AN UNSUPERVISED SEGMENTATION-BASED CODER FOR MULTISPECTRAL IMAGES

M.Cagnazzo, L.Cicala, G.Poggi, G.Scarpa, L.Verdoliva

Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni
Università Federico II di Napoli, via Claudio, 21 – 80125 Napoli, Italy
{cagnazzo, lucacicala, poggi, giscarpa, verdoliv}@unina.it

ABSTRACT

To fully exploit the capabilities of satellite-borne multi/-hyperspectral sensors, some form of image compression is required. The Gelli-Poggi coder [1], based on segmentation and class-based transform coding, has a very competitive performance, but requires some a-priori knowledge which is not available on-board. In this paper we propose a new version of the Gelli-Poggi coder which is fully unsupervised, and therefore suited for use on-board a satellite, and presents a better performance than the original. Numerical experiments on test multispectral images validate the proposed technique.

Key-words: Multispectral image coding, region-based coding, on-board implementation.

1. INTRODUCTION

The performance of satellite-borne sensors increases ever more in terms of spatial resolution, radiometric accuracy, and number of spectral bands. All these aspects, and especially the latter, contribute to increase the data volume that such sensors must transmit to the ground station, to the point that the required data rate largely exceeds the available channel capacity and large chunks of data must be simply discarded. To avoid this loss one can resort to data compression which allows one to reduce the data volume by one/two orders of magnitude without serious effects on the image quality and on their diagnostic value for subsequent automatic processing. To this end, however, one cannot resort to general purpose techniques as they do not exploit the peculiar features of multispectral remote-sensing images, and in fact several ad hoc coding schemes have been proposed in recent years, e.g., [1-4].

One of the most promising such schemes, based on classified transform coding, is the Gelli-Poggi coder, originally proposed in [1]. The image is first segmented, so that each pixel is associated with one of a given number of classes based on its spectral response vector. Then, all vectors of the same class are grouped together and compressed by means of transform coding techniques. This way, transform coding operates on stationary homogeneous sources, thereby maximizing its efficiency, and leading to an excellent overall rate-distortion performance, which is in fact superior to that of other state-of-the-art coders.

The Gelli-Poggi coder, however, relies heavily on a-priori information which is hardly available to both encoder and decoder, and makes the coder unsuited for compression on-board a satellite before transmission to the ground station. In this paper we address this problem, by suitably modifying the various steps of the original coder in order to obtain more practical coding schemes suited for on-board operations. Next Section describes the Gelli-Poggi coder in detail, highlighting its weak points. Section 3 presents the various improvements proposed and Section 4 assesses the performance of the various alternative schemes by means of numerical experiments on test multispectral images. Finally Section 5 draws conclusions.

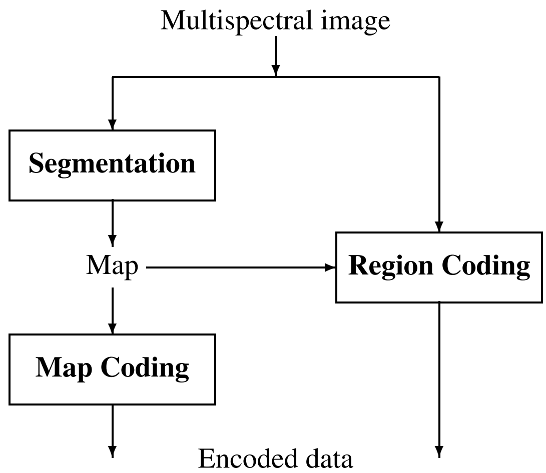

Figure 1: Block diagram of the coding scheme.

2. THE GELLI-POGGI CODER

The coding scheme comprises three main steps (see Fig.1):

1. image segmentation;
2. lossless coding of the segmentation map;
3. lossy coding of the radiometric information.

The original scheme is fully supervised, meaning that all statistical parameters are computed in advance on a training-set. Let us describe these three steps in some more detail.

