Low Complexity Efficient Raw SAR Data Compression

Shantanu Rane\textsuperscript{a}, Petros Boufounos\textsuperscript{a}, Anthony Vetro\textsuperscript{a} and Yu Okada\textsuperscript{b}

\textsuperscript{a}Mitsubishi Electric Research Laboratories, Cambridge, Massachusetts, USA;
\textsuperscript{b}Mitsubishi Electric Space Information Systems Department, Kamakura, Kanagawa, Japan

ABSTRACT

We present a low-complexity method for compression of raw Synthetic Aperture Radar (SAR) data. Raw SAR data is typically acquired using a satellite or airborne platform without sufficient computational capabilities to process the data and generate a SAR image on-board. Hence, the raw data needs to be compressed and transmitted to the ground station, where SAR image formation can be carried out. To perform low-complexity compression, our method uses 1-dimensional transforms, followed by quantization and entropy coding. In contrast to previous approaches, which send uncompressed or Huffman-coded bits, we achieve more efficient entropy coding using an arithmetic coder that responds to a continuously updated probability distribution. We present experimental results on compression of raw Ku-SAR data. In those we evaluate the effect of the length of the transform on compression performance and demonstrate the advantages of the proposed framework over a state-of-the-art low complexity scheme called Block Adaptive Quantization (BAQ).

Keywords: discrete cosine transform (DCT), transform coding, entropy coding, compression, raw SAR data

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is a well established imaging technology that has become critical for remote sensing and observation applications. SAR exploits the motion of a moving platform, such as a plane or a satellite, to create a synthetic aperture much larger than the real aperture of the antenna carried by the moving platform. This allows SAR systems to image very large land swaths with extremely high precision. Furthermore, using two different platform trajectories and interferometric analysis it is even possible to acquire height information for the imaged area.

The success of SAR has resulted to a demand for ever increasing resolution and fidelity from the imaged area. One of the main challenges in satisfying this demand is coping with the amount of data acquired and processed. To manage this data deluge, it is necessary to develop very effective compression algorithms that can efficiently and effectively compress this data. A significant consideration for such algorithms is their computational complexity, as they are required to operate using the limited hardware on-board a satellite.

A important challenge with typical satellite-borne SAR data compression is that the acquired data is not processed on the satellite, but transmitted to a ground station for processing. The reason is that the data processing step, called image formation, is very expensive computationally to be performed on-board. On the other hand, image formation exposes significant structure on the data that makes compression a much simpler task. Instead, the raw acquired SAR data transmitted from the satellite to the ground exhibit much less exploitable structure, and appear uncorrelated at first glance. Furthermore, the distortion metric of interest is not the distortion of the raw data after compression, but the distortion of the reconstructed image after decompression and image formation, making rate-distortion control and evaluation much more difficult.

In this paper we develop an efficient yet lightweight compression framework for raw SAR data compression. Our method is based on transform coding using the Discrete Cosine Transform (DCT) to expose some of the data structure followed by Arithmetic Coding (AC) to exploit that structure in compression. Our approach is sufficiently lightweight to fit on two FPGA chips with specifications typical for space applications. Using simulations we show that our approach has comparable or better performance with state-of-the-art compression schemes with significantly increased complexity. We further show that some light-weight signal processing, such

Send correspondence to \{rane,petrosb,avetro\}@merl.com or Okada.Yu@eb.mitsubishielectric.co.jp
as low-pass filtering, on-board the satellite can expose more data structure and further increase compression performance compared with conventional approaches.

This paper is organized as follows. In the next section we provide a brief background and an overview of existing raw SAR data compression approaches. Section 3 presents our compression framework and its possible adaptations. An experimental evaluation is presented in Sec. 4. We discuss our results and conclude in Sec. 5.

2. BACKGROUND

2.1 SAR Data Acquisition and Image Formation

SAR systems image the area of interest by transmitting pulses as they move and recording their reflection (echo) from the imaged ground. The received echos comprise the raw SAR data that should be processed to produce a complex-valued image of the ground area. The pulses are transmitted at a uniform rate, the pulse repetition frequency (PRF) and they typically are linear frequency modulated (FM) chirps of a certain bandwidth. The received echos are acquired by sampling at the Nyquist rate corresponding to their bandwidth or slightly greater to avoid aliasing due to the leakage of the anti-aliasing filter.

