UNIVERSAL ENCODING OF MULTISPECTRAL IMAGES

Diego Valsesia[*]

Politecnico di Torino
diego.valsesia@polito.it

Petros T. Boufounos

Mitsubishi Electric Research Laboratories
petrosb@merl.com

ABSTRACT

We propose a new method for low-complexity compression of multispectral images. We develop on a novel approach to coding signals with side information based on recent advances in compressed sensing and universal scalar quantization. Our approach can be interpreted as a variation of quantized compressed sensing, where the most significant bits are discarded at the encoder and recovered at the decoder from the side information. The image is reconstructed using weighted total variation minimization, incorporating side information in the weights while enforcing consistency with the recovered quantized coefficient values. Our experiments validate our approach and confirm the improvements in rate-distortion performance.

***Index Terms*—** Compressed sensing, multispectral image compression, universal quantization, distributed coding

1. INTRODUCTION

The ever increasing demand for imaging data transmission and storage has renewed interest in efficient compression methods. However, in several modern applications, such as satellite image transmission and lightweight mobile computing, technological limitations impose complexity constraints that cannot be satisfied by conventional compression approaches. Thus, new approaches are necessary to handle the low-complexity and high efficiency required.

In this paper we propose a low-complexity multispectral image compression method based on the recently developed theory of universal quantization. Multispectral images comprise of a small number of spectral bands—typically 4 to 6—with significant correlation between them. Modern compression techniques are able to exploit these correlations and improve the rate-distortion performance. However, this often comes at the expense of encoder complexity.

Our approach is able to exploit correlations between spectral bands and reduce the bitrate, while maintaining extremely low complexity at the encoder. Instead, the complexity is shifted to the decoder, which in many applications is a big data center with significant processing power. We design the decoder to use information from previously decoded spectral bands as side information to decode the bitstream. Decoding and recovering the image requires solving a sparse optimization problem very similar to a conventional quantized compressed sensing (CS) problem.

Our contribution relies on several realizations. Specifically, while CS can be used to design light-weight encoders, it is not a rate-efficient encoding scheme. The main reason is that the most significant bits (MSBs) encode significantly redundant information. Universal quantization remedies this problem by eliminating MSBs during quantization, thus removing the redundancies. However, this makes the reconstruction problem non-convex possibly with combinatorial complexity. Instead, our approach uses side information to recover the MSBs, thus making reconstruction convex and tractable without compromising rate-efficiency. Our experimental results show that using universal scalar quantization we can achieve significant improvement in the rate-distortion performance, compared to conventional CS approaches.

The next section provides some background and establishes the notation. We describe the details of our approach in Sec. 3. Section 4 provides experimental results validating our approach. In Sec. 5 we discuss our results and conclude.

2. BACKGROUND

2.1. Multispectral Image Compression

While multispectral images can, in principle, be compressed using conventional image compression techniques, in many applications such techniques are often not suitable. In particular, onboard of spacecrafts computational power is very scarce. Thus, spaceborne compression algorithms require significantly lower complexity and, in turn, significantly different designs, than conventional methods such as JPEG and JPEG2000.

Similar to conventional methods, transform coding is often the workhorse of many popular approaches [1,2], albeit with transforms designed to have reduced complexity in the spatial and spectral dimensions. An alternative approach is predictive coding with predictors mainly based on adaptive filters [3].

Distributed source coding, relying on the celebrated Slepian-Wolf and Wyner-Ziv bounds [4, 5], is also a very popular approach due to its encoding simplicity [6–8]. These approaches treat part of the data as side information and code the remaining data assuming this side information is available at the decoder. The side information might be transmitted uncompressed or using low-complexity conventional compression techniques.

The approach we propose in this paper is similar to distributed coding in that it relies on side information during decoding. However, it uses very different methods than conventional distributing coding approaches and enables the use of sparsity and other modern signal models during decoding.