Segmentation amounts to a simple spectral clustering. Specifically, each pixel is classified by computing the Euclidean distances between its spectral vector and a set of template vectors, one for each class, and assigning the pixel to the minimum-distance class. The set of template vectors can be viewed as a VQ codebook, computed off-line on a suitable training set, and the segmentation itself as a vector quantization. In particular, to limit computation complexity, the VQ codebook is tree-structured so that only a few binary comparisons are needed.

The map of class indexes must be sent to the decoder as a side information. Since neighboring pixels are highly correlated the map is significantly compressed, without loss of information, by resorting to a predictive scheme followed by Huffman coding, with the code computed on the training set as well.

Using the selected template vectors for every pixel instead of the original spectral vectors, we have a first VQ approximation of the multispectral image. The difference between the original image and the VQ approximation is the residual image, which is compressed by means of transform coding. First, a classified Karhunen-Loeve Transform (KLT) is performed along the spectral dimension. In order to account for class information, a different transformation matrix for each class is derived off-line from the training-set. Then, a Discrete Cosine Transform (DCT) is used to decorrelate the spatial information within each transformed band. Finally, each transform coefficient is sorted by spectral class, KLT band and DCT frequency, and is included in a quantization set which is quantized by a specific tree-structured Lloyd-Max quantizer designed off-line on the training set. Rate allocation is decided on-line with a greedy bit allocation algorithm.

3. THE UNSUPERVISED VERSION

The obvious weakness of the original Gelli-Poggi coder in view of on-board implementation is that several pieces of information are supposed to be known in advance, that is

- the VQ classifier;
- the class-adapted KLT matrices;
- the set-adaptive Lloyd-Max quantizers.

We will therefore abandon this hypotheses and consider an alternative coding scheme in which all needed parameters are designed on-line based on the very same data to be encoded. Of course, with respect to the original scheme, this entails an increase in computational complexity, an increase in the side information to be transmitted along with the quantized coefficients and, on the pros side, a different and possibly superior compression ability, since all parameters are now tuned on the data. We will examine the new steps in turn under these points of view.

3.1 The VQ classifier

The design of a VQ codebook can be very demanding in terms of CPU power but, since only a limited number of land covers are typically present in a given image, we are interested in a rather small codebook (e.g., 4 to 20 classes [5]), which largely reduces computation time. In addition, our codebook is tree-structured, which further reduces both design and segmentation complexity. Finally, the design needs not be carried out on all the data to be encoded, but only on a sample subset, which can be as small as a few thousands spectral vectors, although extreme subsampling can produce some performance losses. All in all, computational complexity is likely not an issue for the VQ classifier.

As for the side information, the VQ codebook for C classes is composed of C vectors, with B components each if B is the number of bands in the image. For images in the order of 1 Mpixel, and coding rates not unreasonably small, this cost is always negligible, even when 16 bits are spent to encode each vector.

On the contrary, a good codebook designed on-line can be significantly superior to its off-line counterpart, since in the latter case the training set is not guaranteed to fit well the actual data, so we can expect some performance gain here.

3.2 The class-adapted KLT matrices

To compute a KL transform matrix, we must first estimate the $B \times B$ correlation matrix of the data, and then compute its eigenvectors. Since we use class-adaptive KLT, we need C such matrices, one for each class.

The estimation part is not extremely demanding, especially if we resort again (with due care) to some subsampling of the training data. Computing the eigenvectors, instead, has a computational complexity which grows as the third power of the number of bands, and therefore can become a problem if B is very large. In this case the image can be segmented spectrally in smaller groups of bands without significant loss of performance, also because precision could be an issue for matrices that large. On the other hand, if B is large, almost all of the image energy is compacted in the first few transform coefficients, to the point that the less significant coefficients are assigned no bits at all. This suggests us to resort to low-complexity iterative techniques, such as the power method, to compute the B' most relevant eigenvectors which comprise almost all the energy (say, 99.9%). This condition can be tested on-the-fly, and helps limiting complexity in critical cases.

Concerning the side information, for each KLT matrix we must send $B \times (B+1)/2$ parameters in the conventional case, and approximately $B' \times B$ coefficient in the reduced dimensionality version. In some non-typical conditions (small images, very low coding rates, many classes, many bands) this could become significant and some care must be taken to encode all parameters with as few bits as possible, without significant performance losses.

