Title
Tree-structured vector quantization with region-based classification

Authors
Perlmutter, S M
Perlmutter, K O
Cosman, P C
et al.

Publication Date
1992-10-01

Peer reviewed

Tree-Structured Vector Quantization with Region-Based Classification

Sharon M. Perlmutter Keren O. Perlmutter Pamela C. Cosman Eve A. Riskin[†]
Richard A. Olshen[‡] and Robert M. Gray

Information Systems Laboratory
Durand Building
Stanford University
Stanford, CA 94305-4055

†Dept. of Electrical Engineering
FT-10
University of Washington
Seattle, WA 98195

‡Division of Biostatistics
Health Research and Policy Bldg.
Stanford University
Stanford, CA 94305-5092

Abstract

Many classes of images possess a strong degree of spatial stationarity such that particular features of the image reliably appear in certain regions of the image. This spatial information can be used to improve compression. Unbalanced or Pruned Tree-structured Vector Quantization (PTSVQ) is a variable-rate coding technique that tends to use more bits to code active regions of the image, and fewer to code homogeneous ones. The PTSVQ is developed based on a training sequence of typical images. We used a regression tree algorithm to segment the images of the training sequence, using the x,y pixel location as a predictor for the intensity. This segmentation was used to partition the training data by region and generate separate codebooks for each region, and to allocate differing numbers of bits to the regions. Unlike other varieties of classified vector quantization, a region-based classification requires no side information as the decoder knows where in the image the current encoded block originated. These methods can enhance the perceptual quality of compressed images when compared with ordinary PTSVQ. Results are shown on magnetic resonance data.

1 Introduction

It is possible to enhance the perceptual quality of compressed images by applying a region-based classification scheme to exploit the spatial stationarity of certain images. In particular, techniques of generating codebooks and encoding test images can be improved with the *a priori* knowledge that particular features of an image reliably appear in certain regions of the image. This premise is investigated using separate training and test images from a set of 12 mid-sagittal MR brain scans of different individuals. The training sequence of 10 different MR images was segmented into regions using the CART^{TM} algorithm [2], a regression tree algorithm that can be used for predicting intensity from the x,y pixel location. Based on the CART parameters and the type of images used in this study, 7 regions were chosen for each training sequence. Figure 1 shows the average image of the ten images that composed a training sequence with the CART segmentation superimposed. For ease of implementation, we used a version of CART in which the splits were constrained to be parallel to the coordinate axes, but numerous other alternative segmentation algorithms could have been used for providing the initial segmentation. The training vectors are considered to be separated into 7 classes according to the region in which they originated.

In all cases, the quantizer considered was the pruned tree-structured vector quantizer (PTSVQ). The TSVQ was grown one node at a time in a "greedy" fashion as described in [3, 5, 6] and then pruned back using the generalized Breiman, Friedman, Olshen and Stone algorithm [2, 6]. The node centroids were determined by the Generalized Lloyd Algorithm. In all cases the vector was a 2×2 pixel block. The performance was measured by the mean-squared error (MSE) between the input image and the quantizer output at a given rate.

2 Methods

We examined four different approaches that used this region-based classification in codebook generation or encoding. These algorithms attempt to exploit one or both of the following two distinct features. First, since codebook generation algorithms attempt to cluster similar vectors, to the extent that the similarity of vectors depends upon their spatial locations, one can perhaps improve the codebooks. Secondly, in medical images some regions contain less important information than other regions, and thus bits can be allocated unequally to the regions. This idea of unequal bit allocation for different features is similar to the work done by Braccini *et al.* with classified VQ [1]. Their

1058-6393/92 $03.00 © 1992 IEEE

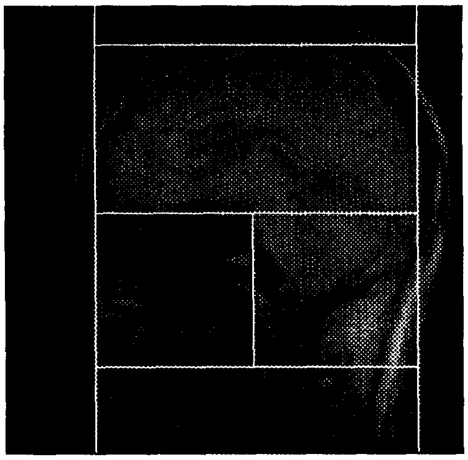

Figure 1: Training image with CART segmentation superimposed

work was applied to videophone images which generally consist of a frontal picture of a person's face, and the eyes and mouth were identified as the most important features of the images. Unlike classified vector quantization, however, the region-based classification scheme does not require side information to be sent to the decoder. The image is scanned as a raster, and so the decoder knows which region it is currently decoding.

