraft are usually not state-of-the-art and are limited in computational capability. Ground computers on the other hand can be supercomputers if need be.

Space-based data systems have limited storage capability. Mass storage is often handled by magnetic tape recorders which cannot provide random access to the data. Start/stop cycles are limited to preserve tape life, which leads to a requirement for sequential processing of the data. On-board computers also have limited random access memory which puts a further restriction on data processing. All of these considerations lead to the need for fast, simple data compression algorithms for use on-board spacecraft.

Another difference is the data that is being compressed. In entertainment applications the quality of the image can be determined by looking at it. The goal is an aesthetically pleasing reconstruction of the original image. For science data, an aesthetically pleasing image may have lost the data of interest to the scientist. Scientists prefer to use lossless compression, if they use any at all, largely because they don't know what they are looking for. If you choose a lossy method that throws out certain features that are assumed to be irrelevant, then you will never be able to look for those features in the future if they suddenly become important.

Science data is often put to multiple uses by large teams with conflicting requirements, while entertainment images are typically single purpose.

Remote sensing systems have special requirements for calibration and fault detection that are not addressed by general-purpose compression methods. It is a good idea to keep track of pixels with saturated or null values as well as the average scene brightness for fault identification.

After the data is compressed, it is prepared for transmission by the telemetry system. Channel coding is used to detect and correct errors in transmission. It may be advantageous to combine source and channel coding so that more important information has better error protection. The data compression method has to be robust to channel errors that cannot be corrected, especially burst errors. Entertainment applications are much more tolerant of errors than scientific applications, so the data compression requirements need to take performance in the presence of errors into account. Entertainment applications can use error concealment techniques to compensate for objectionable errors while scientists do not want to work with fake data.

Lander imaging offers some advantages in data compression over other space-based systems. For a stationary lander, image registration problems are not as severe as for orbiting or fly-by spacecraft. The ability to produce images of the same landscape from the same vantage point allows differential encoding which records the difference between the two images taken at different times. Rather than sending two complete images back, differential coding would send one reference image followed by difference data. When the scene changes sufficiently, or after a period of time, a new reference image would be sent. Imaging the same landscape repeatedly also offers the advantage of cropping the image to remove unneeded areas such as the lander structure. Fly-by images taken in low light conditions require long exposures on the order of a minute. For a color image, three exposures are required through three different filters. The object being imaged will not be located in the same place or be the same size in the separate component images making registration between the three components difficult.

MESUR Pathfinder has baselined the use of data compression for its imaging system. The use of JPEG for image data compression is attractive because of the tight schedule for development. JPEG is a good general purpose compression method and can be tailored somewhat to particular applications. The main modification would be the use of custom quantization tables. Post-processing can also be used to remove blocking artifacts. MESUR Pathfinder will probably use a low compression ratio setting of around 10:1. Modifications to JPEG to use integer transforms in place of the DCT could result in a reduction in computational complexity.

The Cassini project is including JPEG hardware to provide image compression. The Pluto Fast Flyby project, which is in the conceptual design stage, is baselineing lossless data compression to provide enough storage space for the encounter data. It is also looking at using a progressive resolution scheme for transmission of the imaging science data.

The Galileo S-band Mission will be using image data compression to compensate for the lower data rate due to the high gain antenna anomaly. Compression software will be developed and uplinked to the flight computer on-board the Galileo spacecraft. The method that will be used for image data is the integer discrete cosine transform (ICT). This transform is similar to the DCT, but uses only integer values in the transform matrix for ease of computation. Deep space applications suffer from low light conditions, high sensor noise (especially in high radiation environments), and noisy communication channels. The images are very noisy to begin with, making them difficult to compress, and communication errors are a regular occurance.

Mars Observer was designed with a 16x16 discrete cosine transform compression scheme for image data. A valuable lesson was learned from this project in error containment with compressed data. The trade-off between compression efficiency and error susceptibility was made with some assumptions about the types of errors that would be encountered. The actual errors turned out to be somewhat different, resulting in much larger data loss. Since the lossy compression was implemented in software, changes to the system were possible to mitigate errors.

Both Mars Observer and Galileo were designed with hardware implementations of lossless compression along with a way of bypassing the compression. The capability to bypass the lossless compression simplified workarounds to unforseen problems that could be handled with software. Lossy compression was implemented in software on both spacecraft.

Subband/Transform Coding

A promising strategy for image data coding that has emerged in the literature in the last few years is subband coding. This approach entails the splitting of the image data into several frequency bands which can then be optimally encoded using methods tailored to the spectral characteristics of the individual subbands, the lowest of which is a low resolution version of the original image. The low band is a fraction of the size of the original image. It can be made as small as desired depending on the number of bands into which the image is split. The high frequency bands contain details such as edge information, and can be thought of as the high resolution components of the overall image.