Again, barring the case of a poor subsampling of the data, using matrices computed on-line cannot but improve the compression performance.

3.3 The set-adaptive quantizers

The problem, here, is that a very large number of quantizers are needed, $C \times B \times K$ in the most general case, with K the DCT vector length. In fact, an ad hoc quantizer is used for the first DCT coefficient of the first KLT band of the first class, another one for the second DCT coefficient of the first KLT band of the first class, and so on. Even considering that most of these sets of coefficients will be assigned no encoding bits, and no information needs be transmitted for them, so many quantizers remain to be designed and transmitted that this approach becomes clearly unreasonable. We resort therefore to parametric quantizers: each set of coefficients is modeled as a zero-mean generalized Gaussian, characterized by its variance and shape parameters, which univocally identify the optimal quantizer. To preserve the scalability of the original scheme however, we designed embedded quantizers and we made them mid-tread to increase robustness, so that the tree structure has, at every depth level, one ternary node besides the binary ones. The coefficient variances are then used to perform rate allocation by means of the Huang-Schultheiss algorithm [5].

With the new quantization scheme, the computational burden increases very little, because only synthetic statistics

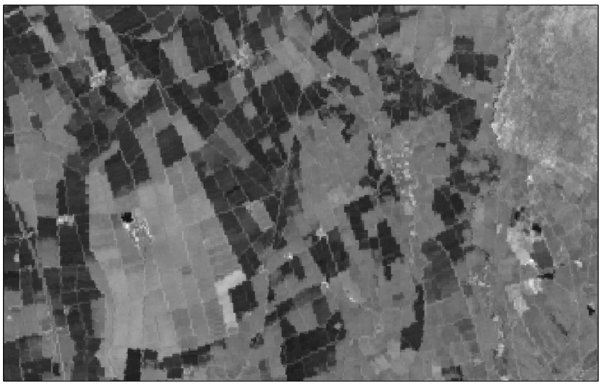

Figure 2: Band 5 of the test image.

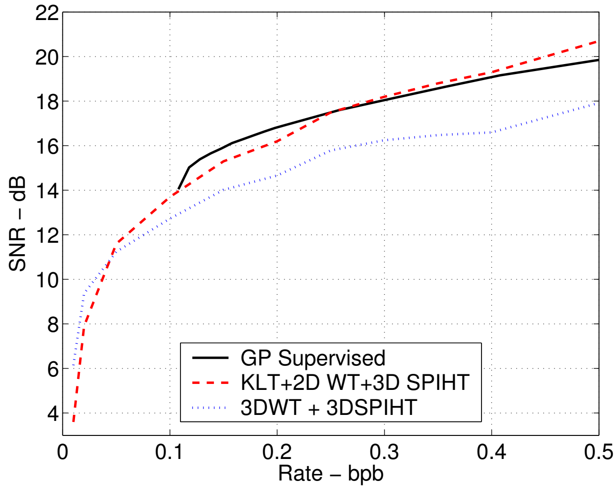

Figure 3: Gelli-Poggi vs. wavelet-based coders.

must be computed for the data. Two pieces of side information must to be sent: the active/inactive bits for each set, and the pdf parameters for the active sets only, which should be quite limited as well.

In terms of performance, parametric quantizers do not guarantee in theory the same accuracy of the optimal Lloyd-Max, but the design on actual data might even offset this theoretical disadvantage.

4. EXPERIMENTAL ANALYSIS

All experiments presented here are carried out on a multispectral image acquired by the LANDSAT sensor (882x448 pixels, 8 bit/pixel, 6 bands) which portrays an agricultural area in Italy near the river Po. A sample band of the test image is shown in Fig.2.

Preliminarily we analyze the absolute performance of the original Gelli-Poggi coder by comparing it with two state-of-the-art coders, one based on 3d-wavelet transform followed by 3d-SPIHT [3], and the other based on spectral KLT followed by 2d-wavelet on the transform bands and 3d-SPIHT.