The utility of the various approaches was tested in a cross-validated way [2, 4] by selecting 6 different training sets of 10 images each from the overall set of 12 images, and in each case testing on the two remaining images. Thus each of the 12 different test images was used one time as a test image and 5 times as a training image (when it was not used as a test image). We investigated the MSE and the perceptual quality at rates from 0.4 bpp to 1.5 bpp. The most important basis of comparison used to judge performance was the perceived quality of the reconstructed image, and in general more emphasis was placed on the quality of the cortex and cerebellum regions (the medically important regions of an MR brain scan). For a quantitative comparison, the overall distortions for each of the 12 test images were averaged and PSNRs were computed. For the unequal bit allocation approaches, the average distortions and PSNRs were also computed for the cortex and cerebellum regions.

2.1 Separate Codebooks for Each Region

Separate codebooks were generated for each of the 7 regions using the training data from that region from all of the training images. We then encoded each of the 7 regions of the test image using the corresponding codebook. The premise was that each region would be encoded with a codebook that had been generated from vectors more similar to its own. The question arose of how to allocate bits among the various regions. One could encode each region at the same average rate. However, when one uses an unbalanced TSVQ on an entire image of this type, the TSVQ tends to devote very few bits to the background, and many bits to the more active regions of the image. We would be foregoing this very desirable property of unbalanced TSVQ if we allocated bits so that background regions had the same average rate as active regions. As a simple but effective improvement upon this, we distributed the bits among the regions to be proportional to the average intensity of the training vectors from the regions. This comes from the *a priori* knowledge that for MR brain scans, important regions such as the cortex possess a relatively high average intensity, and less important regions such as background have a relatively low average intensity. A different approach could of course be used for different image sources.

Using this procedure, there was an improvement in the perceptual quality of the entire compressed image and an increase in PSNR at all bit rates, as compared to ordinary PTSVQ. Furthermore, the perceptual quality and PSNR in the medically important regions of the cortex and cerebellum improved. In general, there were larger improvements in the medically important regions as compared to ordinary PTSVQ at the higher rates. For example, there was a 2.3 dB increase in PSNR at 1.3 bpp.

There are some limitations, however, that may persist when using separate codebooks for each region. For example, the brain may be situated differently within the test image than the "average" brain of the training images. In this case, the regions of the test image may not correspond well to the regions of the training images, and the selected codebooks may not accurately reflect the contents of the test regions. Another issue is that the training sequence size of some regions may be too small. Because the 10 training images are divided into 7 distinct training sets, some sets may not have enough training data, which would entail a marked increase in distortion. [4].

2.2 Ordinary PTSVQ with Intensity-Based Bit Allocation

Only one codebook was generated from all of the training vectors of all 10 training images, and the region information was used purely for purposes of bit allocation. Regions of higher average intensity received more bits than regions of lower intensity. This was accomplished by using different pruned subtrees for the different regions. This method avoids the limitation of insufficient training data, and also largely avoids the problem of having a test brain spatially shifted from the average training sequence brain. An unbalanced tree may result in many bits being devoted to some input vectors, while a low average rate is maintained. If the active region of the test brain is located where the background is expected, even though the algorithm has few bits to devote to the background region, the active input vectors will be able to trace longer paths through the tree and find better codewords. The single codebook approach, however, does not benefit from the similarity of vectors within a region; region information is exploited only in the allocation of bits.

This approach yields higher PSNR in the cortex and cerebellum regions than does ordinary PTSVQ, which is expected because this method is merely a way to shunt bits away from the background and towards the medically important regions. In general, the increase in PSNR for these regions as compared to ordinary PTSVQ ranged from 0.5 to 3.3 dB. The larger improvements were at the higher bit rates.

2.3 Region-Specific Centroids

Only one codebook was constructed from all of the training vectors, but for each terminal node 7 separate region-specific centroids were computed from the training vectors that fell within the Voronoi region of the node. Each leaf stored both the overall centroid for the vectors that mapped into it, and the 7 region-specific centroids as illustrated in Figure 2. When a vector from a specific region of the test image was encoded, the centroid of the corresponding region within the selected terminal node was used as its codeword. If no training vectors from one of the regions mapped into a particular leaf, the overall centroid was substituted for that region centroid.