As the mobile platform moves in a straight line it defines a natural coordinate frame for the ground image. The direction in the line of motion of the platform is referred to as azimuth direction, and the direction across is referred to as range direction. The azimuth direction is effectively sampled at the pulsing rate, with each received echo corresponding to one azimuth position of the mobile platform along its track. The PRF is dictated by system parameters, such as the beamwidth, which also dictate the azimuth resolution. Increasing the resolution for the same amount of ground coverage necessarily increases the PRF. The range resolution is determined by the bandwidth of the transmitted pulse, which also determines the sampling frequency of the received echo. Thus, higher resolution SAR requires higher PRF and higher sampling rate for the echo acquisition. For example, doubling the resolution in both directions results to a four times data increase.

The structure of the acquisition process also defines a natural representation of the raw SAR data. Generally, the raw data are represented in a data matrix, with each row containing the samples of a single received echo and all the echoes stacked in the sequence in which they are received. Thus, each row of the matrix corresponds to an azimuth location. Similarly to the ground image, the vertical direction of this matrix (i.e., the row number) is also referred to as the azimuth direction, while the horizontal direction (i.e., the column number) is also referred to as range direction.

Typical satellite SAR systems process the echos as they acquire them and immediately transmit them to the ground station. On-board memory is at a premium, so buffering large amounts of data is undesirable. Thus, it is very difficult to process more than a part of one echo at a time, which does not allow us to exploit structure and correlation in the azimuth direction.

Often some light signal pre-processing is performed on the mobile platform before the data is transmitted to the ground. A very common step is de-chirping, which demodulates the received echo by a linear chirp. This step changes the characteristics of the data, and, as we discuss later, also improves the compression performance.

After the data are transmitted to the ground, they are processed using one of the many available image formation algorithms. These algorithms invert the linear acquisition system implied by the pulsing and reception model to reconstruct a complex-valued ground image. The best known algorithms are the Range-Doppler Algorithm (RDA) and the Chirp Scaling Algorithm (CSA), with the latter being more efficient and, therefore, more popular.\(^1\)

Both algorithms use pulse compression as a fundamental step in they reconstruction. The basis of pulse compression is that convolution of a chirp with its complex conjugate inverts the chirp and produces an impulse, i.e., compresses the chirp pulse to an impulse that has the same energy as the pulse. This method allows to transmit higher energy to the ground, which improves the signal-to-noise ratio, without using high instantaneous power, by extending a lower power pulse for a longer time.

Both the phase and the magnitude of the reconstructed complex image are of interest. The magnitude represents the reflectivity of the ground features, i.e., the ground image. The phase is useful in interferometric applications, in which the ground is imaged from two different tracks (either by the same or two different platforms) and the phase is used to determine the elevation of other features of the imaged area.
2.2 Related Work on Raw SAR Data Compression

Raw SAR data is especially challenging to compress because of its statistics. Typically, the inphase and quadrature components follow a nearly Gaussian distribution, have low inter-pixel correlation in both the range and azimuth directions, and have a low correlation between the inphase and quadrature components. However, the standard deviation of the data changes slowly in both range and azimuth directions, and this fact has been exploited in raw SAR data compression.

The state-of-the-art in raw SAR data compression is driven by the requirement of high compression ratio at low complexity. The first well-known method to achieve a noteworthy reduction in the data rate of raw SAR data was Block-Adaptive Quantization (BAQ) which was used in the NASA Magellan mission to Venus. In this method blocks of raw SAR data are processed with a given set of quantization parameters and the quantization parameters are updated from block to block. First, the standard deviation of the raw data pixel values is determined and used to adjust a Lloyd-Max or a uniform quantizer and obtain the optimal quantization thresholds. Then, quantization is carried out, and the encoder transmits the quantization indices and the standard deviation for that block, so that the decoder can use the appropriate inverse quantizer. In each block of raw data pixels, the in-phase and quadrature components are processed separately, though there are some works in which the data is converted into polar co-ordinates prior to performing BAQ.

