

HIGH ORDER ENTROPY-CONSTRAINED RESIDUAL VQ FOR LOSSLESS COMPRESSION OF IMAGES

Faouzi Kossentini and Mark J. T. Smith

School of ECE, Georgia Institute of Technology, Atlanta GA 30332-0250

Allen Scales

Nichols Research Corporation, MS 1-1-174, 4040 S. Memorial Pkwy, Huntsville AL 35802

Abstract—High order entropy coding is a powerful technique for exploiting high order statistical dependencies. However, the exponentially high complexity associated with such a method often discourages its use. In this paper, an entropy-constrained residual vector quantization method is proposed for lossless compression of images. The method consists of first quantizing the input image using a high order entropy-constrained residual vector quantizer and then coding the residual image using a first order entropy coder. The distortion measure used in the entropy-constrained optimization is essentially the first order entropy of the residual image. Experimental results show very competitive performance.

I. INTRODUCTION

A common approach to lossless image coding is to preprocess the data, in a way that removes statistical dependencies among the input symbols, and code those symbols with an entropy coder. Individual systems differ in their choice of statistical models for removing redundancies and their choice of entropy coders, like arithmetic and Huffman for example. Simple statistical models such as DPCM can remove some of the dependencies but usually are ineffective in handling high order dependencies.

High order statistical models have been proposed previously for lossless compression of binary images [1], and were shown to be very effective. Unfortunately, they cannot be translated efficiently to the gray-scale case. The computational and storage demands can be prohibitive. For example, a typical first order conditional statistical model might require that 65535 conditional probabilities be computed and stored. This number grows exponentially with increasing model order. Compounding the problem is the fact that many of the probability tables cannot be populated even when large training sequences are used, making high order entropy coding a very difficult task.

Several methods have been proposed recently for re-

ducing the complexity of these statistical models [2, 3, 4, 5]. Most employ quantization or merging principles to reduce the number of conditioning states or tables of conditional probabilities, usually leading to orders of magnitude reductions in complexity while sacrificing only a small loss in performance. Others involve decomposition of the original signal into binary signals, which increases the accuracy of estimating the statistical model and thus improves the compression performance. In this paper, we introduce a new method that is based on both decomposition and probability table reduction techniques. Statistical modeling is performed through high order conditional entropy-constrained residual vector quantization (CEC-RVQ) [6, 7]. The entropy-based distortion measure employed in the CEC-RVQ optimization coupled with the high order entropy coding of the CEC-RVQ output result in substantial reductions in the entropy of the residual signal. This design framework, leads to high compression performance relative to other competing approaches.

II. PROPOSED FRAMEWORK

The hybrid technique of quantization and entropy coding of the residual signal has been shown to yield good compression performance [8, 9, 10]. This is due to the fact that quantization often produces a structure where high order statistical dependencies can be exploited. Moreover, since the output alphabet of the quantizer can be made smaller than that of the original signal, the complexity of high order statistical modeling is reduced. This is especially the case when structurally constrained quantizers are employed. In particular, the structure of the multistage residual vector quantization (RVQ) used here has been shown [11] to be very successful in providing more accurate estimates of the statistical dependencies of the original signal while also reducing drastically the complexity of high order statistical modeling. Multistage RVQ produces multiresolution approximations of the input signal, and allows high order statistical conditioning to be performed between the stage sub-signals.

As shown in Figure 1, we employ a CEC-RVQ to

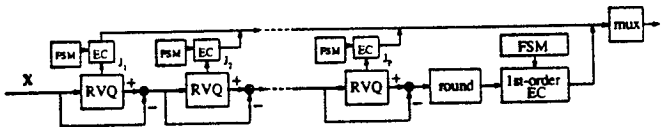


Figure 1: Proposed CEC-RVQ lossless coder.

quantize the input signal, where the output of the stage RVQ is then fed into a statistical-model-driven entropy coder (EC). The high order stage statistical model is represented by a finite-state machine (FSM) where the state transitions are based on previously coded symbols. The quantized signal is *rounded* to the nearest integer, and the residual signal, formed by subtracting the rounded quantized signal from the original one, is then coded using a first order entropy coder. Empirical work has shown that using higher order entropy coding does not lead to significant reductions in output entropy of the residual signal. In the final stage of the encoder, the bits emanating from the stage entropy coders as well as the residual entropy coder are combined together into a uniquely decodable bit stream, which is sent to the channel.

There are two important ideas, unique to this framework, that exemplify the novelty of this lossless approach. First, since the overall system is lossless, it is potentially better to employ the entropy of the residual signal as a distortion measure in the design of the CEC-RVQ. Using conventional distortion measures such as the squared error measure does not lead to minimization of the residual entropy. To elaborate, let \mathbf{x} be the input and $\hat{\mathbf{x}}$ be the output of the CEC-RVQ. The new distortion measure used in the design of the CEC-RVQ is $d(\mathbf{x}, \hat{\mathbf{x}}) = -\log_2[\text{pr}(I(\mathbf{x} - \hat{\mathbf{x}}))]$, where $I(\alpha)$ is the integer closest to the real α . The distortion is essentially the self-information of the integer-converted residual signal, and is used as an estimate of the length of the codeword that would be used to encode the symbol $I(\mathbf{x} - \hat{\mathbf{x}})$. In other words, the CEC-RVQ designed to minimize such a distortion measure also minimizes the entropy¹ of the residual signal.

