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

Multispectral Image Compression Using Universal Vector Quantization

Valsesia, D.; Boufounos, P.T.

TR2016-119 September 2016

Abstract

We propose a new method for low-complexity compression of multispectral images based on universal vector quantization. Our approach generalizes the recently developed theory of universal scalar quantization to vector quantization, and uses it in the context of distributed coding. We exploit the availability of side information on the decoder to reduce the encoding rate of a vector quantizer, applied to compressed measurements of the image. The encoding reuses quantization labels to label multiple quantization cells and leverages the side information to select the correct cell at the decoder. The image is reconstructed using weighted total variation minimization, incorporating side information in the weights while enforcing consistency with the recovered quantization cell.

Information Theory Workshop

Multispectral Image Compression Using Universal Vector Quantization

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 based on universal vector quantization. Our approach generalizes the recently developed theory of universal scalar quantization to vector quantization, and uses it in the context of distributed coding. We exploit the availability of side information on the decoder to reduce the encoding rate of a vector quantizer, applied to compressed measurements of the image. The encoding reuses quantization labels to label multiple quantization cells and leverages the side information to select the correct cell at the decoder. The image is reconstructed using weighted total variation minimization, incorporating side information in the weights while enforcing consistency with the recovered quantization cell.**

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

I. Introduction

Data compression is a widely studied topic and considered a relatively mature technology. However, interest has been renewed due to recent developments in the area of compressive sensing, as well as ever evolving demand for signal acquisition, transmission and storage. 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. In many of these applications, decompression is performed in large data centers, where computational complexity is a lesser concern. Thus, a number of conventional approaches, which either balance complexity or require more computation at the encoder side, are not appropriate. 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. Specifically, we extend our existing work [1] by generalizing scalar universal quantization to a vector formulation that further improves performance. While we only describe the generalization to a 3-dimensional vector space, further generalization is possible, at the expense of decoding complexity. Our generalization has connections to coset codes [2], [3], which we defer to subsequent publications. However, our work, in addition to coding, exploits and incorporates the most recent advances in signal models, stemming out of the compressive sensing literature.

Our focus is on multispectral images, which 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 requires significant encoder complexity. In a number of application, especially satellite-borne imaging systems, encoder complexity can be prohibitive. Instead, it is desirable to shift the complexity to the decoder, which is typically a big data center with significant processing power.

Indeed, our approach exploits correlations between spectral bands to reduce the bitrate, while maintaining extremely low complexity at the encoder. The correlations are exploited at the decoder, which is designed to use information from previously decoded spectral bands as side information to augment and 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 a key realization: while CS can be used to design light-weight encoders, it is not a rate-efficient encoding scheme. In particular, the most significant bits (MSBs) encode redundant information [4]. Universal quantization removes redundancies by eliminating the MSBs. However, this makes the reconstruction problem non-convex possibly with combinatorial complexity. Generalizing to vector quantization further exacerbates the problem.

As a remedy, our approach uses side information to make the reconstruction convex and tractable, without compromising rate-efficiency. In particular, similarly to most distributed coding schemes, the encoder only transmits information on how to refine the prediction from the side information, i.e., the relative location of compressive measurements of the signal with respect to their prediction. The decoder, thus, uses the prediction to generate a convex quantization cell in which the measurements belong, and uses a sparse recovery algorithm to decode the signal with measurements in that cell. In addition, the decoder uses the side information to bias the compressive recovery algorithm to a solution closer to the encoded signal.

The next section provides some background and establishes notation. The approach is described in Sec. III. Finally, Section IV provides experimental validation and discussion.

II. Background

A. Multispectral Image Compression

Multispectral images can, in principle, be compressed using conventional image compression techniques. However, when

encoding should have low complexity, such techniques are often not suitable. In particular, onboard of spacecrafts computational power is very scarce. Thus, spaceborne compression algorithms require different designs than conventional methods, such as JPEG and JPEG2000.