The original BAQ methods operated in the pixel domain and, therefore, did not leverage the benefits of energy compaction provided by transform coding. Combining frequency-domain methods with BAQ increases the on-board encoding complexity, but provides gains in compression performance. Further improvements have been demonstrated by first transforming the data into the frequency domain and then applying an entropy constrained block-adaptive quantizer. In addition to having blocks that are large enough to obtain an accurate estimate of the standard deviation, these frequency domain techniques have explored the option of using 2-dimensional transforms. Some methods developed for ENVISAT SAR data exploited 1-dimensional transforms only, while other methods assume that future methods will have enough memory and computational resources to consider transforms blocks as large as 32 × 32 samples. In this work, we take the low-complexity frequency domain approach, with the aim of constructing a scheme that would run on essentially the same hardware as a conventional (pixel-domain) BAQ system. However, while we do modulate the quantization parameters in response to the standard deviation of the transform blocks, the scheme differs significantly from frequency-domain BAQ in two aspects: (1) the bit rate can change during the compression process in response to a rate control algorithm (2) efficient entropy coding is performed using an arithmetic encoder.

Another advantage of using frequency-domain techniques is the option of performing filtering at the on-board module. Since the useful frequency range is known apriori, it is possible to filter out the frequency components outside this range. Before inverting the transform at the decoder, these frequency coefficients outside the useful band are replaced by zeros. This method provides compression gains, with respect to BAQ.

Some methods apply standardized image compression codecs to raw SAR data compression, which can take advantage of efficient and optimized hardware implementations. For example, a JPEG2000-based scheme has been used for separate compression of the inphase and quadrature components of the raw SAR data. However, in the evaluation of the rate-distortion performance, the distortion is computed on the raw SAR data itself rather than on the actual image formed by applying the RDA or CSA algorithm. Hence, the effect of JPEG2000 on the quality of the formed image remains unexplored. In our work, we do consider elements that appear in standardized image codes, such as the discrete cosine transform (1-dimension as opposed to the 2-dimensional transform used in JPEG), quantization and arithmetic coding. However, different from the past work, our distortions are computed using the formed image rather than directly using the raw SAR data.

3. COMPRESSION OF RAW SAR DATA

3.1 Statistical Properties of Raw SAR Data

In this section we describe the architecture of the on board compression module, which is depicted in Fig. 1. As explained in the background section, visualize the raw SAR data as a 2-dimensional matrix of complex values, with each row representing the range signal for a particular azimuth position. The raw data is read into the compression module row by row. We consider the complex-valued raw data as being composed of two
Figure 1. Encoder architecture. In the proposed encoder, inphase and quadrature images are transformed using a 1-dimensional DCT, and then uniformly quantized. Then, several blocks of quantized transform coefficients are compressed into a bit stream by an arithmetic coder. In parallel, the quantized transform coefficients are used to update the probability distribution that will be utilized by the arithmetic coder during subsequent entropy coding operations.

separate raw data images, the inphase component and the quadrature component. Histograms of the inphase and quadrature components are nearly identical, shaped roughly like a Gaussian distribution with zero mean, and nearly equal variance. Therefore, in what follows, identical operations are performed separately on the inphase and quadrature components.

3.2 Signal Transforms
In order to compress the signal we transform it to a domain in which signal correlation can be exploited by the subsequent compression steps. To do so, we first apply a 1-dimensional DCT to each component of the raw data. In order to retain roughly the same hardware complexity as that incurred by BAQ-based architectures, we avoid 2-dimensional block transforms, which would require buffering multiple rows of a very large matrix, and therefore significant amount of memory and processing power. The length of the transform is a variable parameter in our experiments. The longer the transform, the greater is the resolution achievable in the frequency domain, the higher is the computational complexity, and the greater is the buffering requirement. In our work we have considered transforms of length 8, 16, 32, 64 and 128. Even though the samples appear uncorrelated at first glance, some energy compaction is observed in the transform coefficients. i.e., some coefficient magnitudes are significantly larger than others. This energy compaction is not nearly as pronounced as that seen in natural images, but the discrepancy in coefficient magnitudes hints at the fact that an entropy coder can compress them more efficiently than uniformly distributed transform coefficients.