For the Gelli-Poggi coder, all needed parameters are supposed to be known a-priori and are actually evaluated on a small training section of the same LANDSAT image. The rate-distortion curves are reported in Fig.3, and show that the Gelli-Poggi coder always outperforms the wavelet-based coder, and has a performance very close to that of the KLT-wavelet one, even outperforming the latter for some values of the bit-rates. Note that, contrary to the reference techniques, very small encoding rates are not allowed, since the first piece of information, the segmentation map, is encoded without loss of information. However, for remote-sensing applications such low-rates are not of interest and, in addition, the availability of an accurate segmentation map computed on the uncompressed data might be a valuable side product for many users.

We now turn to analyze the effects of the new unsupervised coder on computational complexity, side information and overall performance.

4.1 Complexity

First of all we must select the size of all training sets used in the design phase. After a series of preliminary experiments, we opted for a training set of $100 \cdot C$ vectors to design the VQ classifier (discovering again a well-known rule of thumb) which guarantees a loss always within 0.1 dB with respect to the full-size training set. Likewise, we decided to use 100 vectors per cluster to estimate the KLT matrices, for a loss always inferior to 0.1 dB, and finally, found that 25-50 samples are more than appropriate to estimate the variance of the quantization sets, with similarly negligible losses.

With these parameters, and considering C=20 classes and DCT blocks of K=64 pixels, we estimated the increase in complexity due to the on-line design in about 1.5 multiplications/pixel (m/p), almost evenly distributed among VQ, KLT and SQ design, to be compared to about 130 m/p required for the encoding phase. The additional cost is therefore in the order of 1%, obviously negligible. It goes by itself that this fraction changes with the image and coding parameters: for example, the additional cost increases (in relative terms) with decreasing image size, and increases with the number of classes and especially bands. In fact, the KLT design, as said before, becomes quite expensive for blocks of many bands and, for example, the numbers change to about 40 m/p for the design phase and 1800 m/p for the encoding phase if we encode jointly 36 bands, keeping all other parameters fixed. The design phase now costs proportionally more, about 2.5% of the total, but the main observation is that the encoding phase itself becomes very expensive, suggesting that such a situation should be avoided. These results are summarized in Fig.4, which plots the complexity of the various design steps in log scale vs. the number of bands, compared with the complexity of the encoding phase. We can therefore conclude this subsection saying that, for reasonable parameters, the on-line design phase has a fully affordable computational cost.

4.2 Side Information

A first piece of side information is the segmentation map, which however was already present in the original scheme and therefore will not be considered here. We are interested instead in the additional costs, namely, the cost of sending the VQ codebook, the KLT matrices, and the parameters of the active scalar quantizers (we use Laplace quantizers and therefore do not estimate the shape parameters). Considering the same experimental setting as before, and assuming that all parameters are encoded to 16 bit precision (not re-

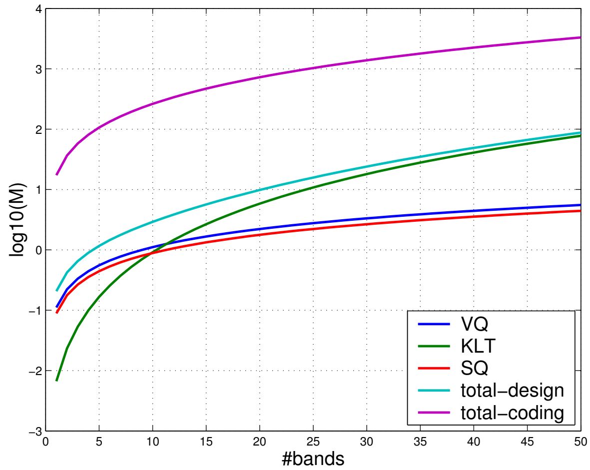

Figure 4: Complexity of design and encoding phases vs. number of bands.

