MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Distributed Coding of Multispectral Images

Goukhshtein, M.; Boufounos, P.T.; Koike-Akino, T.; Draper, S.C.

TR2017-080 June 2017

Abstract

Compression of multispectal images is of great importance in an environment where resources such as computational power and memory are scarce. To that end, we propose a new extremely lowcomplexity encoding approach for compression of multispectral images, that shifts the complexity to the decoding. Our method combines principles from compressed sensing and distributed source coding. Specifically, the encoder compressively measures blocks of the band of interest and uses syndrome coding to encode the bitplanes of the measurements. The decoder has access to side information, which is used to predict the bitplanes and to decode them. The side information is also used to guide the reconstruction of the image from the decoded measurements. Our experimental results demonstrate significant improvement in the rate-distortion trade-off when compared to coding schemes with similar complexity.

IEEE International Symposium on Information Theory (ISIT)

Distributed Coding of Multispectral Images

Maxim Goukhshtein*, Petros T. Boufounos†, Toshiaki Koike-Akino†, and Stark C. Draper*
* Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada,
Email: maxim.goukhshtein@mail.utoronto.ca, stark.draper@utoronto.ca
† Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA, Email: {petrosb, koike}@merl.com

***Abstract*—Compression of multispectal images is of great importance in an environment where resources such as computational power and memory are scarce. To that end, we propose a new extremely low-complexity encoding approach for compression of multispectral images, that shifts the complexity to the decoding. Our method combines principles from compressed sensing and distributed source coding. Specifically, the encoder compressively measures blocks of the band of interest and uses syndrome coding to encode the bitplanes of the measurements. The decoder has access to side information, which is used to predict the bitplanes and to decode them. The side information is also used to guide the reconstruction of the image from the decoded measurements. Our experimental results demonstrate significant improvement in the rate-distortion trade-off when compared to coding schemes with similar complexity.**

I. Introduction

Recent advances in remote sensing technology have significantly increased imaging resolution and, consequently, the amount of raw data that needs to be transmitted. As a result, there is renewed interest in lightweight data compression algorithms, e.g., for airborne platforms. In such platforms, conventional compression approaches based on transform coding, such as JPEG and JPEG2000, may not be suitable. Instead it is necessary to develop rate-efficient low-complexity encoders, shifting complexity to the decoder.

In this work we focus on high-resolution multispectal images, typically comprised of four to eight spectral bands. Inspired by [1]–[4] we present a lightweight distributed coding approach that exploits structural correlations among bands. Specifically, we assume that one band is available as side information at the decoder, along with statistical information that enable coarse prediction of the other bands.

Our approach uniquely and efficiently combines recently developed quantized compressed sensing (CS) and randomized sampling techniques with conventional distributed coding, to enable an accurate control of the prediction error and the compression rate. In addition, our approach exploits structural correlations between bands, both in decoding—by forming an accurate prediction from the side information—and in the CS-based reconstruction—by influencing the weights in the optimization. Thus, it can exploit more complex structural models of the image, beyond second order statistics, such as sparsity and structured sparsity.

In particular, encoding first linearly measures the signal using a randomized measurement matrix. The measurements are quantized and separated into bitplanes from the least to the most significant. Given that side information is available at the decoder, each bitplane is separately compressed by computing and transmitting a syndrome at the appropriate rate. The decoder decodes each bitplane iteratively, starting from the least significant and iterating to the most significant. At each iteration, a prediction of the bitplane is formed using the signal prediction and the bitplanes decoded in the previous iterations. The syndromes are used to correct the bitplane prediction, to recover the quantized measurements. In turn, these are used to reconstruct the signal using a sparse optimization informed by the side information.

The first author performed this work while at MERL.

Our method can be considered a significant generalization of [3], exhibiting strong theoretical connections with universal quantization [5], [6]. In comparison to [3], it exhibits lower complexity, while experimentally delivering up to 6dB improvement in peak signal-to-noise ratio (PSNR) for the same encoding rate. Furthermore, it enables better theoretical understanding of the coding rate required. This work also raises several interesting theoretical questions regarding the redundancy of compressive measurements, and the optimal way to exploit it.

The next section provides some background on CS and distributed coding. Section III describes the details of the proposed approach. Simulation results validating our approach are presented in Section IV. Finally, Section V discusses our results and concludes the paper.