Very often, raw SAR data are sampled slightly higher than the Nyquist rate to ensure the acquired data are not aliased even if the anti-aliasing filter is not very tight. While this helps the acquisition process, it also increases the data rate with data that do not contribute in image formation. However, we can exploit this knowledge to further reduce the data rate at this stage. This can be achieved by removing the high-frequency coefficients of the DCT before transmitting the transformed block, and reintroducing them at the decompression stage by setting them to zero. This effectively low-pass filters the acquired data, albeit with a very non-ideal low-pass filter. The number of high-pass coefficients to remove should depend on the size of the DCT and the amount of oversampling of the acquisition system. The coefficients removed should contain no useful signal components.

3.3 Uniform quantization
The next step in the compression process is uniform quantization of the transform coefficients. This results in quantization indices which can be fed to the entropy coder. The initial quantization step size, referred to as the “base QP”, is chosen based on the target bit budget. As compression proceeds, the base QP is altered in two ways before encoding any block of transform coefficients.

First, the base QP is altered depending on the variance of the block of pixels before the transform is computed. The rationale behind this is that, if the variance is too low, there will be a large number of very small DCT coefficients in the transformed block. These would be quantized to zero, and significant details in the raw data would be lost. Thus, for coefficient blocks with low variance in the pixel domain, the base QP is reduced by a specified amount. On the other hand, if the variance is high, the absolute value of many of the quantization indices
obtained after quantizing the DCT coefficients will be large. Since large index values have a low probability, they generate a large number of bits after entropy coding, which might result in buffer overflow. Thus, for coefficient blocks with a high variance in the pixel domain, the base QP is reduced by a specified amount. This notion of modulating the base QP based on the variance of the raw data pixels is borrowed from traditional Block Adaptive Quantization (BAQ). The thresholds to determine a variance value that is “too high” or “too low” and the corresponding amounts by which the base QP is updated, are determined from extensive experimentation with raw data. These thresholds are sensitive to the frequency ranges encountered, the SAR acquisition technique and equipment.

Second, the base QP is altered for the purpose of rate control. At regular intervals, i.e., after a fixed number of transform coefficient blocks, the base QP is altered slightly in response to the bit rate consumed while encoding the previous block of transform coefficients. In our implementation, after each block of transform coefficients is quantized and entropy coded, the bit rate consumed is recorded. If, after a specified number of transform blocks, the output bit rate is less than the target bit rate, the quantization step size is reduced by 1. On the other hand, if the total bit rate consumed so far is greater than the target bit rate, the quantization step size is increased by 1. This modulation of the quantization step size ensures that the consumed bit rate remains close to the target bit rate throughout the compression process. While encoding blocks of transform coefficients, the value of the selected quantization step size is also entropy coded using the arithmetic coder and transmitted as part of the compressed bit stream.

3.4 Efficient Entropy Coding

The final step in the compression process is entropy coding of the quantization indices. This is accomplished by an arithmetic encoder.\textsuperscript{13} Arithmetic coding is used because of its compression efficiency compared with Huffman coding, even though it is significantly more complex to implement. When provided with the correct probability distribution of the quantization indices, the arithmetic encoder compresses the indices to a bit rate that is very close to their information entropy.

The arithmetic coder maintains a separate probability distribution of the quantization indices for each DCT coefficient. The initial distribution is generated from training data and provided to the arithmetic encoder. As compression proceeds and new data is gathered, the probability tables are constantly updated. Thus, quantized coefficients in a given row (or a set of rows) are used to update the probability distribution that will be used to compress the coefficients in the next row (or the next set of rows). This ensures causality at the ground station; since the decoder at the ground station also begins with the same initial probability distribution, and updates the probability distribution from the recently decoded DCT coefficients, encoder-decoder synchronization is maintained at no additional signaling cost for updating the distributions.

A significant amount of resources is devoted in storing and maintaining the appropriate probability distributions. In fact, in our architecture, the largest memory footprint in the FPGA is occupied by arithmetic coding modules, especially by the probability tables stored for each transform coefficient. The transform length, discussed in the previous section, severely impacts the memory use of the encoder; doubling the transform length requires double the number of tables to be stored and maintained.