Similar to conventional methods, transform coding is often the workhorse of many popular approaches [5], [6], albeit with transforms designed to reduce computation. An alternative approach is predictive coding with predictors mainly based on adaptive filters [7]. A popular alternative is distributed source coding, relying on the celebrated Slepian-Wolf and Wyner-Ziv bounds [8], [9]. Its appeal is mostly due to encoding simplicity [10]–[12]. 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.

Our approach, which generalizes [1], bears similarities to distributed coding in relying on side information during decoding. However, it uses very different methods than conventional distributing coding and enables the use of sparsity and other modern signal models during decoding.

B. Compressed Sensing

Compressed sensing is a well-established, by now, theory for signal acquisition, providing the ability to undersample signal and still successfully reconstruct them [13], [14]. Unique and correct reconstruction is possible using additional knowledge about the signal and exploiting appropriate models.

The canonical CS problem considers a sparse signal $\mathbf{x} \in \mathbb{R}^n$. The signal is measured using random projections $\mathbf{y} = \Phi\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$ acquired by a sensing matrix $\Phi \in \mathbb{R}^{m\times n}$. For natural images, sparsity in the gradient, i.e., low total variation (TV), is usually the preferred signal model. Recovery enforces the model through TV minimization [15]:

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

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

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

Using appropriate sensing matrices only $m = O(k \log n) \ll n$ are required for signal reconstruction—where k measures the model sparsity—compared to the n required if the signal is not known to be sparse. Thus, CS acquisition preforms an implicit compression of the signal during acquisition. A number of matrix constructions have been shown to work, including fully randomized ones as well as more structured ones exploiting fast transforms, e.g., [16]–[18]. Fast transforms are appealing for lightweight compression, because they significantly reduce memory and computational requirements.

While CS is a very effective acquisition approach, it does not perform well as a compression method, if implemented in a straightforward manner. In particular CS-based compression methods suffer from poor rate-distortion performance compared to transform coding, despite the significant undersampling factor [4], [19]. The reason is that, fundamentally, the CS-based measurements oversample the signal once the sparsity pattern sparsity is taken into account. In hindsight, this is expected and in-line with well-established results on scalar quantization of oversampled signals (e.g., see [20]).

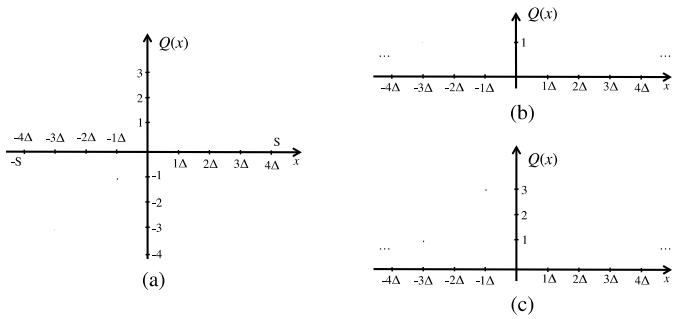


Fig. 1: (a) Conventional 3-bit uniform linear quantizer and corresponding (b) 1-bit and (c) 2-bit universal quantizers, equivalent to the uniform linear quantizer with, respectively, 2 or 1 most significant bits removed.

C. Universal Scalar Quantization

Universal scalar quantization (USQ) has been recently proposed as an alternative quantization approach to improve the coding efficiency of CS-based systems [21]. In particular, while uniform scalar quantization of CS measurements only achieves linear reduction in distortion as the number of measurements increases, consistent reconstruction from universally quantized measurements can achieves exponential reduction.

The improved performance is achieved using a non-monotonic scalar quantizer which eliminates the MSB of a conventional uniform scalar quantizer. 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.

The resulting reconstruction problem is non-convex. While consistent reconstruction with sufficient number of measurements guarantees accurate solution, it is a computationally challenging problem with combinatorial complexity in general. Computationally tractable reconstruction can be achieved by designing the quantization intervals to form a hierarchy of convex problems, often at the expense of rate efficiency [22]. Instead, [1] exploits the side information to fill-in the missing MSBs and solve a convex problem.