2.2. Compressed Sensing

Compressed sensing is a well-established, by now, theory for signal acquisition, able to measure and reconstruct signals using fewer linear measurements than dictated by the classical Nyquist-Shannon theory [9, 10]. Reconstruction is possible using additional knowledge about the signal and exploiting appropriate models.

The canonical CS problem considers a signal $\mathbf{x} \in \mathbb{R}^n$ having a sparse representation $\boldsymbol{\theta}$ under some basis $\Psi \in \mathbb{R}^{n\times n}$: $\mathbf{x} = \Psi\boldsymbol{\theta}$,

[*]The first author performed this work while at MERL

with $\|\boldsymbol{\theta}\|_0 = k \ll n$. Here, the ℓ_0 norm $\|\boldsymbol{\theta}\|_0$ counts the number of nonzero coefficients of $\boldsymbol{\theta}$. Measurements are acquired as a vector of random projections $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\boldsymbol{\theta}$, $\mathbf{y} \in \mathbb{R}^m$, using a sensing matrix $\Phi \in \mathbb{R}^{m\times n}$, typically implemented directly in hardware. The original sparse signal can be recovered from the measurements using a convex optimization problem, minimizing the l_1 norm of the solution, or a greedy algorithm such as [11–14].

Using appropriate, typically randomized, sensing matrices only $m = O(k \log n) \ll n$ are required for signal reconstruction to be possible, compared to n required if the signal is not known to be sparse. Thus, CS acquisition preforms an implicit compression of the signal during acquisition. The most studied sensing matrices have entries that are i.i.d. realizations of subgaussian random variables [15]. However, more structured sensing matrices, such as partial Hadamard [16] or circulant [17], allow computation of the measurements and reconstruction using fast transforms while still exhibiting good dimensionality reduction properties. Fast transforms are appealing in computationally constrained systems, because they significantly reduce memory and computational requirements.

For many commonly encountered signals, such as natural images, sparsity is often not an appropriate signal model. Thus, a number of signal models have been proposed as alternatives. In particular, natural images typically exhibit compressible gradients, i.e., very few of their gradient coefficients pack most of the energy. Thus, total variation (TV) minimization can be used for recovery [18]:

$$\widehat{X} = \arg\min_{X} \ \mathrm{TV}(X) + \lambda\|\mathbf{y} - \Phi\mathbf{x}\|_2^2, \tag{1}$$

where X is a two-dimensional image, $\mathbf{x}$ is a vectorized version of the image X and the isotropic TV is defined as

$$\mathrm{TV}(X) = \sum_{i,j} \sqrt{|X_{i+1,j} - X_{i,j}|^2 + |X_{i,j+1} - X_{i,j}|^2}. \tag{2}$$

Compressive sensing is a very effective acquisition approach. However, the simplicity in acquisition suggests it can also be used for efficient and low-complexity signal compression, simply by digitally implementing the measurement process. Unfortunately, CS-based compression methods suffer from poor rate-distortion performance, despite the significant undersampling factor. Fundamentally, the undersampling performed during CS acquisition behaves like oversampling when sparsity is taken into account and performs significantly worse than transform coding [19, 20]. In hindsight, this is expected and in-line with well-established results on scalar quantization of oversampled signals (e.g., see [21]).

2.3. Universal scalar quantization

Universal scalar quantization (USQ) has been recently proposed as an alternative scalar quantization technique that, coupled with consistent reconstruction, promises efficient coding of signals [22]. In particular, uniform scalar quantization of CS measurements is limited by a linear reduction in distortion as the number of measurements increases. In contrast, consistent reconstruction from universally quantized measurements achieves an exponential reduction.

The key ingredient of universal scalar quantization is a non-monotonic quantizer having non-contiguous quantization regions. It can be though of as a uniform linear quantizer with the MSBs removed. Figure 1 shows examples of (a) a 3-bit conventional uniform linear quantizer with step-size Δ and corresponding (b) 1-bit and (c) a 2-bit universal quantizers. Disjoint intervals that share the same 2 or 1 least significant bits in the conventional uniform quantizer will quantize to the same value using a uniform 1- or 2-bit quantizer.

