GENERALIZED GAIN-SHAPE VECTOR QUANTIZATION FOR MULTISPECTRAL IMAGE CODING

Gerardo R. Canta, Luigi Paura**, Giovanni Poggi**

*D.I.E., Università di Napoli, via Claudio 21, I-80125 Napoli, Italy
Tel: +39 81 7683151; Fax: +39 81 7683149; E-mail: poggi@nadis.dis.unina.it

**D.I.I., II Università di Napoli, via Roma 29, I-81031 Aversa, Italy
Tel: +39 81 5044035; Fax: +39 81 5045804; E-mail: paura@nadis.dis.unina.it

ABSTRACT

This paper proposes a new encoding scheme that generalizes the Gain-Shape Vector Quantization technique and takes advantage of the distinctive features of multispectral images to encode them at very low bit rate, with a satisfactory reproduction quality and low complexity. Each codevector is obtained as the Kronecker product of a gain codevector and a shape codevector, which reduces both the memory requirements and codebook design complexity. Besides, the encoding complexity is also greatly reduced by resorting to a fast encoding algorithm.

1 INTRODUCTION

The use of multispectral images for a wide variety of applications concerning earth resources (geology, pollution monitoring, meteorology, to cite some) has been increasing steadily in recent years. A multispectral image is composed by a large number of bands each of which portrays the same subject, typically a region of the earth, in different spectral windows. Their transmission and/or storage is a major problem due to the huge amount of data involved. To obtain an appropriate data reduction, one has to resort to lossy compression techniques, introducing some degradation in image quality. Several coding schemes have been proposed in the last few years which try to exploit both the high level of spatial redundancy and the strong dependency among different spectral bands to achieve high compression levels. Many of them [1-3] rely on transform coding to exploit the inter-intraband dependency: Karhunen-Loewe transform, DCT, and wavelet transform have all been experimented with in various combinations. Other algorithms [4,5] resort to vector quantization (VQ) [6] which is known to be the optimal among all block coding techniques, including so all transform coding techniques. Unfortunately, its high computational complexity imposes a severe limitation on the block size that can be used and, hence, on the performance actually achievable.

To overcome such a limitation, one can resort to some form of constrained VQ, renouncing to strict optimality in favor of reduced complexity. As a consequence, one can use larger blocks, and exploit the statistical dependence among many more pixels, with the goal of overcompensating the lack of optimality.

In this paper, we propose a generalized gain-shape VQ (GSVQ) [7] technique that takes full advantage of the peculiarities of multispectral images to provide a satisfactory reproduction quality at a very low bit rate, and with a limited computational complexity. In particular, we build upon the working hypothesis that the shape of a block does not vary significantly from band to band and is well approximated by a constant vector, while all variations are concentrated in the gain term. In such an hypothesis, there is only a small-size shape codevector to be selected, common to all bands, plus a gain codevector that encodes the highly dependent gains, and is selected jointly with the shape codevector. The reduction of complexity with respect to unconstrained VQ is remarkable, also thanks to an ad hoc fast encoding procedure proposed here, and allows for the use of quite large blocks with beneficial effects on the rate-distortion performance.

2 BACKGROUND

A vector quantizer groups input samples in vectors[1], and represents each input vector by means of a template vector (or codevector) chosen from a finite collection called codebook. Let $\mathbf{x}$ be a generic input vector, and let $\mathcal{X} \equiv \{\widehat{\mathbf{x}}_1, \widehat{\mathbf{x}}_2, \ldots, \widehat{\mathbf{x}}_N\}$ be the codebook, then VQ performs the mapping

$$\mathcal{Q} : \mathbf{x} \rightarrow \widehat{\mathbf{x}}_m$$

where $\widehat{\mathbf{x}}_m$ minimizes a suitable distortion measure $d(\cdot,\cdot)$, usually the squared error, over the whole codebook

$$d(\mathbf{x}, \widehat{\mathbf{x}}_m) \leq d(\mathbf{x}, \widehat{\mathbf{x}}_n), \qquad n = 1, \ldots, N$$

Despite its simplicity, VQ can be shown to be the optimum block coding technique, namely, it provides

[1] We will often use block as a synonymous of vector in the following.