III. Compression of multispectral images

A. Universal Vector Quantization

Our compression approach fundamentally exploits a simple premise: universal vector quantization can be thought of as conventional vector quantization with missing information such that quantization labels are reused for different parts of the quantized vector space. In the subsequent development we use a 3D lattice. However, for ease of visualization, a 2D

example of the principle is shown in Fig. 3(a). Of course, higher dimensional lattices could also be used.

As shown in the figure, the quantizer tiles the space similarly to a conventional vector quantizer. However, quantization labels are reused, ensuring that the same label is not assigned to bordering cells. Ideally, cells with the same label are as separated as possible. Quantizer design and label assignment are two very interesting problems, with several connections to coset codes [2], [3], that we defer to future publications.

In this work we use the D_3 lattice and the quantization label of a vector $\mathbf{y}$ is computed using

$$\mathbf{y}_q = D_3\left(\frac{\mathbf{y}}{\Delta} + \mathbf{w}\right) \mod 2^B, \tag{3}$$

where the $D_3(\cdot)$ is the D_3 lattice quantizer, operating on triplets in its argument, Δ is a scaling factor, $\mathbf{w}$ is an optional dither drawn uniformly over a canonical quantization cell, and mod is taken along each of the 3 coordinates separately. The Voronoi regions of the D_3 lattice quantizer are dodecahedra, forming a space-filling packing in 3D. The mod function ensures that quantizer labels are reused and that centers with the same label have offsets equal to integer multiples of $\Delta 2^B$ along each coordinate.

Since the decoder has access to reliable side information, it can resolve the ambiguities resulting from label reuse 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. IV.

B. Encoding

Similar to [1], [10], [11], 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 in sets of 3 using the D_3 lattice with a scaling of Δ, according to (3).

When the universal vector quantizer uses the D_3 lattice, each block of three measurements can be mapped to one of 2^{3B-1} points. Therefore, $B - \frac{1}{3}$ bits per measurement are required. Taking into account the subsampling factor, the actual rate is $\frac{m}{n}\left(B - \frac{1}{3}\right)$ bits per pixel (bpp).

Similarly to [1], the side information is used at the decoder to predict the original quantization cell, thus resolving the ambiguity arising from the mod operation, and making consistent reconstruction a convex problem. However, the prediction might make errors depending on its quality. A group is affected by a first order error when the error vector has norm equal to $\Delta 2^B$, i.e., only one of the entries of the error vector contains either $\pm\Delta 2^B$. The decoder itself can be made robust to some sparse errors, but a low quality prediction can cause more errors than the decoder can correct.

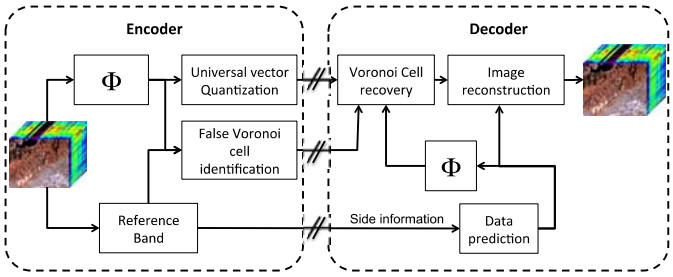


Fig. 2: High-level encoding and decoding architecture for the proposed compression approach.

Fortunately, the encoder also has access to the other bands and 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 prediction errors, they require lower rate for 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 errors.

Our approach is a compromise: we explicitly encode first order errors because their correction in the reconstruction program is not very effective due to their small norm. This is done in the following way. Each group has a label ranging from 0 to $\frac{m}{3} - 1$. The labels of groups exhibiting first order errors are sorted by increasing order and differentially encoded using a universal Exp-Golomb code [23]. There can only be 6 types of errors, so a code requiring at least $\log_2(6)$ bits per group is enough to distinguish the correct error. Concerning higher order errors, the group labels are the only information that is coded, again with a differential Exp-Golomb code.