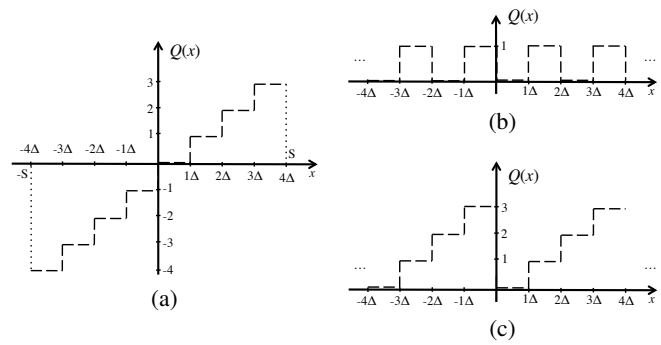

Fig. 1: (a) Conventional 3-bit uniform linear quantizer and corresponding (b) 1-bit and (c) 2-bit universal quantizers, equivalent to the uniform linear quantizer with, respectively, 2 or 1 most significant bits removed.

Before non-monotonic quantization, the signal $\mathbf{x}$ to be encoded is sampled using a sensing matrix Φ to obtain measurements $\mathbf{y} = \Phi\mathbf{x}$. Those measurements are dithered and quantized to produce $\mathbf{y}_q$:

$$\mathbf{y} = \Phi\mathbf{x}, \ \ \mathbf{y}_q = \lfloor \mathbf{y}/\Delta + \mathbf{w} + 0.5 \rfloor \mod 2^B, \tag{3}$$

where $\mathbf{w}$ is additive dither with entries w_i drawn from an i.i.d. distribution uniform over $[0, 1)$. The quantizer parameters are the step size Δ and the rate B, resulting to a quantizer with 2^B levels.

As evident in Fig. 1(b) and (c), the resulting quantization function is periodic. More importantly, the consistency set corresponding to any quantized measurement is composed of disjoint intervals, one per period, forming a non-convex set. Thus, consistent reconstruction is a computationally challenging problem with combinatorial complexity in general. Computationally tractable reconstruction is possible by designing the quantization intervals to form a hierarchy of convex problems, often at the expense of rate efficiency [23].

3. COMPRESSION OF MULTISPECTRAL IMAGES

Our compression approach fundamentally exploits a simple premise: if universal scalar quantization can be thought of as uniform quatization with missing bits and we have access to reliable side information, then the decoder can use the side information to predict and fill in the missing bits and solve a conventional quantized CS problem.

The high-level encoder and decoder architecture is shown in Fig. 2, the components of which we describe in the next two sections. Some experiment-specific details of our implementation are described more thoroughly in Sec. 4.

3.1. Encoding

Similar to [6, 7], we assume that the encoder transmits one of the bands as side information, encoding it using a standard technique. This side information, maybe combined with additional statistics transmitted by the encoder, is used to predict the other bands.

The remaining bands are partitioned into non-overlapping blocks of size $n_x \times n_y$. A small number of random projections $\mathbf{y} \in \mathbb{R}^m$ is computed for each block $\mathbf{x} \in \mathbb{R}^n$, $n = n_x \times n_y$ using a partial Hadamard matrix Φ, obtained by randomly subsampling the rows of the Hadamard transform. The measurements are quantized with a universal scalar quantizer of rate B and step size Δ as in (3). This accounts for $\frac{m}{n}B$ bits per pixel (bpp).