Transform coding has been around longer than subband coding, but can be thought of as a subset of subband coding.

Subband/Transform coding has inherent advantages for remote sensing applications, both in space-based compression and in ground-based processing. The low frequency band can be used for quick look data in a space-based system or for browsing in a data archiving application. There is no sense in taking the time or trouble to transmit or acquire data that is not useful, for example a remote sensing image of Earth that is obscured by cloud cover. The use of 64 uniform or 10 octave bands can provide a thumbnail image less than 2% of the original image size. The subbanding can be made lossless by using perfect reconstruction filters or transforms. The low band provides inherent scaling of image resolution. The size of the low band can be varied depending on the number of stages of subbanding.

Because the transforms can be implemented with adders and shifts (no multipliers and accumulators are needed), the subband hardware required is simple allowing real-time, space-based compression. After an image is subbanded, straightforward compression coding can be applied to the bands. The subbands can be coded separately for compression using techniques geared towards the characteristics of the particular subband. The use of cascading tree structures provides the capability to split an image into many subbands using the same simple hardware. The use of octave-band trees is especially useful for progressive resolution reconstruction. Subbanding provides multiresolution and multirate possibilities by selectively recombining bands for image reconstruction.

The well-known Walsh-Hadamard Transform (WHT, also known as the Discrete Hadamard Transform) makes use of the Hadamard matrix (H) for 2x2 blocks[1] (ignoring a scaling factor of $\frac{1}{\sqrt{2}}$ for simplicity):

$$H(1) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (1)$$

An implementation of the WHT in 2x2 blocks can also be viewed as a subband analysis bank using separable, two-dimensional filters. If all the low frequency terms (DC coefficients) are collected in one group, the low-horizontal/low-vertical frequency subband is formed. The other subbands are likewise formed by grouping results by frequency band. Collecting one result from each of the four outputs and grouping them together (maintaining the block organization) is equivalent to a block transform. One advantage to organizing the results as subbands is that the low band is a good low resolution representation of the original image. The higher frequency bands contain edge information. A typical two-dimensional block transform would operate on a block (submatrix) of data values, **D**, with a transform matrix, **T**, to give coefficients, **C**, as follows[1]:

where **T'** is the transpose of **T**. The inverse transform is:

$$C = T * D * T' \quad (2)$$

$$D = t' * C * t \quad (3)$$

Cascading the 2x2 subbanding in a uniform band tree structure is also equivalent to performing larger size block transforms using Kronecker product expansions of the matrix. For example, a 2x2 WHT of data that has already been processed by a 2x2 WHT and organized into subbands is equivalent to performing a 4x4 WHT on the original image. This can easily be shown by comparing the 16 permutations of the Kronecker tensor product of the basis pictures of the 2x2 WHT with the basis pictures of the 4x4 WHT. If the first stage transform coefficients are not organized as subbands, but are left in the same block structure, applying the WHT again results in the original data (because the transform matrix is the same for both the forward and inverse case).

The Kronecker product (also direct product or tensor product) is the operation which creates a larger matrix from two smaller matrices by using the product of the components of one matrix with the other matrix as submatrices. For example, for the Hadamard matrix[1]:

$$H(1) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (4)$$

$$H(2) = \frac{1}{\sqrt{2}} \begin{bmatrix} H(1) & H(1) \\ H(1) & -H(1) \end{bmatrix} \quad (5)$$

giving the 4x4 matrix:

$$H(2) = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad (6)$$

This matrix can be rearranged into the following format by moving the second column to the fourth column position:

$$WHT = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad (7)$$

The 4x4 Discrete Cosine Transform (DCT) matrix is:

$$DCT \approx \frac{1}{2} \times \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1.307 & 0.541 & -0.541 & -1.307 \\ 1 & -1 & -1 & 1 \\ 0.541 & -1.307 & 1.307 & -0.541 \end{bmatrix} \quad (8)$$

Another 4x4 matrix which is derived from the discrete cosine transform, but has only integer elements is the Integer Cosine Transform (ICT)[2]:

$$ICT(4x4) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad (9)$$

A similar integer transform matrix, suggested by comparing the WHT and ICT, is:

$$G4T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 1 & -1 & -4 \\ 1 & -1 & -1 & 1 \\ 1 & -4 & 4 & -1 \end{bmatrix} \quad (10)$$

Another transform matrix that more closely approximates the DCT is suggested by comparing the ICT and DCT. The Approximate Cosine Transform (ACT) uses elements that are factors of 1/2 instead of factors of 2:

$$ACT = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1.5 & 0.5 & -0.5 & -1.5 \\ 1 & -1 & -1 & 1 \\ 0.5 & -1.5 & 1.5 & -0.5 \end{bmatrix} \quad (11)$$

The apparent advantage of the ICT over the DCT is the use of integer elements which simplifies the computations needed to calculate the transform. Multiplication by 2 is equivalent to a shift left, so it should be an easier matter to design a digital circuit to perform the ICT. It should also be faster for a simple processor to evaluate the ICT. There are fast algorithms that minimize the number of multiply-add operations for several transforms, but they are not necessarily simple algorithms. Fast algorithms show their efficiency at larger block sizes, require more storage, and have a complex internal organization[1].