II. Background

A. Compressed Sensing

The development of CS in the last decade has demonstrated that it is possible to undersample signals significantly and still be able to reconstruct them, as long as the reconstruction exploits information on the signal structure [7], [8]. Typically, the structure exploited is sparsity in some domain, such as a basis transform or in the signal derivative. To ensure information is preserved, incoherent linear measurements of the signal are obtained, typically randomized, of the form $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{y}$ is the measurements vector, $\mathbf{x}$ is the signal of interest, and $\mathbf{A}$ is the measurement matrix.

The signal can be recovered by solving a sparse inverse problem. One of the most useful sparsity-promoting metrics in imaging is the total variation norm, which quantifies the sparsity of the image gradient. Additional prior information about the image can be incorporated by adjusting the weights of a weighted total variation (WTV) norm. The isotropic two-dimensional (2D) WTV is defined as

$$WTV(X) = \sum_{s,t} \sqrt{W^x_{s,t}(X_{s,t} - X_{s-1,t})^2 + W^y_{s,t}(X_{s,t} - X_{s,t-1})^2}, \quad (1)$$

where X is a 2D image, W^x and W^y are 2D sets of weights, and (s,t) are image coordinates. Larger weights penalize edges more in the respective location and direction, while smaller weights reduce their significance. In abuse of notation, we use X and $\mathbf{x}$ interchangeably to refer to the same signal. The former denotes the signal arranged as an $N \times N$ 2D image, while the latter as a column vector of length $n = N \times N$.

One of the most common approaches to reconstruct the image is to solve the following regularized convex optimization:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda WTV(\mathbf{x}), \quad (2)$$

where λ is an appropriately chosen regularization parameter. There are several heuristics to set the weights in $WTV(\cdot)$. The simple approach of switching between a large and a small weight, depending on the prior information, has been shown to perform well [9].

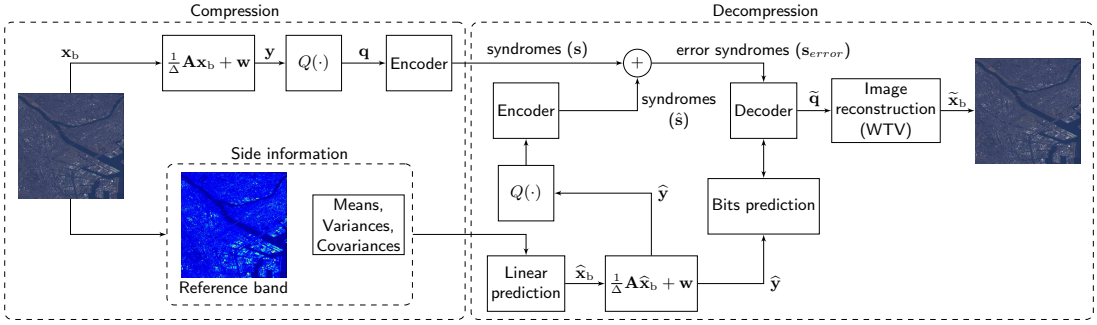

Fig. 1. System diagram for the proposed compression-decompression method.

It should be noted that, despite the reduced number of measurements, when quantization and bit rate are taken into account, quantized CS does not produce as good a compression as transform coding [10]. To improve performance, the approach we use here combines CS with distributed coding, enabling more efficient use of the available bitrate.

B. Distributed Source Coding

Distributed source coding, first developed in [11], is a source coding approach that exploits side information available at the decoder to reduce the rate required for reliable transmission. Remarkably, the encoder can optimally encode the signal without requiring access to the side information.

In the setting we encounter in this paper, a sequence of bits $\mathbf{q} \in \mathbb{F}_2^m$ should be encoded and transmitted, knowing that the decoder has access to a prediction $\widehat{\mathbf{q}} = \mathbf{q} + \mathbf{e}$ of the same sequence. In the prediction, each predicted bit may be incorrect with some probability p, i.e., the typical error vector $\mathbf{e}$ has Hamming weight pm. To encode $\mathbf{q}$, the encoder computes a syndrome using a linear code $\mathbf{s} = \mathbf{H}\mathbf{q}$, where $\mathbf{H} \in \mathbb{F}_2^{M\times m}$ is a parity check matrix [12]–[15].