the lowest possible distortion for any given block size and bit rate. On the other hand, VQ is characterized by an encoding/memory complexity that, for any given bit rate, grows exponentially with the block size. This forces one to use only small blocks, and represents ultimately a limit on the performance. It turns out that, for a given computational complexity, suboptimal and hybrid VQ techniques can often outperform ordinary VQ by simply using much larger blocks. GSVQ is one of the most popular such techniques. It belongs to the class of product-code VQs in which the codebook is obtained as the cartesian product of two or more smaller codebooks. In gain-shape VQ, the norm (gain) of the input vector is first extracted and scalar quantized using a codebook of gains, while the unit-norm shape vector is quantized separately using a codebook of shapes. The input vector is then represented as the product of optimal gain and shape. This structuring leads to a significant saving in computation but also to some loss in performance with respect to unconstrained VQ. The bulk of the complexity lies in the VQ of the shape, but here a much smaller codebook than in unconstrained VQ can be used because of the reduced variability of the shape vectors. If gain and shape are only weakly dependent of each other, as is usually the case, the performance loss is quite limited [6] and amply compensated for by the opportunity of using much larger blocks.

3 PROPOSED TECHNIQUE

When applying GSVQ to multispectral images, one can take advantage of the fact that all the objects, edges, textures, are present in all the bands in the same positions. Therefore, it is reasonable to expect that the shape vector corresponding to a block in a given spatial position does not change much from band to band, while the gain term will change according to the variation of the reflectivity across the bands. This consideration leads us to a natural extension of GSVQ in which a multispectral block is represented by a fixed intraband shape vector modulated by an interband gain vector. To be more specific, let $\mathbf{X}$ be a multispectral input vector (we consider row vectors)

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_B)$$

obtained by collecting the vectors (blocks) $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})$ drawn from the same spatial location in the B bands of the image. We can easily extract a gain vector

$$\mathbf{g} = (g_1, g_2, \ldots, g_B)$$

with $g_i = \| \mathbf{x}_i \|$, and a multispectral shape vector

$$\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_B)$$

such that $\mathbf{x}_i = g_i \mathbf{s}_i$. Then we quantize both the gain vector and the multispectral shape vector, using a gain codebook $\mathcal{G} = \{\hat{\mathbf{g}}_i \ , \ i = 1, \ldots, N_G\}$ and a shape codebook $\mathcal{S} = \{\hat{\S}_j \ , \ j = 1, \ldots, N_S\}$; for the latter however we constrain the codevectors to have the form

$$\widehat{\mathbf{S}}_{\mathbf{j}} = (\widehat{\mathbf{s}}_j, \widehat{\mathbf{s}}_j, \ldots, \widehat{\mathbf{s}}_j)$$

namely, to consist of a single intraband shape replicated B times. This constraint is justified by the shape-invariance assumption (to be validated by numerical experiments) and leads to a significant reduction of both the memory requirement and the design complexity for the shape codebook. Gain and shape codebooks can be jointly optimized using a well-known algorithm [7] derived by the generalized Lloyd algorithm [8].

Therefore, for each input vector $\mathbf{X}$ the encoder has to choose the pair of codevectors $(\widehat{\mathbf{g}}_i, \widehat{\mathbf{s}}_j)$ that minimize the squared error distortion measure:

$$\| \mathbf{X} - \widehat{\mathbf{X}} \|^2 = \| \mathbf{X} - \widehat{\mathbf{g}}_i \otimes \widehat{\mathbf{s}}_j \|^2$$

where $\otimes$ denotes Kronecker product. In other words, we are carrying out VQ by means of a product code in which the $N_S \times N_G$ codevectors are all obtained as Kronecker products of some gain and shape codevectors. This codebook structuring appears to model accurately the source and allows one to use a large block size and to obtain, consequently, a significant performance improvement.

It is important to realize that, even for relatively high values of the multispectral block size, this encoding strategy has an affordable complexity. Indeed, we can write the distortion measure as follows:

$$\begin{aligned} \| \mathbf{X} - \widehat{\mathbf{X}} \|^2 &= \| \mathbf{X} \|^2 + \| \widehat{\mathbf{g}}_i \|^2 - 2 < \mathbf{X}, \widehat{\mathbf{g}}_i \otimes \widehat{\mathbf{s}}_j > \\ &= \| \mathbf{X} \|^2 + \| \widehat{\mathbf{g}}_i \|^2 - 2 < \widehat{\mathbf{z}}_j, \widehat{\mathbf{g}}_i > \end{aligned}$$

where the fact that $\| \widehat{\mathbf{s}}_j \| = 1$ has been accounted for, and

$$\widehat{\mathbf{z}}_j \triangleq (< \widehat{\mathbf{s}}_j, \mathbf{x}_1 >, < \widehat{\mathbf{s}}_j, \mathbf{x}_2 >, \ldots, < \widehat{\mathbf{s}}_j, \mathbf{x}_B >)$$