3.5 Measurement of decoded image quality

At the ground station, the above compression steps are reversed to obtain the reconstructed inphase and quadrature components of the raw SAR data, as depicted in Fig. 2. Similar to conventional image compression schemes, the bit rate is simply the number of bits generated by the arithmetic encoder. However, contrary to conventional image compression schemes, the distortion is not measured between the original and reconstructed raw data. Instead, as explained in the background section, the raw data image is converted to the Single Look Complex (SLC) domain. The distortion is then measured between this formed SLC image and a similar SLC image recovered from the uncompressed raw data. The SLC images are complex valued, and have different distortion measures for the amplitude and phase, which are explained below.

Let the SLC image matrix associated with the uncompressed raw data image be denoted by $I$, and its corresponding real and imaginary parts by $X$ and $Y$ respectively. Let the size of the image be $M \times N$ pixels in
the SLC domain. Then the amplitude image \( A \) and phase image \( \phi \) are given by

\[
A(i, j) = \sqrt{(X(i, j))^2 + (Y(i, j))^2} \quad \text{and} \quad \phi(i, j) = \tan^{-1}\left(\frac{Y(i, j)}{X(i, j)}\right), \quad \forall \ i \in \{1, 2, ..., M\}, \ j \in \{1, 2, ..., N\}
\]

Similarly, let the SLC image associated with the uncompressed raw data image be denoted by \( \hat{I} \), and its corresponding real and imaginary parts by \( \hat{X} \) and \( \hat{Y} \) respectively. Then, the amplitude and phase images \( \hat{A} \) and \( \hat{\phi} \) are computed exactly as above from \( \hat{X} \) and \( \hat{Y} \). Then, compute amplitude and the phase error measures as follows:

\[
E_A(i, j) = 20 \log_{10}\left(\frac{\hat{A}(i, j)}{A(i, j)}\right) \quad \forall \ (i, j) \in \{1, 2, ..., M\} \times \{1, 2, ..., N\} \text{ for which } A(i, j) \neq 0, \hat{A}(i, j) \neq 0
\]

\[
E_{\phi}(i, j) = \phi(i, j) - \hat{\phi}(i, j)
\]

Finally, the distortion measures for amplitude and phase are given by the standard deviations of the above error measures. Thus,

\[
D_A = \text{std}\{E_A\} \quad D_{\phi} = \text{std}\{E_{\phi}\}
\]

Note that there isn’t a straightforward relation between the distortion measured between the SLC images and the corresponding distortion between the raw data, that can be exploited at low complexity inside the onboard module. While it is certainly true that highly compressed raw data will give a poor quality SLC image, it is much harder to pinpoint which DCT coefficient in the raw data ought to be quantized heavily without affecting the image quality in the SLC domain. This makes the system design and parameter selection very challenging.

4. EXPERIMENTAL EVALUATION OF RATE-DISTORTION PERFORMANCE

As described in Section 3, the process of evaluating the compression algorithm is fundamentally different from that used for natural images. Our system compresses the raw SAR data, while we are actually interested in the quality of the formed SAR image. Furthermore, the magnitude and the phase of the complex-valued image are used differently, and our evaluation takes this into account using different amplitude and phase distortion metrics described earlier. Again, this differs from simple mean-squared-error metrics used for traditional image compression. Our experiments are based on raw Ku-SAR data which contains 4096 range samples per row, and 2080 azimuth positions. Parameters of the Ku-band SAR data are provided in Table 1.

Specifically, to evaluate the performance of our system we measure distortion after forming the SAR image from the compressed raw SAR data. The conversion from the reconstructed (respectively uncompressed) raw data to the corresponding reconstructed (respectively uncompressed) SAR data is performed using the CSA algorithm. The formed SAR image is also referred to as a SLC image. The distortion is measured with respect to the SAR image formed from the uncompressed raw data, which we consider as the ground truth.