The side information is used at the decoder to predict the MSBs missing from the universally quantized data, thus making consistent

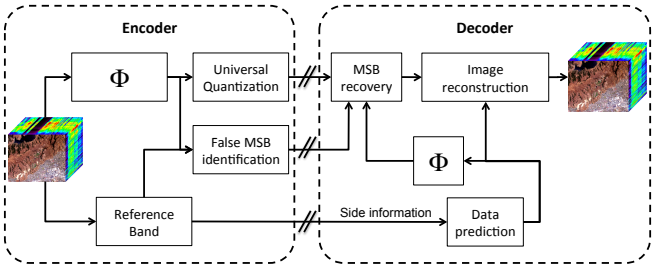

Fig. 2: High-level encoding and decoding architecture for the proposed compression approach.

reconstruction a convex problem. However, MSB prediction might make errors. Specifically, when the predicted MSB is combined with the universally quantized measurements, an error of order κ in the predicted MSB causes an error equal to $\pm\kappa\Delta 2^B$ in the recovered coefficient. The quality of the prediction is affected by the number of MSBs dropped, or, equivalently, by the size of the quantization step size Δ and the rate B used. Larger Δ and higher B result in fewer errors, at the expense of rate-distortion efficiency. On the other hand, errors may introduce distortion at the decoder if not corrected.

Fortunately, the encoder knows the side information and, therefore, can compute, encode, and transmit additional side information indicating where such errors occur. Thus, the trade-off in designing Δ and B to introduce more or fewer errors now manifests as additional rate required to encode the side information with the location of the errors. Since larger Δ and higher B result in fewer MSB errors, they require lower rate for MSB error encoding, at the expense of larger reconstruction error and higher universal quantization rate, respectively. Correspondingly, smaller Δ and B reduce the reconstruction distortion and universal quantization rate, respectively, but increase the rate to encode side information on MSB errors.

Since the decoder itself can be made robust to some sparse errors in the measurements, our approach is a compromise: we explicitly encode first order errors, i.e., for which $\kappa = 1$, by sorting their locations in increasing order and then differentially coding them using an Exp-Golomb universal code [24], together with their sign. For higher order errors, i.e., for which $\kappa > 1$, only their location is encoded, but not their multiplicity or their sign.

3.2. Decoding

The decoder has available the universally quantized CS measurements $\mathbf{y}_q$ and a real-valued prediction of them $\mathbf{y}_{\text{pred}}$ obtained from the side information. Typically this prediction is obtained by first predicting the signal $\mathbf{x}_{\text{pred}}$ and then measuring it using the measurement matrix, the dither and scaling: $\mathbf{y}_{\text{pred}} = \Phi\mathbf{x}_{\text{pred}}/\Delta + \mathbf{w}$.

Given universally quantized measurements, the prediction is used to identify the quantization point on a uniform linear quantizer that is closest among the multiple possible candidates. An example is shown in Fig. 3, where the universally quantized value of 1 is resolved to -3 given a prediction slightly less than -1Δ. Formally, the quantization point is resolved using minimum distance decoding:

$$\widehat{\mathbf{y}} = \Delta\left(\mathbf{y}_p + \mathbf{y}_q + \widehat{\mathbf{t}}\right),\ \mathbf{y}_p = 2^B \cdot \left\lfloor \frac{\mathbf{y}_{\text{pred}}}{\Delta 2^B} + \frac{0.5}{2^B} \right\rfloor \tag{4}$$

where each component $\widehat{t}_i$ is an offset that minimizes the distance to the prediction

$$\widehat{t}_i = \arg \min_{t \in \{-\Delta 2^B, 0, \Delta 2^B\}} \left| y_{\text{pred},i} - \left(y_{p,i} + t + y_{q,i}\right)\Delta \right| \tag{5}$$

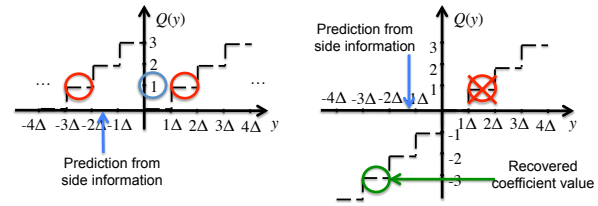

Fig. 3: Given a universal quantization point, the prediction is used to select the closest corresponding point on a uniform linear quantizer out of the multiple candidates.