Given $\widehat{\mathbf{q}}$ and $\mathbf{s}$, the decoder finds the error vector $\mathbf{e}^*$ with the minimum Hamming weight such that $\mathbf{s} = \mathbf{H}(\widehat{\mathbf{q}} + \mathbf{e}^*)$ and estimates $\mathbf{q}$ as $\overline{\mathbf{q}} = \widehat{\mathbf{q}} + \mathbf{e}^*$. Successful decoding implies that $\overline{\mathbf{q}} = \mathbf{q}$, i.e., all the errors have been corrected. The probability of an error in decoding can be made arbitrarily small as m increases, as long as the code used to generate the syndrome has rate $r < C(p) = 1 - H_\mathrm{B}(p)$, where $H_\mathrm{B}(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy for probability p and $C(\cdot)$ is the capacity of the corresponding binary symmetric channel. Equivalently, the syndrome needs to have length $M > m(1 - C(p)) = mH_\mathrm{B}(p)$. Of course, for finite blocklengths a slightly lower rate is necessary to guarantee reliable decoding.

In this work we use the most popular capacity approaching class of codes based on low-density parity-check (LDPC). The LDPC parity check matrix $\mathbf{H}$ is sparse, making it efficient to store on board a satellite. They also allow computationally efficient decoding based on belief propagation. To design the parity check matrix of irregular LDPC codes, we use the Pareto optimization approach proposed in [16], where the highest coding gain and lowest decoding complexity are achieved at the same time by analyzing the extrinsic information transfer (EXIT) [17] trajectory across decoding iterations. We consider check-concentrated triple-variable-degree LDPC codes for designing the degree distribution, while the girth is maximized in a greedy fashion by means of progressive-edge-growth (PEG) [18].

III. Proposed Approach

A. Overview

Our approach relies on compressively measuring the band to be encoded, henceforth denoted as $\mathbf{x}_\mathrm{b} \in \mathbb{R}^n$, using a randomized measurement matrix, and uniformly quantizing the measurements. We assume that the decoder has access to a reference band as side information, denoted as $\mathbf{x}_\mathrm{ref} \in \mathbb{R}^n$, as well as sufficient prediction information to obtain a good prediction $\widehat{\mathbf{x}}_\mathrm{b} \in \mathbb{R}^n$ of $\mathbf{x}_\mathrm{b}$.

A system diagram is presented in Fig. 1. In this paper we sidestep the question of how the reference band is coded. For example, a conventional lossy or lossless compression method might be used. We assume that such a reference band is available at the decoder for simple analysis. Furthermore, we assume that the decoder has access to prediction information that help the prediction of the encoded band from the reference band. In particular, as shown in the figure, we assume that the encoder, which has access to both bands, encodes and transmits linear prediction coefficients, i.e., means, variances and covariances, to the decoder. To manage complexity on-board the satellite, the image is segmented in blocks of size $n = N \times N$. Each block can be treated separately, both at encoding and at decoding, enabling massive parallelization of both steps.

A key realization, formalized and quantified in Thm. 1 in Sec. III-B, is that the quantized measurements of the encoded band can be predicted from the measurements of the image prediction. In particular, while the least significant bits of the measurements will be difficult to predict, as the significance of the bits increases, prediction becomes more reliable. Furthermore, if the first k least significant bit levels of the measurements are known or have been reliably decoded at the decoder, then the prediction of the $(k+1)$th least significant bit level becomes easier. In other words, we can encode each bitplane with a distributed source code at a rate appropriate for the reliability of the prediction. The decoder iteratively decodes each bitplane, from the least significant to the most significant, updating the prediction of the not-yet-decoded bitplanes as each bitplane is decoded.

One issue with commonly used distributed coding schemes is that it is difficult to decide what code rate to use. In particular, it is often not straightforward to quantify how prediction quality affects the encoding of the signal and the reliability of each bit of the encoding. Conveniently in our case, Thm. 1 exactly quantifies the probability that the $(k+1)$th LSB will be different between prediction and true measurement as a function of the prediction error, and assuming that the first k LSBs have been correctly decoded. Thus, the rate required for successful decoding of the code can be exactly computed if the prediction quality is known.

B. Coding

For each block, the band to be encoded, $\mathbf{x}_\mathrm{b}$, is measured, scaled, and dithered according to

$$\mathbf{y} = \frac{1}{\Delta}\mathbf{A}\mathbf{x}_\mathrm{b} + \mathbf{w}, \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^m$ are the measurements, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a measurement operator, $\Delta \in \mathbb{R}$ a scaling parameter, and $\mathbf{w} \in \mathbb{R}^m$ is a dither vector with i.i.d. elements drawn uniformly in $[0, 1)$. Typically, although not necessarily, the operator $\mathbf{A}$ is compressive, i.e., it reduces the dimensionality at the output to $m < n$.