C. Decoding

The decoder has available the universally quantized CS measurements $\mathbf{y}_q$ and a real-valued prediction of them $\mathbf{p}$ 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: $\mathbf{p} = \Phi\mathbf{x}_{\text{pred}} + \mathbf{w}$.

Similarly to the scalar case, the prediction is used to the determine the correct Voronoi cell in which the universally quantized measurements belonged before universal quantization. The prediction $\mathbf{p}$ is quantized to the closest lattice point $\mathbf{p}_q = D_3\left(\frac{\mathbf{p}}{\Delta}\right)$, and then universally quantized to obtain $\mathbf{p}_q^{uni} = \mathbf{p}_q \mod 2^B$. In addition to $\mathbf{p}_q^{uni}$, all the neighbouring consistent points are checked to recover the correct quantization region. The goal is to choose the region that is consistent with the universally quantized measurements and closest to the prediction, as shown in Fig. 3(b) for the 2D lattice. For the D_3 lattice, there are 27 neighbouring points generated as:

$$\mathbf{c}_i = \mathbf{p}_q - \mathbf{p}_q^{uni} + \mathbf{s}_i, \tag{4}$$

where $\mathbf{s}_i$ are all the possible combinations of elements $0, 2^B, -2^B$ in a three-dimensional vector.

Finally, the estimated reconstruction point is:

$$\hat{\mathbf{y}} = (\mathbf{c}_{\hat{i}} + \mathbf{y}_q)\,\Delta \tag{5}$$

where

$$\hat{i} = \arg\min_i \left\| \frac{\mathbf{p}}{\Delta} - (\mathbf{c}_i + \mathbf{y}_q) \right\| \tag{6}$$

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

$$\hat{\mathbf{y}} = \Phi\mathbf{x} + \mathbf{w} + \mathbf{e} + \nu \tag{7}$$

where ν is the 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.

Given a good prediction and suitable values of Δ, then $\mathbf{e}$ tends to be group-sparse. Furthermore, as mentioned in Sec. III-B, 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 labels of the groups containing known errors that have not been corrected, typically of order 2 or higher.

Consistency with the quantization lattice should be enforced in the reconstruction program. However, the Voronoi regions of the D_3 lattice are dodecahedra which can make the decoding quite complex. An approximation is therefore used, where the dodecahedra are replaced by spheres of radius Δ.

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:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \mathrm{WTV}(X) + \lambda f\left(\Phi\mathbf{x}\right) \\ \text{s.t. } &\left\| \mathbf{y}_{\mathcal{G}_i} - (\Phi\mathbf{x} + \mathbf{w})_{\mathcal{G}_i} \right\| \leq \Delta \text{ if } \mathcal{G}_i \nsubseteq \mathcal{S}, \end{aligned} \tag{8}$$

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

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

$f(\cdot)$ penalizes decoding errors using a penalty with a mixed ℓ_1/ℓ_2 norm, due to the group-sparsity of the errors,

$$f\left(\Phi\mathbf{x}\right) = \sum_{\mathcal{G}_i \subseteq \mathcal{S}} \max\left\{ \left\| \mathbf{y}_{\mathcal{G}_i} - (\Phi\mathbf{x} + \mathbf{w})_{\mathcal{G}_i} \right\| - \Delta, 0 \right\} \tag{9}$$

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

In addition to resolving quantization ambiguities, 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 [24]. 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.

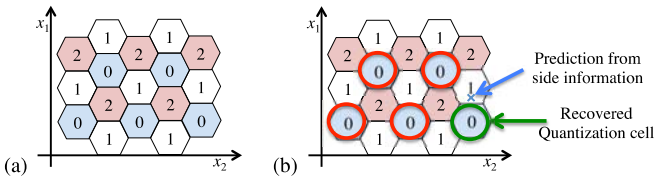


Fig. 3: (a) 2D vector universal quantizer example. (b) Given a universal quantization level, the prediction is used to select the closest corresponding cell out of the multiple candidates.