Depending on the quality of the prediction and on the chosen value of B and Δ, prediction errors might be more or less frequent. Thus, the recovered quantized measurements can be modeled as

$$\hat{\mathbf{y}} = \Phi\mathbf{x} + \mathbf{e} + \nu \tag{6}$$

where ν is the uniform quantization error and $\mathbf{e}$ is a vector with elements drawn from a finite alphabet of integer multiples of $\Delta 2^B$ capturing the decoding errors.

An error of order κ is captured by entries of $\mathbf{e}$ with value $\pm\kappa\Delta 2^B$. Given a good prediction and suitable values of Δ, then $\mathbf{e}$ tends to be sparse. Furthermore, as mentioned in Sec. 3.1, the encoder might include information on $\mathbf{e}$ which can be used to correct some or all of the errors and reset the corresponding coefficients of $\mathbf{e}$ to 0. We use $\mathcal{S}$ to denote the set containing the locations of known errors that have not been corrected, typically of order 2 or higher.

To recover the image, the decoder uses the recovered measurements, aided by the image prediction. Specifically, recovery solves a weighted TV minimization with consistent reconstruction:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \text{WTV}(x) + \lambda f\left(\Phi\mathbf{x}\right) \\ &\text{s.t. } \left|y_i - \left(\Phi\mathbf{x}\right)_i\right| \leq \frac{\Delta}{2} \text{ if } i \notin \mathcal{S}, \end{aligned} \tag{7}$$

where $\text{WTV}(\cdot)$ is the isotropic weighted total variation

$$\text{WTV}(x) = \sum_{i,j} \sqrt{W_{i,j}^{(x)}\left(X_{i,j} - X_{i-1,j}\right)^2 + W_{i,j}^{(y)}\left(X_{i,j} - X_{i,j-1}\right)^2}, \tag{8}$$

$f(\cdot)$ penalizes decoding errors using an ℓ_1-type penalty,

$$f\left(\Phi\mathbf{x}\right) = \sum_{i \in \mathcal{S}} \max\left(\left|y_i - \left(\Phi\mathbf{x}\right)_i\right| - \frac{\Delta}{2}, 0\right), \tag{9}$$

and $W_{i,j}$ are weights that determine how gradients in each pixel of the image should be penalized.

In addition to filling in the MSBs in the encoded data, the prediction obtained from the side information is also used to derive the weights $W_{i,j}$. Low weights are used when the gradient magnitude of the prediction is higher than a predefined threshold and high weights when the gradient is lower, a setting similar to the weighted ℓ_1 minimization in [25]. The resulting model penalizes edges that do not exist in the prediction more than the ones that do exist. Since prediction is derived from the other spectral band, the model reinforces correlations between spectral bands, especially among the edges.

The function $f(\cdot)$ promotes data consistency for the part of the data in $\mathcal{S}$, i.e., where we suspect there is an MSB decoding error. While a quadratic penalty is a more common data consistency penalty, the sparsity of the errors in our case suggests that an ℓ_1-based penalty is more appropriate [26]. Since the measurements

Table 1: MSB decoding errors and the corresponding transmission overhead.

	$B=3$				$B=4$				$B=5$			
	band 2	band 3	band 4	bpp	band 2	band 3	band 4	bpp	band 2	band 3	band 4	bpp
$\Delta=80$	6.40 %	25.63 %	26.61 %	0.527	0.18 %	3.56 %	4.76 %	0.097	0 %	0.04 %	0.19 %	0.11
$\Delta=100$	2.60 %	16.45 %	17.66 %	0.346	0.03 %	1.21 %	1.97 %	0.044	0 %	0.01 %	0.05 %	0.009
$\Delta=120$	1.08 %	10.16 %	11.50 %	0.228	0.01 %	0.41 %	0.88 %	0.023	0 %	0 %	0.01 %	0.009