The measurements are quantized element-wise, using a scalar uniform integer quantizer $Q(\cdot)$. The quantizer rounds its input to the nearest integer, using B bits, producing quantized measurements $\mathbf{q} = Q(\mathbf{y}) \in \mathbb{Z}^m$. We assume that the quantizer does not saturate, i.e., that B is selected sufficiently large. It should be noted that changing the scalar parameter Δ in (3) is equivalent to using unscaled measurements and setting the quantization interval to Δ.

In the remainder of this paper we use $\mathbf{q}_{(k)}, k = 1, \ldots, B$ to denote each bitplane of the measurement, from least significant to most significant. In other words, $\mathbf{q}_{(k)} \in \mathbb{F}_2^m$ is a binary vector containing the kth significant bit of all quantized measurements $\mathbf{q}$, with $k = 1$ being the LSB and $k = B$ the MSB.

Having access to the corresponding block in the reference band, $\mathbf{x}_\mathrm{ref}$, the encoder also computes prediction statistics to be transmitted. These will enable linear prediction of $\mathbf{x}_\mathrm{b}$ from $\mathbf{x}_\mathrm{ref}$ at the decoder. In particular, the following statistics are computed and transmitted separately

$$\mu_\mathrm{b} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_\mathrm{b})_i, \tag{4}$$

$$\sigma_\mathrm{b}^2 = \frac{1}{n}\sum_{i=1}^{n}\left((\mathbf{x}_\mathrm{b})_i - \mu_\mathrm{b}\right)^2, \tag{5}$$

$$\sigma_\mathrm{ref,b} = \frac{1}{n}\sum_{i=1}^{n}\left((\mathbf{x}_\mathrm{ref})_i - \mu_\mathrm{ref}\right)\left((\mathbf{x}_\mathrm{b})_i - \mu_\mathrm{b}\right). \tag{6}$$

The overhead of this transmission is small. For example, assuming 8 bits per parameter, blocks of size 64×64, each band requires 3 parameters, totaling 24 bits, i.e., 0.0059 bits per pixel (bpp). Note that the mean and variance of the reference band can be computed at the decoder and do not need to be transmitted.

The linear prediction parameters also inform the encoder of the prediction error, which is equal to

$$\epsilon^2 = \|\mathbf{x}_\mathrm{b} - \widehat{\mathbf{x}}_\mathrm{b}\|^2 = n\sigma_\epsilon^2 = n\left(\sigma_\mathrm{b}^2 - \frac{\sigma_\mathrm{ref,b}^2}{\sigma_\mathrm{ref}^2}\right), \tag{7}$$

where $\widehat{\mathbf{x}}_\mathrm{b}$ is the prediction. For each bitplane k, the linear prediction error also quantifies the probability that a measurement bit will be different when comparing the prediction to the correct measurements, according to the following theorem.

Theorem 1: Consider a signal $\mathbf{x}_\mathrm{b}$ measured using a random matrix $\mathbf{A}$ with i.i.d. $\mathcal{N}(0, \sigma^2)$ entries according to (3). Also consider its prediction $\widehat{\mathbf{x}}_\mathrm{b}$ with prediction error $\epsilon = \|\mathbf{x}_\mathrm{b} - \widehat{\mathbf{x}}_\mathrm{b}\|$ and assume that bitplanes $k = 1, \ldots, K-1$ have been correctly decoded. Then $\mathbf{q}_{(K)}$ can be estimated with probability of bit error equal to

$$p_K = \frac{1}{2} - \sum_{l=1}^{+\infty} e^{-\frac{1}{2}\left(\frac{\pi\sigma\epsilon l}{2^{K-1}\Delta}\right)^2} \operatorname{sinc}\left(\frac{l}{2}\right)\operatorname{sinc}\left(\frac{l}{2^K}\right). \tag{8}$$

The proof, which we defer to an extended version of the paper, exploits and extends techniques introduced in [5]. Furthermore, although we can bound (8) above and below using the first few terms of the summation, we omit those bounds, and discussion thereof, in the interest of space. Still, it is straightforward to compute (8) to arbitrary precision, even when computation is scarce. It can also be approximated well as a piecewise linear function.