Therefore, to encode a size KB multispectral vector, the encoder has to:

1. evaluate the N_S vectors $\widehat{\mathbf{z}}_j$ ($N_S BK$ products);
2. maximize over j the scalar product $< \widehat{\mathbf{z}}_j, \widehat{\mathbf{g}}_i >$ for all i's ($N_G N_S B$ products);
3. minimize over i the sum $\| \widehat{\mathbf{g}}_i \|^2 - 2 < \widehat{\mathbf{z}}_{opt}, \widehat{\mathbf{g}}_i >$ (negligible complexity).

The overall per-pixel complexity is therefore proportional to $N_S N_G / K + N_S$ as opposed to $N_S N_G$ for an unstructured codebook of the same size, certainly a considerable reduction. As an example, if K=16 (blocks of 4x4 pixels) a saving of about an order of magnitude is obtained.

Even so, the encoding complexity can be prohibitive if one is interested in very large block sizes, namely, if one tries to encode many bands at once. However, it is possible to reduce further the encoding complexity by resorting to a suitable fast procedure.

Fast encoding

The second step of the encoding procedure, namely the evaluation of the $N_S N_G$ scalar products $< \widehat{\mathbf{z}}_j, \widehat{\mathbf{g}}_i >$, is responsible for most of the overall computational complexity. However, one can avoid computing but a small part of these products.

We look for the minimum over i and j of the function

$$V_{ij} \stackrel{\triangle}{=} \| \widehat{\mathbf{g}}_i \|^2 - 2 < \widehat{\mathbf{z}}_j, \widehat{\mathbf{g}}_i >$$

Applying the Schwartz inequality it results

$$V_{ij} \geq \| \widehat{\mathbf{g}}_i \|^2 - 2 \| \widehat{\mathbf{g}}_i \| \cdot \| \widehat{\mathbf{z}}_j \| \stackrel{\triangle}{=} Q_{ij}$$

so, given the current minimum $V_{\min}$ it is possible to skip all the pairs i, j for which

$$V_{ij} \geq Q_{ij} \geq V_{\min}$$

Of course, Q_{ij} is simpler to evaluate than V_{ij}. Moreover, for $z > 0$ and $g > 0$, the function $g^2 - 2gz$ increases with decreasing z and, for each z, with increasing $|g - z|$. Based on these considerations, it is possible to implement many different fast search algorithms. We proceed as follows:

- sort the $\widehat{\mathbf{z}}_j$ for decreasing norm (this can be done with little overhead while they are evaluated);
  sort the $\widehat{\mathbf{g}}_i$ for increasing norm (this can be done off-line);
- find $V_{\min}$, the minimum over i of V_{i1};
- find the first and last indexes i, say i' and i'', for which $Q_{i1} \geq V_{\min}$; all i's outside this interval can be discarded since for them $V_{ij} \geq Q_{ij} \geq Q_{i1} \geq V_{\min}$;
- for each index $i' < i < i''$, find the first index j, say $j'(i)$, for which $Q_{ij} \geq V_{\min}$; again, all j's larger than $j'(i)$ can be discarded.

Using this algorithm, the complexity of the second and third steps of the encoding procedure is very much reduced. In fact, by evaluating V_{ij} for the maximum-norm $\widehat{\mathbf{z}}$ vector, one finds very often the global minimum of V_{ij} and always a good approximation of it, so the majority of the candidates can be discarded right away. Therefore, the overall encoding complexity reduces about to that of the first step alone, namely, to that of a full-search VQ with a codebook of size N_S.

Moreover, having observed that the maximum-norm $\widehat{\mathbf{z}}$ vector is very often optimal, and always close to the optimal, we also implemented an approximate search procedure consisting only of the first two steps listed above, namely, we select j such that $\| \widehat{\mathbf{z}}_j \|$ is maximum and, given j, we select i so as to minimize V_{ij}.

4 NUMERICAL RESULTS

To assess the performance of the proposed coding scheme, we carried out some experiments using an hyperspectral image acquired by the GER (Geophysical Environmental Research) airborne sensor which portrays an agricultural area in Germany near the river Rhein. The image is composed by 63 bands in the range 0.4-2.5 μm, each band consisting of 1953 lines of 512 pixels quantized at 16 bit per pixel (bpp). In the experiments we use 8 bands (from 10 to 17) extracted from the first 24 bands that have constant spectral resolution and 8 bpp of meaningful information, and consider only a square 512x512 region as a test set, using the rest of the image as a training set to design all the codebooks. Fig.1 shows band 12 of the test image. The encoding performance, is evaluated in terms of average SNR (on all the bands) versus average bit rate.