Table 2: Decoding PSNR at 2 bpp

	band 2	band 3	band 4
(a) Prediction	33.68 dB	28.87 dB	28.35 dB
(b) Classic CS	33.24 dB	31.18 dB	33.58 dB
(c) Weighted CS	34.40 dB	31.63 dB	33.95 dB
(d) **Proposed**	**37.79 dB**	**32.76 dB**	**34.24 dB**

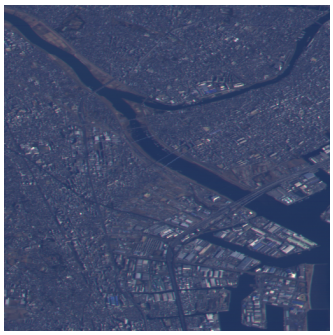

Fig. 4: Test multispectral image acquired using the AVNIR-2 instrument of the ALOS satellite [27]. This image of a coastal city exhibits complexity and details that challenge compression algorithms.

are quantized, we include a deadzone of width Δ, promoting consistency to the quantization intervals. Consistency on the rest of the data is enforced using the hard constraint $|y_i - (\Phi\mathbf{x})_i| \leq \Delta/2$ in (7).

4. EXPERIMENTAL RESULTS

We tested the scheme on the $512 \times 512 \times 4$ multispectral image shown in Fig. 4. The first band is used as side information to predict the content of the other bands and it is compressed losslessly in our experiments. We separate each image in blocks of 32×32 and code each block separately.

To predict band b for each block we use classical linear prediction from the same block in band 1:

$$\hat{\mathbf{x}}_b = \frac{\sigma_{1b}}{\sigma_1^2}(\mathbf{x}_r - \mu_1) + \mu_b, \tag{10}$$

where μ_b is the block mean for band b, σ_b^2 is the variance, and σ_{1b} the covariance of b with block 1. The parameters are computed at the encoder and transmitted as side information. Assuming 16-bit values in the worst case, the overhead is 0.047 bits per pixel (bpp).

Using the prediction we decode the universally quantized measurements as described in Sec. 3.2. First order errors and their sign are detected at the encoder, transmitted and corrected at the decoder. For second and higher order errors, only their location is transmitted to form the set $\mathcal{S}$ during decoding. The total overhead to transmit the errors is variable and depends on the choice of Δ and B.

Table 1 reports the percentage of measurements ($m = 512$) affected by first or higher order errors as a function of the quantization step size Δ and rate B, as well as the cost in bpp for coding these errors as described above. As expected, larger Δ and B result to lower error rate and corresponding coding overhead. Furthermore, bands further away from the reference band 1, i.e., bands 3 and 4, are more difficult to predict, and, therefore, the errors increase.

Since the reconstruction algorithm can tolerate some decoding errors, perfect decoding is not our goal. In our experiments we found that $\Delta = [100, 140, 120]$ and $B = [2, 3, 3]$, with $m = [640, 448, 390]$ measurements, was a good parameter choice to encode bands 2 to 4, respectively, performing consistently in a variety of similar images. Since errors increase for bands further away from the reference, we expect the corresponding optimal parameter choices to have higher Δ and B to reduce the errors.

Table 2 compares the PSNR obtained by (a) simple linear prediction, a classic CS encoder using a uniform scalar quantizer at a compression rate of 2 bpp and reconstruction using (b) TV minimization or (c) WTV minimization using weights obtained from the reference band, and (d) the proposed method at the same compression rate. The rate reported for our proposed scheme includes the overhead due to prediction and error coding. For comparison, classic CS coding and reconstruction at 2 bpp, regularized using TV minimization, sometimes performs worse even than simple linear prediction, which requires less than 0.047 bpp. Reconstruction of the same data using WTV minimization, i.e., exploiting side information at the decoder, provides a small gain of 0.5–1dB, depending on the band.