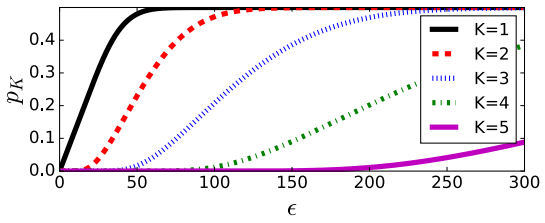

Fig. 2. Probability p_K of bit error in decoding as a function of the prediction error, assuming $K-1$ bitplanes have been correctly decoded.

Figure 2 shows how the probability behaves as a function of the prediction error, for different values of $K-1$ correctly decoded bitplanes. As evident in the figure, the range of errors for which the bitplane prediction is accurate increases rapidly with K, making syndrome coding unnecessary for large K.

For each bitplane $\mathbf{q}_{(k)}$, and given the prediction error (7), the encoder uses (8) to compute the probability p_k that each bit will be flipped in the bitplane prediction. The bitplane must therefore be encoded via the appropriately chosen parity check matrix $\mathbf{H}_{(k)}$ as the syndrome $\mathbf{s}_k = \mathbf{H}_{(k)}\mathbf{q}_{(k)}$ of length $M_\mathrm{S} > mH_\mathrm{B}(p_k)$ bits, which is transmitted to the receiver.

When implementing the encoder, there are several practical considerations. In particular, a truly Gaussian matrix $\mathbf{A}$ is not easy to implement because of storage and computation requirements, especially since the encoder is typically implemented in finite-precision arithmetic. In practice, similar to [3], [4], we use a binary ± 1 matrix, implemented by randomly permuting the signal, taking the Hadamard transform and randomly subsampling the output. Thus, we have a $\mathcal{O}(n \log n)$ complexity operator, instead of $\mathcal{O}(nm)$, which further requires only $\mathcal{O}(n)$ storage, instead of $\mathcal{O}(nm)$. In our experiments, the behavior of this ensemble is similar to the behavior of a Gaussian ensemble with respect to the probability of error in Thm. 1.

In addition, due to the complexity of computing codes of arbitrary rates on the fly, we pre-compute and store codes for a fixed number of rates in the range $r = 0.05, 0.1, \ldots, 0.95$. Note that rate 1, i.e., $p_k = 0$ means that there is no need to send any data since there is no decoding error. In contrast, for $p_k = 0.5$ we need a rate 0 code, i.e., the syndrome will have the same length as the bits to be encoded. In that case, instead of computing a syndrome we simply transmit the entire bitplane $\mathbf{q}_{(k)}$ as is.

Another practical consideration is the effect of finite blocklength, which limits the rate of the code we use to be strictly below capacity. In practice, after computing the required rate for the code, instead of selecting the next lower rate code from the available rates, we heuristically select a code rate at least 0.1 less. While more sophisticated approaches could be used to select the rate, we found that their benefit is minor in terms of the compression rate achieved, while their implementation complexity is significant.

C. Decoding

Using the side information and the reference band, the decoder first computes an estimate $\widehat{\mathbf{x}}_\mathrm{b}$ using a simple linear minimum mean squared error (LMMSE) estimator

$$\widehat{\mathbf{x}}_\mathrm{b} = \frac{\sigma_\mathrm{b,ref}}{\sigma_\mathrm{ref}^2}\left(\mathbf{x}_\mathrm{ref} - \mu_\mathrm{ref}\right) + \mu_\mathrm{b}. \tag{9}$$

The estimate is measured, scaled and dithered, as with the original data in (3), to produce predictions of the measurements $\widehat{\mathbf{y}}$. The measurement predictions will be used to decode the syndromes and to recover the quantized measurements $\mathbf{q}$ exactly.

The quantized measurements are iteratively recovered starting with the least significant bitplane $k = 1$. At iteration k, a new estimate of the quantized measurements $\widehat{\mathbf{q}}$ is computed, incorporating all the new information from the previous iterations. From that estimate, the kth bitplane, $\widehat{\mathbf{q}}_{(k)}$, is extracted and corrected using the syndrome $\mathbf{s}_k$ to recover the corrected bitplane $\widetilde{\mathbf{q}}_{(k)}$. If the syndrome has been properly designed at the correct rate, decoding is successful with high probability and $\widetilde{\mathbf{q}}_{(k)} = \mathbf{q}_{(k)}$.