TABLE I: 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) USQ	37.79 dB	32.76 dB	34.24 dB

The function $f(\cdot)$ promotes data consistency for the part of the data in $\mathcal{S}$, i.e., where we suspect there is a decoding error. While a quadratic penalty is a more common data consistency penalty, in [1] we showed that a similar penalty for the universal scalar quantizer enabled the recovery of some decoding errors thanks to their sparsity. The penalty we propose in this paper is a generalization to block-sparse vectors, as it fits the kind of decoding errors arising in vector universal quantization.

IV. Experimental results

We tested the scheme on the $512 \times 512 \times 4$ multispectral image shown in Fig. 4. The first band, i.e., blue, 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.

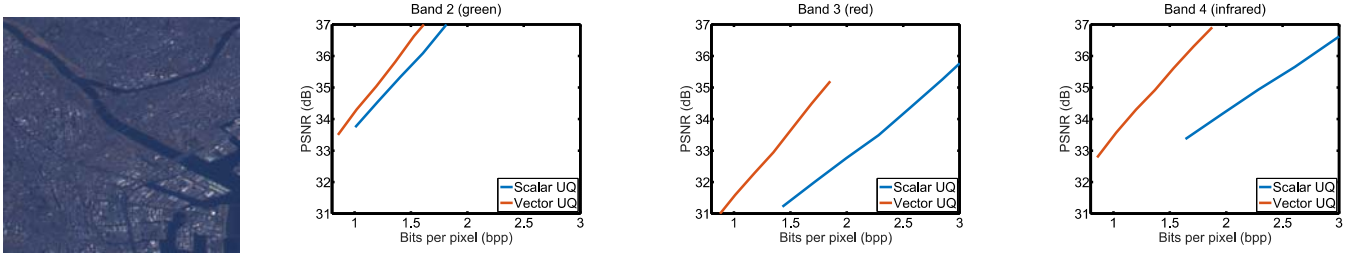
To predict band b for each block we use classical linear prediction from the same block in band 1:

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

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. III-C. 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 I, replicated from [1], 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



(a) (b) (c) (d)

Fig. 4: (a) Test multispectral image acquired using the AVNIR-2 instrument of the ALOS satellite [25]. This image of a coastal city exhibits complexity and details that challenge compression algorithms. The experimental rate-distortion curves, in figures (b), (c) and (d), compare universal scalar quantization (blue curve) with universal vector quantization (red curve) for bands 2, 3, and 4, assuming band 1 is used as reference.

minimization using weights obtained from the reference band, and (d) universal scalar quantization (USQ) at the same rate. It is evident from the table that universal quantization significantly improves performance over simpler approaches. It should also be noted that bands 3 and 4, (red and infrared), are more difficult to predict from band 1 (blue) compared to band 2 (green), and therefore, performance suffers.

Figures 4(b), (c), and (d) demonstrate the performance the proposed universal vector quantization on spectral bands 2, 3, and 4, respectively. The figures plot the rate-distortion trade-off for universal vector quantization (red), compared to universal scalar quantization (blue). As evident, universal vector quantization further improves the reconstruction PSNR by 1-2 dB. In particular, more improvement is exhibited when prediction is harder, i.e., bands 3 and 4.

In summary, the proposed approach significantly improves performance, especially when universal scalar quantization suffers because of lower prediction quality. It should be noted that the method used in the experiments 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.

REFERENCES

[1] 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.

[2] G David Forney Jr, "Coset codes. i. introduction and geometrical classification," *IEEE Trans. Info. Theory*, vol. 34, no. 5, pp. 1123–1151, 1988.

[3] G David Forney Jr, "Coset codes. ii. binary lattices and related codes," *IEEE Trans. Info. Theory*, vol. 34, no. 5, pp. 1152–1187, 1988.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

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

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

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

[24] 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.

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