The second idea is that only entropy is a measure of performance. Since the distortion measure is the entropy, the CEC-RVQ design algorithm produces an operational *entropy-entropy* curve where each point represents a pair of entropies, the first being the high order entropy h_o of the CEC-RVQ and the second being the entropy h_r of the residual. The high order entropy h_o is obtained by $h_o = \mathcal{H}(h_r)$, where \mathcal{H} is the operational entropy-entropy function. It can be easily shown that the function $\mathcal{H}(h_r)$ is continuous and differentiable (except for some points). However, it is generally not convex, and its convexity depends on the source as well as the entropy measure used to estimate the information content in the residual sig-

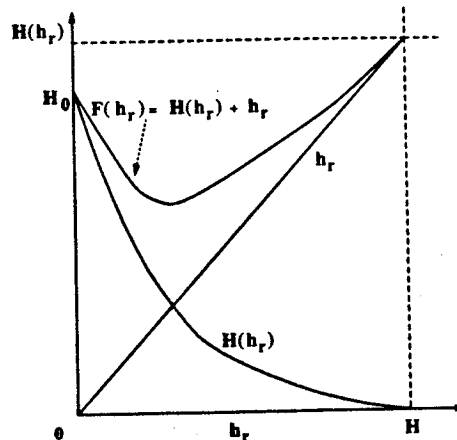


Figure 2: Illustration of an operational entropy-entropy curve.

nal. Fortunately, experimental work shows that for natural images and the first-order entropy, the function $\mathcal{H}(h_r)$ is convex with endpoints H_o and H , as illustrated in Figure (2). In the figure, the right endpoint H is the first-order entropy of the original signal. The left endpoint $H_o = \mathcal{H}(0)$ is the high order entropy of CEC-RVQ which results in perfect reconstruction after the CEC-RVQ output is rounded to the nearest integer. Due to the monotonicity of the CEC-RVQ (i.e., distortion will, on the average, only decrease by adding RVQ stages), H_o is finite. In other words, there is a point beyond which all of the real components of the residual signal lie in the real interval $(-0.5, 0.5)$. The problem at hand is to find an $(\mathcal{H}(h_r), h_r)$ pair such that the function $\mathcal{F}(h_r) = \mathcal{H}(h_r) + h_r$ is minimized. As shown in the figure, the minimum occurs at h_r^* such that $\mathcal{H}'(h_r^*) = -1$. As will be shown later, the CEC-RVQ algorithm is based on a Lagrangian minimization where λ is the slope of the operational entropy-entropy function \mathcal{H} . Thus, the problem translates into designing the EC-RVQ with corresponding Lagrangian parameter lying in the neighborhood of 1.

Note that \mathcal{F} would not necessarily have a minimum at h_r^* if \mathcal{H} were not convex. Moreover, it is implied in Figure 2 that $H_o < H$. This is not true in general, since H depends on the source and H_o depends on the source, quantizer, and quantizer output statistical model. If $H_o \geq H$, the minimum may be larger or equal to the entropy H , and quantization becomes useless. However, by using CEC-RVQ, it is observed that H_o is usually significantly smaller than H . Thus, CEC-RVQ has the potential of achieving rates that are substantially lower than those obtained by first order entropy coding the original signal.

III. DESIGN AND COMPLEXITY ISSUES

The CEC-RVQ design algorithm proposed here iteratively minimizes the Lagrangian

$$J_\lambda = E[-\log_2 \text{pr}(I(\mathbf{X} - \hat{\mathbf{X}}))] + \lambda E[\ell(\mathbf{L}(J|U))],$$

¹This is the first-order entropy. For higher order entropies, high order probabilities should be used in the distortion measure.

where U is the state random variable [6], \mathbf{L} is the high order conditional entropy mapping, and $\ell(\mathbf{L}(J|U))$ is the length of the variable length codeword $\mathbf{L}(J|U)$. The Lagrangian parameter λ controls the entropy-entropy trade-offs and is used in the design process to locate on the operational entropy-entropy curve the point where the sum of the entropies is a minimum or close to a minimum.