In particular, for $k = 1$, the predicted measurements are quantized to $\widehat{\mathbf{q}}$. Their least significant bitplane $\widehat{\mathbf{q}}_{(1)}$ is the prediction corrected using by the syndrome $\mathbf{s}_1$. For $k > 1$, assuming $k-1$ bitplanes have been successfully decoded, $\widehat{\mathbf{q}}$ is estimated by selecting the uniform quantization interval consistent with the decoded $k-1$ bitplanes and closest to the prediction $\widehat{\mathbf{y}}$. Having correctly decoded the first $k-1$ bitplanes is equivalent to the signal being encoded with a $(k-1)$-bit universal quantizer. Thus, recovering $\widehat{\mathbf{q}}$ is the same as the decoding performed in [3].

An example of $k-1 = 2$ is shown in Fig. 3. The left hand side of the figure plots a 2-bit universal quantizer, equivalent to a uniform scalar quantizer with all but the 2 least significant bits dropped. The right hand side shows the corresponding 3-bit uniform quantizer used to produce $\mathbf{q}$. In this example, the two least significant bits decode to the universal quantization value of 1, which could correspond to $\mathbf{q}_i = 1$ or -3 in the uniform quantizer. However, the prediction of the measurements $\widehat{\mathbf{y}}_i$ is closer to the interval corresponding to $\mathbf{q}_i = -3$, and, therefore $\widetilde{\mathbf{q}}_i = -3$ is recovered.

Formally, let a temporary estimate of the yet undecoded bits be $\overline{\mathbf{q}} = Q(\widehat{\mathbf{y}})_{(k:B)}$. Then the estimate for $\mathbf{q}_i$ is

$$\widehat{\mathbf{q}}_i = 2^k(\overline{\mathbf{q}}_i + c) + \left(\widetilde{\mathbf{q}}_{(1:k-1)}\right)_i, \tag{10}$$

where $c \in \{-1, 0, 1\}$ is chosen to minimize the distance of $\widehat{\mathbf{q}}_i$ to $\widehat{\mathbf{y}}_i$.

Finally, the kth bitplane $\widehat{\mathbf{q}}_{(k)}$ is corrected by decoding the syndrome to produce the corrected estimate $\widehat{\mathbf{q}}_{(k)}$. As long as the syndrome satisfies the rate conditions of Thm. 1, the decoding is reliable. Decoding continues iteratively until all B bitplanes have been decoded. After decoding each bitplane, the next bitplanes become increasingly reliable. At some point in the decoding process the remaining bitplanes are sufficiently reliable that $\widehat{\mathbf{q}}$ will stop changing from iteration to iteration, and decoding can stop early. Note that this point is already known at the transmitter, because of Thm. 1. At this point no additional syndromes are transmitted. In our experiments, with the parameters used, typically 1 or 2 least significant bitplanes were transmitted as is, i.e., with a rate 0 syndrome code. A maximum of 3 additional bitplanes were transmitted for which syndromes were required, i.e., for which the rate was greater than 0 but less than 1.

Once all bitplanes have been successfully decoded, the quantized measurements $\widetilde{\mathbf{q}}$ are used to reconstruct the image by solving

$$\widetilde{\mathbf{x}}_{\mathrm{b}} = \arg\min_{\mathbf{x}} \left\| \widetilde{\mathbf{q}} - \frac{1}{\Delta}\mathbf{A}\mathbf{x} - \mathbf{w} \right\|_2^2 + \lambda WTV(\mathbf{x}), \tag{11}$$

where $\lambda = 0.1$ was tuned experimentally using a small part of the data. Several approaches exist to solve (11); we use the fast iterative shrinkage thresholding algorithm (FISTA)-based approach in [19].

The optimization in (11) should be guided by the reference image since the two images exhibit the same structure. Specifically, the two spectral bands image the same area and, therefore, we expect edges to be colocated. Thus, the gradient in the reference image X^{ref} can

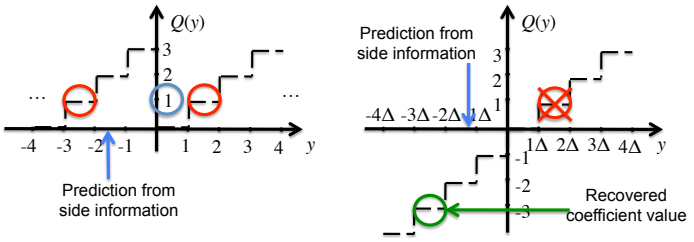

Fig. 3. Minimum distance decoding of a quantization point using the prediction measurement.