As evident in the table, the proposed approach significantly improves performance, especially if the prediction is of higher quality. Furthermore, most of the gain is due to the use of universal quantization instead of WTV. It should also be noted that the method used in the experiment is not particularly optimized. A number of improvements could further boost performance such as better choice of the parameters m, Δ and B that could be optimized per block, instead of per band, or better prediction schemes, to name a few.

5. DISCUSSION AND CONCLUSIONS

The proposed compression scheme is the first application of universal quantization with side information, a new approach to distributed compression. In contrast to conventional distributed coding, our approach enables decoding using the sophisticated signal models recently emerged in the context of CS. Thus, it significantly outperforms conventional quantized CS approaches that use scalar quantization. The method provides very light-weight encoding, with more complex but computationally tractable decoding.

Our results validate our approach. Our method is not competitive in rate-distortion performance with highly refined state-of-the-art multispectral compression methods, which require much more complex encoders. However, we do believe that, with sufficient refinement, the gap can be bridged. Our work only scratches the surface of this rich topic. A key issue is the design of the encoder parameters. There is currently no guidance on how these should be selected, other than intuitive general principles. Furthermore, the importance of prediction cannot be understated. Transmitting and using optimally the right kind of side information can bring notable improvements in the performance of the current scheme.

6. REFERENCES

[1] Consulative Committee for Space Data Systems (CCSDS), "Image Data Compression," *Blue Book*, November 2005.

[2] I. Blanes and J. Serra-Sagristà, "Pairwise orthogonal transform for spectral image coding," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 3, pp. 961–972, 2011.

[3] D. Valsesia and E. Magli, "A novel rate control algorithm for onboard predictive coding of multispectral and hyperspectral images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 10, pp. 6341–6355, Oct 2014.

[4] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471–480, Jul 1973.

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

[6] 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, December 7-10 2010, IEEE.

[7] 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.

[8] A. Abrardo, M. Barni, E. Magli, and F. Nencini, "Error-resilient and low-complexity onboard lossless compression of hyperspectral images by means of distributed source coding," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 4, pp. 1892–1904, April 2010.

[9] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[10] E.J. Candes and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[11] T. Blumensath and M. E Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[12] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.

[13] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec. 2010.

[14] Anastasios Kyrillidis and Volkan Cevher, "Recipes on hard thresholding methods," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on*. IEEE, 2011, pp. 353–356.

[15] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[16] J. A. Tropp, "Improved analysis of the subsampled randomized hadamard transform," *Advances in Adaptive Data Analysis*, vol. 03, no. 01n02, pp. 115–126, 2011.

[17] H. Rauhut, J. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242–254, 2012.

[18] Deanna Needell and Rachel Ward, "Stable image reconstruction using total variation minimization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, 2013.

[19] P. T. Boufounos and R. G. Baraniuk, "Quantization of sparse representations," in *Rice University ECE Department Technical Report 0701. Summary appears in Proc. Data Compression Conference (DCC)*, Snowbird, UT, March 27-29 2007.

[20] Petros T. Boufounos, Laurent Jacques, Felix Krahmer, and Rayan Saab, "Quantization and compressive sensing," in *Compressed Sensing and its Applications*, Holger Boche, Robert Calderbank, Gitta Kutyniok, and Jan Vybíral, Eds., Applied and Numerical Harmonic Analysis, pp. 193–237. Springer International Publishing, 2015.

[21] V.K. Goyal, M. Vetterli, and N.T. Thao, "Quantized overcomplete expansions in irn: analysis, synthesis, and algorithms," *Information Theory, IEEE Transactions on*, vol. 44, no. 1, pp. 16–31, Jan 1998.

[22] P.T. Boufounos, "Universal rate-efficient scalar quantization," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1861–1872, March 2012.

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

[24] J. Teuhola, "A compression method for clustered bit-vectors," *Information Processing Letters*, vol. 7, no. 6, pp. 308–311, 1978.

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

[26] J.N. Laska, M.A. Davenport, and R.G. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, Nov 2009, pp. 1556–1560.

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