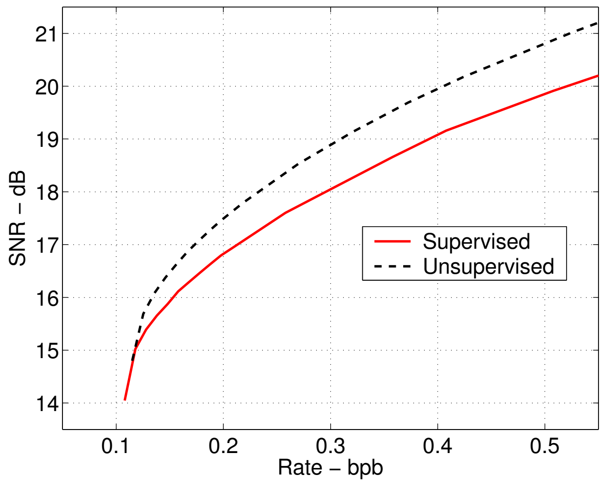

Figure 5: Supervised vs. unsupervised coders.

ally necessary), the cost for VQ codebook is less than 0.001 bit/pixel/band (b/p/b), KLT matrices cost about 0.003 b/p/b, and the active variances less than 0.005 b/p/b. At a coding rate of 0.25 b/p/b, which is rather low for remote-sensing application which require high fidelity, the side information accounts for less than 4% of the total coding cost. Such a cost obviously decreases when working at higher rates.

Again, it is interesting to consider the dependence on the coding parameters, and especially on the number of bands. Although an increase in the costs can be observed, it is not really significant. As an example, with C=20, and B=36, the side information requires less than 0.015 b/p/b which, at a coding rate of 0.25 b/p/b, is just 6% of the total rate.

4.3 Encoding quality

Finally, let us consider the variation in encoding quality when the supervised coder is used. In this case, rather than analyzing the various sources of differences in performance, which include the rate increase due to side information, the training set subsampling, the different quality of the VQ, and SQ codebooks and of KLT matrices, and are therefore quite complicate to tell apart, we will examine the overall rate-distortion performance when encoding the test image. Fig.5 compares the rate-distortion performance of the supervised Gelli-Poggi coder (solid line) and of the new fully unsupervised version (dashed line), in the experimental setting already described above. It can be seen that, despite the increased cost for side information, the unsupervised case exhibits a significant and increasing gain of about 1 dB w.r.t. the unsupervised coder. The reason for this success must be ascribed to the better quality of the codebook and of the KLT matrices which are now designed for the data they operate on. In particular, the scalar quantizers appear to be much better adapted to the true statistics of the image, and responsible for much of the performance gain

5. CONCLUSIONS

We set to implement an unsupervised version of the Gelli-Poggi coder for multispectral images with the goal of making it suitable for use on-board a satellite and thus reduce the problems encountered in the transmission to the ground stations. Although experiments have not been extensive thus far, they are very encouraging. Of course, the overall encoding time increases, but never more than a few percents w.r.t. the original supervised coder in the reasonable cases considered. In addition, we obtain a better rate-distortion performance since the increase in side information is very limited and more than compensated by the improved quality of encoding. Moreover, typical remote-sensing images are larger than the 882x448 section considered here, which goes in the direction of further reducing the cost of side information.

Experiments are under way to further improve the Gelli-Poggi coder by using a wavelet-based coder in the spatial domain for the coefficients produced by the classified KLT.

REFERENCES

[1] G.Gelli, G.Poggi, "Compression of multispectral images by spectral classification and transform coding", *IEEE Transactions on Image Processing*, Apr.1999, pp.476-489.

[2] J.A.Saghri, A.G.Tescher, J.T.Reagan, "Practical transform coding of multispectral imagery", *IEEE Signal Processing Magazine*, Jan.1995, pp.32-43.

[3] P.L.Dragotti, G.Poggi, A.R.P.Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm", *IEEE Transactions on Geoscience and Remote Sensing*, Jan. 2000, pp.416-428.

[4] H.S.Lee, N.H.Younan, R.L.King, "Hyperspectral image cube compression combining JPEG-2000 and spectral decorrelation", *2002-IEEE IGARSS*, Toronto (Canada), June 2002, pp.3317-3319.

[5] A.Gersho, R.M.Gray, *Vector quantization and signal compression*. Kluwer Academic Publisher, Boston, 1992.