be used to set the weights in a WTV reconstruction, as described in Sec. II-A. In particular, given X^{ref}, we set the weights in (1) as

$$W^x_{s,t} = \begin{cases} 0.2, & \text{if } |X^{\mathrm{ref}}_{s,t} - X^{\mathrm{ref}}_{s-1,t}| > t, \\ 1, & \text{otherwise,} \end{cases} \tag{12}$$

$$W^y_{s,t} = \begin{cases} 0.2, & \text{if } |X^{\mathrm{ref}}_{s,t} - X^{\mathrm{ref}}_{s,t-1}| > t, \\ 1, & \text{otherwise,} \end{cases} \tag{13}$$

where t is an appropriate threshold chosen to qualify which gradients are considered significant. This choice of weights is common in a number of weighted sparsity measures. It has been shown in practice to have similar performance to more complicated schemes [9].

IV. Results

To validate our approach we used [3] as our benchmark on multispectral images acquired by the AVNIR-2 instrument of the ALOS satellite [20]. Specifically, we performed experiments on a 4-band, 512×512 crop of an image, compressed at 2bpp. The image was compressed using non-overlapping blocks of size $n = N \times N = 64 \times 64$. A total of $m = 4000$ measurements of each block were obtained for each spectral band. Each measurement was quantized to $B = 10$ bits. Similarly to [3], the blue band was chosen as the reference, $\mathbf{x}_{\mathrm{ref}}$, to compress the other three bands, green, red and infrared.

The choice of the scaling parameter Δ affects the reconstruction quality and the bit budget needed to encode an image band. In particular, reducing Δ results in finer measurement quantization which translates to improved reconstruction PSNR at the cost of higher encoding rate. In our experiments, we considered two scenarios. First, we kept Δ equal among all three bands, ensuring similar reconstruction quality but at variable rate for each band. Second, we set Δ such that all bands use the same rate but have different reconstruction quality. It is expected that the easier to predict bands will consume lower rate in the first scenario or exhibit better reconstruction quality in the second. For the benchmark approach [3], the only available results use the same rate for each band.

The results are tabulated in Table I. The top row lists the performance of simple linear prediction using the prediction parameters in the side information, quantifying the similarity of each band to the reference, blue band. As expected, the green band is easier to linearly predict from the blue, followed by the red band and then the infrared band. This also corresponds to the spectral distance of the three bands from the blue one.

As evident in the table, our approach significantly outperforms [3], especially in the more difficult to predict cases. When coding using the same rate of 2bpp for each band, our approach exhibits a 4–5dB improvement across all bands, as can be deduced from the third row of the table. However, similar to [3], and as expected, the quality decreases as the prediction performance decreases, i.e., for the red

TABLE I
DECODING PSNR AT 2 BPP (512×512 IMAGE CROP)

	PSNR (green) (dB)	PSNR (red) (dB)	PSNR (infrared) (dB)	BPP (green)	BPP (red)	BPP (infrared)	BPP (overall)
Linear prediction	33.46	28.53	27.52	—	—	—	—
Benchmark [3]	37.79	32.76	34.24	2.00	2.00	2.00	2.00
$\Delta_{green} = 7.5; \Delta_{red} = 13.1; \Delta_{infrared} = 13.35$	41.70	37.65	37.98	2.00	2.01	2.04	2.02
$\Delta_{green} = \Delta_{red} = \Delta_{infrared} = 11$	39.13	38.71	39.11	1.51	2.21	2.30	2.01

TABLE II
DECODING PSNR AT 1.68 BPP (FULL 7040×7936 IMAGE)

	PSNR (green) (dB)	PSNR (red) (dB)	PSNR (infrared) (dB)	BPP (green)	BPP (red)	BPP (infrared)	BPP (overall)
Linear prediction	37.05	31.67	27.32	—	—	—	—
Benchmark [3]	39.06	37.60	35.80	—	—	—	1.68
$\Delta_{green} = \Delta_{red} = \Delta_{infrared} = 10.35$	40.96	40.25	39.31	1.15	1.65	2.09	1.64
$\Delta_{green} = 5.6; \Delta_{red} = 10.25; \Delta_{infrared} = 15.75$	44.49	40.30	37.39	1.68	1.66	1.67	1.68

and infrared bands. On the other hand, when Δ is kept fixed for all bands, reconstruction quality also becomes approximately the same. This is shown in the fourth row of the table. Instead, the encoding rate is now variable, still averaging 2bpp, with the easier to predict green consuming a lower rate than the red, which, in turn, consumes lower rate than the infrared band. Still, reconstruction is better in all three bands, with the gain ranging from approximately 1.5dB to 6dB.