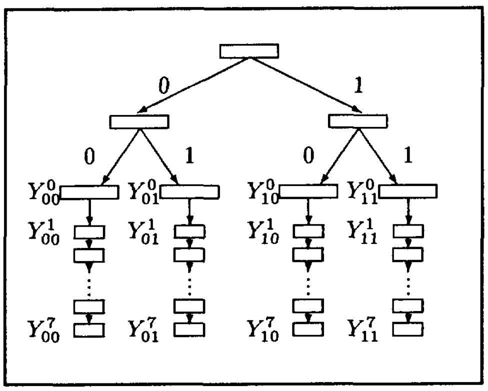

Figure 2: A VQ decoder with nodes containing both overall centroids (Y_i^0) and region-specific centroids (Y_i^1 to Y_i^7)

This method can be guaranteed to do no worse than ordinary PTSVQ for the price of only 7 additional bits per image. For each of the 7 regions, the encoder can encode the entire region using the overall centroids of the leaves (ordinary PTSVQ) and again using the region-specific centroids. The encoder can use a single bit for each region to inform the decoder which codeword set produced a lower MSE.

The region-specific centroid method performs better than ordinary PTSVQ for the entire image and for the cortex and cerebellum regions. In particular, at 0.4 bpp there was a 1.09 dB increase in PSNR of the entire image compared to ordinary PTSVQ. There is greater benefit to using region-specific centroids at lower rates than at higher rates, where the Voronoi regions of the leaves are small. Generating separate centroids within the small space does not produce codewords very different from the overall centroid for the node. Furthermore, since at high rates there are few training vectors mapping into a leaf, partitioning these vectors further into regions may provide too few training vectors to produce good codewords. At lower rates, however, the large Voronoi regions allow the centroids for particular image regions to differ significantly from the overall centroids. In addition, there are many more training vectors lying within each leaf.

2.4 Region-Specific Centroids with Intensity-Based Bit Allocation

The region segmentation was used both for making region specific centroids and for allocating bits with the intensity-based scheme described previously. As expected, this method improved the perceptual quality of the compressed test image in the cortex and cerebellum. Greater improvements in these regions occurred at the higher rates, as compared to ordinary PTSVQ. In particular, there was a 3.3 dB improvement of the average PSNR in the cortex and cerebellum regions at 1.47 bpp.

3 Discussion

Figures 3 and 4 compare the results obtained using ordinary PTSVQ with the approaches presented in this paper. In Figure 3 the x-axis indicates the average rate in bpp for the entire image and the y-axis indicates the average PSNR in dB of the cortex and cerebellum regions. In Figure 4 the x-axis indicates the average rate in bpp for the entire image and the y-axis represents the average PSNR in dB of the entire image.

Figures 3 and 4 also indicate that a tradeoff exists between the average PSNR obtained for the entire image and the average PSNR of the cortex and cerebellum regions when the intensity-based bit allocation schemes were employed. This is particularly evident when region-specific centroids are used, since only at the lower rates can an improvement in PSNR for both the entire image and the cortex and cerebellum be obtained. At the higher rates, because the benefits of region-specific centroids are considerably diminished, it is not possible to obtain simultaneously a higher PSNR for both the entire image and the medically important regions. The tradeoff is less significant when separate codebooks were generated for each region.

The determination of which method would be more effective also depends upon the type of images to be compressed. If the images contain regions of unequal importance, the region-specific centroids with

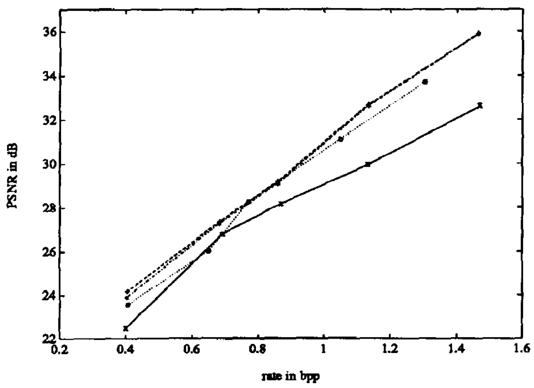

Figure 3: Comparison of PSNR of cortex and cerebellum regions vs. rate for the different methods: ordinary PTSVQ (solid line), separate codebooks (dotted line), ordinary PTSVQ with intensity-based bit allocation (dash-dot line), region-specific centroids with intensity-based bit allocation (dashed line)