Fig.1: band 12 of test image

In Fig.2 we report numerical results obtained using blocks of 4x4 pixels (K=16). We designed gain codebooks of size $N_G = 16, 32, \ldots, 256$, and shape codebooks of size $N_S = 8, 16, \ldots, 1024$, and used them in all possible combinations. The rate goes from 0.054 to 0.140 bpp. At a rate of about 0.12 bpp (a compression ratio of almost 70) the average SNR already exceeds 23 dB, to reach a maximum of about 23.3 dB at 0.14 bpp. Note that for the large majority of the blocks the optimal shape corresponded to the maximum-norm $\widehat{\mathbf{z}}$ vector: as a consequence, the fast encoding algorithm turned out to be very effective resulting in a perfectly manageable encoding complexity. Moreover, the lossy version produced results virtually identical to those of the exact version.

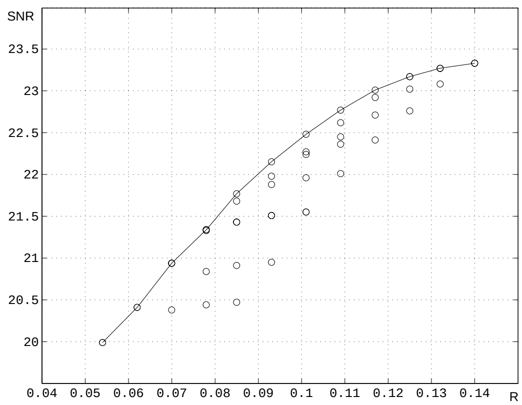

Fig.2: rate-distortion performance

To get some insight about the significance of these results, we implemented and run on the same set of bands the FPVQ algorithm proposed by Gupta and Gersho in [5]. To reach the same level of SNR obtained by the proposed technique, FPVQ required more than double the rate, and both design complexity and encoding complexity were considerably higher.

As the subjective judgement is often a more valuable (and precise) measure of the encoding quality than the SNR is, in Fig.3 and Fig.4 we show band 12 reconstructed after quantization at rate 0.14 bpp (SNR of 23.3) and 0.11 bpp (SNR of 22.6 dB).

Fig.3: band 12 encoded at 0.14 bpp

It seems safe to say that both images exhibit a good reproduction quality, which justifies a posteriori the shape invariance assumption on which the whole technique relies. In particular, there is so little difference between the two images that it appears more advantageous to use the generalized GSVQ algorithms at very low rates, whenever the quality requirements are not too stringent.

Fig.4: band 12 encoded at 0.11 bpp

REFERENCES

[1] J.A.Saghri, A.G.Tescher, "Near lossless bandwidth compression for radiometric data," *Optical Engineering*, vol.30, pp.934-939, July 1991.

[2] B.R.Epstein, R.Hingorani, J.M.Shapiro, M.Czigler, "Multispectral KLT-wavelet data compression for Landsat thematic mapper images," *Proc. Data Compression Conf.*, pp.200-208, Apr. 1992.

[3] G.P.Abousleman, M.W.Marcellin, B.R.Hunt, "Compression of hyperspectral imagery using the 3-d DCT and hybrid DPCM/DCT," *IEEE Trans. Geosci. Remote Sensing*, vol.33, pp.26-34, Jan 1995.

[4] R.L.Baker, Y.T.Tse, "Compression of high spectral resolution imagery," *Proc. SPIE,* no.974, pp.255-264, 1988.

[5] S.Gupta, A.Gersho, "Feature predictive vector quantization of multispectral images," *IEEE Trans. Geosci. Remote Sensing*, vol.30, pp.491-501, May 1992.

[6] A.Gersho, R.M.Gray, *"Vector quantization and signal compression,"* Kluwer Academic Publishers, Boston, 1992.

[7] M.J.Sabin, R.M.Gray, "Product code vector quantizers for waveform and voice coding," *IEEE Trans. Acoust. Speech Signal Process.*, vol.32, pp.474-488, June 1984.

[8] Y.Linde, A.Buzo, R.M.Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol COM-28, pp.84-95, Jan. 1980.