We also performed a similar test on an entire 7040×7936 aligned image from the instrument—from which the above 512×512 image crop originated—targeting compression rate of roughly 1.68bpp. The results are shown in Table II. As expected, the behavior is similar to the smaller cropped image. A common Δ for all bands leads to similar reconstruction quality at different compression rates, whereas varying Δ for each band such that the rate is the same leads to variations in reconstruction quality. As with the smaller crop, the proposed approach outperforms [3].

V. DISCUSSION

Our approach combines ideas from compressed sensing and distributed source coding, resulting in a very low-complexity, rate-efficient, source encoder. This approach is therefore well-suited for use in computationally constrained systems, such as remote sensing systems on satellites. Our experimental results demonstrate a significant improvement in terms of reconstruction quality when compared to similar lightweight approaches, such as [3].

In this paper we sidestep the question of how to encode the reference band at low complexity and defer it to a future publication. We also do not explore which band is best used as reference. Nor do we attempt to use already decoded bands to further improve prediction performance further. In general we expect that the better the prediction is, the greater the compression we will achieve, as is predicted by Thm. 1.

A key contribution of our approach is the development of the theoretical connection between prediction performance and bit error probability. However, a better understanding, especially in the case of non-Gaussian matrices, is still an open question of practical interest. Our analysis hits at strong connections between universal quantization and distributed coding, also of theoretical importance.

REFERENCES

[1] S. Rane, Y. Wang, P. Boufounos, and A. Vetro, "Wyner-Ziv coding of multispectral images for space and airborne platforms," in *Proc. Picture Coding Symposium (PCS)*. Nagoya, Japan: IEEE, December 7-10 2010.

[2] Y. Wang, S. Rane, P. T. Boufounos, and A. Vetro, "Distributed compression of zerotrees of wavelet coefficients," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Brussels, Belgium, Sept. 11-14 2011.

[3] D. Valsesia and P. T. Boufounos, "Universal encoding of multispectral images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, March 20-25 2016.

[4] ——, "Multispectral image compression using universal vector quantization," in *Proc. IEEE Info. Theory Workshop*, Cambridge, UK, Sept 11-14 2016.

[5] P. Boufounos, "Universal rate-efficient scalar quantization," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1861–1872, March 2012.

[6] P. T. Boufounos, "Hierarchical distributed scalar quantization," in *Proc. Int. Conf. Sampling Theory and Applications (SampTA), Singapore*, 2011.

[7] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[9] M. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1122–1134, Feb 2012.

[10] P. T. Boufounos, L. Jacques, F. Krahmer, and R. Saab, "Quantization and compressive sensing," in *Compressed Sensing and its Applications*, ser. Applied and Numerical Harmonic Analysis, H. Boche, R. Calderbank, G. Kutyniok, and J. Vybral, Eds. Springer International Publishing, 2015, pp. 193–237.

[11] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[12] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, Jan 1974.

[13] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar 2003.

[14] T. Ancheta, "Syndrome-source-coding and its universal generalization," *IEEE Trans. Inf. Theory*, vol. 22, no. 4, pp. 432–436, Jul 1976.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[16] T. Koike-Akino, D. S. Millar, K. Kojima, K. Parsons, Y. Miyata, K. Sugihara, and W. Matsumoto, "Iteration-aware LDPC code design for low-power optical communications," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 573–581, 2016.

[17] S. Ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, 2004.

[18] H. Xiao and A. H. Banihashemi, "Improved progressive-edge-growth (PEG) construction of irregular LDPC codes," *IEEE Commun. Lett.*, vol. 8, no. 12, pp. 715–717, 2004.

[19] U. S. Kamilov, "A parallel proximal algorithm for anisotropic total variation minimization," *IEEE Trans. Image Processing*, vol. 26, no. 2, pp. 539–548, 2017.

[20] Japan Aerospace Exploration Agency Earth Observation Research Center. "About ALOS - AVNIR-2". [Online]. Available: http://www.eorc.jaxa.jp/ALOS/en/about/avnir2.htm