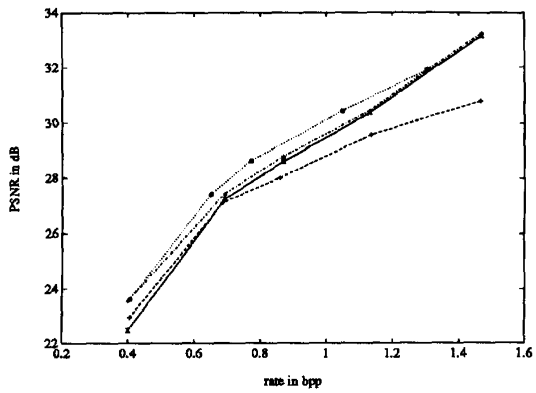

Figure 4: Comparison of PSNR of entire image vs. rate for the different methods: ordinary PTSVQ (solid line), separate codebooks (dotted line), region-specific centroids (dash-dot line), region-specific centroids with intensity-based bit allocation (dashed line)

intensity-based bit allocation scheme would be better suited. On the other hand, if it is important to maintain high quality for all regions of an image, separate codebooks for each region would be more applicable.

Figures 5, 6 and 7 illustrate a qualitative comparison of the perceptual quality of the reconstructed images. Figure 5 displays the original test image. Figure 6 depicts the image compressed to .870 bpp using ordinary PTSVQ. Figure 7 shows the image compressed to .875 bpp using region-specific centroids with intensity-weighted bit allocation.

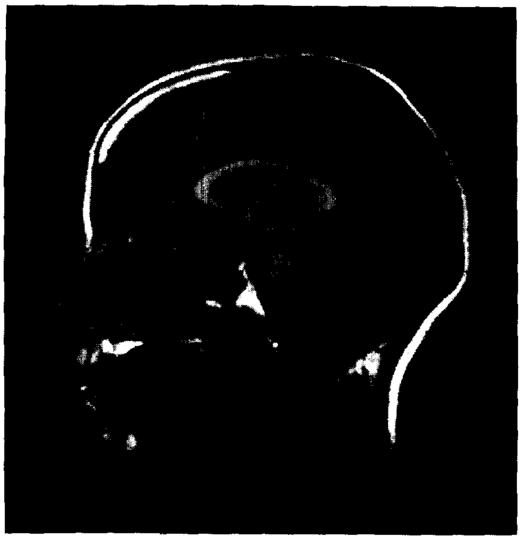

Figure 5: Original uncompressed image

4 Conclusions

Intensity-based bit allocation worked well in targeting the more medically important regions and assigning them more bits, and as such, there was a significant improvement in the cortex and cerebellum regions as compared to ordinary PTSVQ. When this bit allocation scheme for the encoding step was combined with methods for subdividing the training vectors into more similar groups for the codebook generation process, it was possible to obtain better performance in both the medically important regions and in the entire image, as compared to ordinary PTSVQ. Thus, it is possible to improve the quality of compressed images for those images that are spatially stationary, such as medical and videophone images, by incorporating region information into the processes of generating codebooks and encoding.

Figure 6: Reconstructed image using ordinary PTSVQ

Figure 7: Reconstructed image using region-specific centroids with intensity-based bit allocation

5 Acknowledgments

This work was supported in part by the National Institutes of Health under Grants CA49697-02 and CA55325-01, by the National Science Foundation under Grants DMS-9101548 and MIP-9110508 and by National Science Foundation Graduate Fellowships.

References

[1] C. Braccini, A. Grattarola, F. Lavagetto, and S. Zappatore. VQ coding for videophone applications adopting knowledge-based techniques: Implementation on parallel architectures. *European Transactions on Telecommunications*, 3(2):137–144, Mar.-Apr. 1992.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series. Wadsworth, Belmont, California, 1984.

[3] P. A. Chou, T. Lookabaugh, and R. M. Gray. Optimal pruning with applications to tree-structured source coding and modeling. *IEEE Transactions on Information Theory*, 35(2):299 – 315, March 1989.

[4] P.C. Cosman, K.O. Perlmutter, S.M. Perlmutter, R.M. Gray, and R.A. Olshen. Training sequence size and vector quantizer perfomance. In *Proc. Twenty-fifth Asilomar Conference on Signals, Systems and Computers*, 1991. Pacific Grove, CA, Nov. 1991.

[5] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, 1992.

[6] E. A. Riskin and R. M. Gray. A greedy tree growing algorithm for the design of variable rate vector quantizers. *IEEE Transactions on Signal Processing*, 39:2500–2514, November 1991.