Figs. 3 and 4 provide the rate-distortion performance of our scheme for two Ku-SAR datasets. The evaluation is performed for multiple transform distortions, with and without the Low Pass Filtering operation discussed in
Section 3. In this experiment, the quantization step-size was held constant and the updating procedures described in Section 3 were not used. This was done in order to isolate the impact of low-pass filtering on the compression performance. In particular, the quantization step sizes used were 5, 10, 15, 20, 25, 30, and 35. Further, on all plots, the performance of pixel-domain Block Adaptive Quantization (BAQ) is shown for settings that give a compressed file size that is 25% and 50% of the original data size. All other datasets that we tested show similar trends. Note that, because of the way the distortion metrics are defined, lower values indicate better performance. We make the following observations.

1. The transform coding framework provides superior performance to BAQ. The performance benefit increases as longer transforms are used.

2. At a given quantization step-size, low pass filtering applied to the raw data reduces the bit rate required but incurs higher distortion in the reconstructed SAR image. It is still more efficient than BAQ for long transform sizes, and can achieve compressed data sizes much lower than 12.5% of the original size, which is the smallest BAQ is capable of. However, it is inefficient in a rate-distortion sense, when compared against direct compression without low-pass filtering.

As explained in Section 3, the quantization step size is modulated according to the variance of a block of raw data pixels. Additionally, after a constant number of blocks, the quantization step size is altered depending upon whether the bit rate consumed at given stage of decoding is greater than or less than the target bit rate desired. The effect of modulating the quantization parameters is that the rate distortion curves are not smoothly varying, as in the constant step size case of Figs. 3 and 4.

Figure 5 plots the amplitude and phase distortion for the case of changing quantization step-size, along with the respective data points for BAQ for comparison. In addition, we show the ideal performance curve, in which the quantization step size is held constant and the bit rate is just the calculated entropy of the quantization indices. The difference between the two curves captures the compression inefficiency which results from the fact that quantization indices resulting from a new quantization step-size have to be compressed with the arithmetic coder that has a sub-optimal initial distribution. If the quantization step-size remains constant for a large number of blocks, the stored initial distribution of quantization indices gets updated, resembles the true distribution more closely, and improves the compression efficiency of the arithmetic encoder. Thus, the interval (measured in the number of transform coding blocks) after which the quantization step size is varied, and the interval after which the probability distribution of the quantization indices is changed, should be properly selected to achieve the best possible compression performance.

5. SUMMARY AND DISCUSSION

We presented an efficient, low-complexity algorithm for compression of raw SAR data. This algorithm is based on the discrete cosine transform (DCT), uniform quantization and parallel arithmetic coding, and can be implemented with a small onboard FPGA footprint. The compression efficiency of the proposed method exceeds that of pixel-domain block-adaptive quantization (BAQ), the state of the art in on-board raw SAR data compression. Unlike BAQ, which is limited to providing compressed streams at integer bits per pixel rates, the
Figure 3. Rate distortion performance in the amplitude and phase portions of Ku-SAR images compressed using the proposed framework.
Figure 4. Rate distortion performance for a second image in the amplitude and phase portions of Ku-SAR images compressed using the proposed framework.
Figure 5. Rate distortion performance for the amplitude and phase components of the formed SLC image for the case in which the quantization step size is updated in response to the bit budget and the variance of the raw SAR data. These plots are for an experiment with DCT transforms of length 8; our results are similar with transforms of length 16 and 32.

proposed method can generate an arbitrary number of intermediate datapoints on the rate-distortion curve. It is observed that longer transforms increase the compression efficiency at the expense of increased storage and increased computation overhead particularly for parallel arithmetic coding. Further, low-pass filtering prior to compression potentially improves the rate-distortion performance.

As we discuss in Sec. 2, demodulation with a linear chirp might be performed by the system upon reception of the signal, a process known as de-chirping. Our preliminary experiments demonstrated that de-chirping exposes some of the signal structure which compression can exploit to improve compression performance. However, a full dechirping is expensive to perform on the airborne or space platform. Instead, we found that an approximate dechirping also helps slightly in compression. Specifically, we demodulate each block processed separately, with the same chirp, before applying the compression algorithm. This allows the use of a smaller chirp length and, therefore, has smaller memory requirements than full dechirping. On the receiver side we invert this operation; the signal is re-chirped by re-modulating with the same chirp.

Acknowledgments

The authors wish to thank Steve Burgess for his input on various aspects of the algorithm design based on implementation constraints.

REFERENCES