The ACT and G4T also benefit from having elements that are factors of 2 or 1/2. The WHT, with matrix elements of only 1 or -1, has been around for many years and has fast algorithms. Despite the apparent advantage of the simpler transform matrices over the DCT in computation, the DCT is the standard transform in use today. There are commercial chips available which allow direct implementation of the DCT (in the 8x8 version, no less). The place where the simple transforms might have their place is in applications where simple algorithms or extremely fast implementations are needed. The space application is one where simplicity is key. Often, space electronics lag the commercial state of the art by more than 10 years[3]. It is much easier to find adders and shifters on the standard parts list than DCT or JPEG chips. The computational power of on-board computers is also less than is commercially available on the ground.

The question that remains is: What is lost by going to a simpler transform over the DCT? Table I shows a comparison of the peak signal-to-noise ratio
(PSNR) of the reconstructed data for the image "Lenna" with only three transform coefficients kept out of sixteen (at full accuracy). This is an indication of the energy compaction for the transforms. Software functions for performing these calculations are available[4].

The results for the simple transforms fall between the WHT and the DCT. It should be remembered, however, that numerical metrics for image data are not well developed, and the accuracy of the PSNR numbers presented here is misleading in that a difference of a fraction of a dB may not be meaningful. This simple comparison indicates that further work on simple transforms may be fruitful.

Concluding Remarks

Data compression is often thought of by scientists as something that will corrupt their data. In a system with hard limits on data rates, compression is a solution that will allow greater acquisition of useful data. In a future of limited resources, compression is a tool which must be considered.

TABLE I
"Lenna" Reconstructed from Three Coefficients of 4X4 Transform

Transform	PSNR (dB)
WHT	30.76
G4T	31.85
ICT	31.95
ACT	31.98
DCT	32.01

References

1. Clarke, R. J.: Transform Coding of Images. London: Academic Press, 1985.

2. Cheung, K.-M. and K. Tong: "Proposed Data Compression Schemes for the Galileo S-Band Contingency Mission." In: NASA Conference Publication 3191, 1993

Space and Earth Science Data Compression Workshop, Snowbird, Utah, April 2, 1993, pp. 99-109.

3. Glover, D.: "Design Considerations for Space Flight Hardware." NASA TM-102300, January 1990.

4. Glover, D.: "Subband/Transform Functions for Image Processing." NASA TM-106183, June 1993.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
October 1993

3. REPORT TYPE AND DATES COVERED
Technical Memorandum

4. TITLE AND SUBTITLE
Transform Coding for Space Applications

5. FUNDING NUMBERS
WU–144–10–10

6. AUTHOR(S)
Daniel Glover

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135–3191

8. PERFORMING ORGANIZATION REPORT NUMBER
E–8160

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, D.C. 20546–0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA TM–106362
AIAA–93–4670

11. SUPPLEMENTARY NOTES
Prepared for the Computing in Aerospace 9 Conference sponsored by the American Institute of Aeronautics and Astronautics, San Diego, California, October 19–21, 1993. Responsible person, Daniel Glover , (216) 433–2847.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified - Unlimited
Subject Category 32

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 words)***

Data compression coding requirements for aerospace applications differ somewhat from the compression requirements for entertainment systems. On the one hand, entertainment applications are bit rate driven with the goal of getting the best quality possible with a given bandwidth. Science applications are quality driven with the goal of getting the lowest bit rate for a given level of reconstruction quality. In the past, the required quality level has been nothing less than perfect allowing only the use of lossless compression methods (if that). With the advent of better, faster, cheaper missions, an opportunity has arisen for lossy data compression methods to find a use in science applications as requirements for perfect quality reconstruction runs into cost constraints. This paper presents a review of the data compression problem from the space application perspective. Transform coding techniques are described and some simple, integer transforms are presented. The application of these transforms to space-based data compression problems is discussed. Integer transforms have an advantage over conventional transforms in computational complexity. Space applications are different from broadcast or entertainment in that it is desirable to have a simple encoder (in space) and tolerate a more complicated decoder (on the ground) rather than vice versa. Energy compaction with new transforms are compared with the Walsh-Hadamard (WHT), Discrete Cosine (DCT), and Integer Cosine (ICT) transforms.

14. SUBJECT TERMS
Data compression; Spacecraft

15. NUMBER OF PAGES
10

16. PRICE CODE
A02

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	