In this work, a training sequence that is representative of the source output to be encoded is used in the design process. Let \mathbf{x}^i be the i th k -dimensional vector taken from the training sequence of size N . An optimal encoding optimization step generally requires exhaustively searching the reproduction vector $\hat{\mathbf{x}}^*$ that minimizes the Lagrangian $-\log_2 \text{pr}(I(\mathbf{x}^i - \hat{\mathbf{x}}^*)) + \lambda(-\log_2 \text{pr}(j|u))$, where j is the current output of the CEC-RVQ and $u \in U$ is the current conditioning state. This typically yields large encoding complexity. To reduce complexity, non-exhaustive stage searching algorithms are usually used, leading to a good balance between complexity and encoding accuracy. In particular, the dynamic M -search algorithm [12], which is shown to generally perform better than the conventional M -search algorithm, is used here to search the CEC-RVQ.

The decoder optimization step consists of using the Gauss-Seidel algorithm [6] to iteratively minimize the average output entropy of the residual signal subject to fixed stage encoding partitions. Suppose the CEC-RVQ contains P stage VQ codebooks, each containing N_p ($1 \leq p \leq P$) k -dimensional code vectors. Also, let $\mathcal{V}(j_p)$ denote the j_p th non-causal partition cell that corresponds to the j_p th code vector in the p th stage codebook. The partition cell $\mathcal{V}(j_p)$ is formed of all stage-removed residual vectors $\gamma^i(j_p) = \mathbf{x}^i - \mathbf{z}_p^i(j_p)$, where \mathbf{z}_p^i is given by

$$\mathbf{z}_p^i(j_p) = \sum_{l=1}^{p-1} \mathbf{y}_l(j_l^i) + \sum_{i=p+1}^P \mathbf{y}_i(j_i^i),$$

where j_1^i, \dots, j_p^i are the corresponding encoding decisions for the input vector \mathbf{x}^i . Each iteration of the Gauss-Seidel algorithm consists of sequentially replacing for each stage partition cell the old stage code vector $\mathbf{y}(j_p)$ with the centroid vector $\mathbf{c}(j_p)$ given by

$$\mathbf{c}(j_p) = \arg \min_{\mathbf{u} \in \mathbb{R}^k} \sum_{\gamma^i(j_p) \in \mathcal{V}(j_p)} -\log_2 \text{pr}(I(\gamma^i(j_p) - \mathbf{u})). \quad (1)$$

The centroid vector $\mathbf{c}(j_p)$ is very difficult (if not impossible) to determine analytically. Thus, a numerical optimization procedure is used in this work. This further complicates the decoder optimization, but such iterative optimization is only performed in the design process and therefore does not affect the encoder/decoder complexity.

The entropy coder optimization consists of simply updating the finite-state machine (FSM) and the corresponding state tables of conditional probabilities [6]. Only the

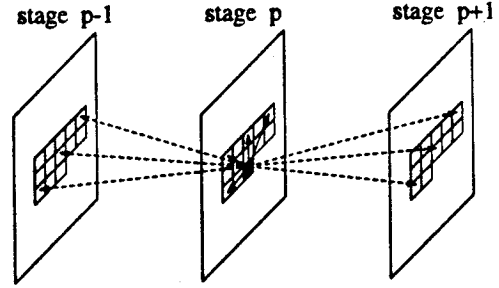


Figure 3: Illustration of a conditioning structure for CEC-RVQ.

stage high order statistical models are optimized, and no actual entropy coders are embedded in the design loop. This simplifies the design process, but the complexity of the stage statistical models must still be addressed. Like VQ, high order statistical modeling provides a way to exploit high order statistics while also requiring complexity that is exponentially dependent on the parameters of the model. RVQ drastically reduces the complexity of the high order model and improves our estimates of the dependencies by generating multistage approximations of the input signal, where the output alphabets of the subspaces are small (e.g., 2, 3, or 4).

Complexity-constrained statistical modeling for the output of the stage RVQs can be divided into three tasks. The first task is to locate a small number m_p of conditioning symbols (or previous outputs of some stage RVQs), given an initial region of support containing \mathcal{R}_p conditioning symbols, such that the m_p th order conditional entropy is minimized. This is illustrated in Figure 3 for the case of image coding, where the shaded block in the middle is the stage vector upon which conditioning is being performed. A total of m (12 in this case) neighboring blocks is utilized for conditioning. These blocks define the spatial region supporting the conditioning. The solid arrows show these neighboring blocks at the p th stage. In addition to the spatial dimension, conditioning is based on corresponding blocks at different stage levels, which is illustrated in Figure 3 by the dashed arrows, showing these conditioning blocks at the $(p-1)$ th and $(p+1)$ th stages. By building a conditioning tree as described in [7] and using the dynamic M -search algorithm, one can find the best stage statistical models of orders 1, 2, 3, etc.

The second task to be performed is to determine the best orders for each of the stage statistical models subject to a constraint on overall complexity. For this purpose, a tree with P branches is built and populated with a sufficiently large number of complexity-entropy pairs in each branch. The well-known generalized BFOS algorithm [13] is then used to prune the tree to find the best stage orders subject to a limit T_1 on the number of conditional probabilities, used here as a measure for complexity.

Since relatively high orders are usually required to achieve a very low entropy, the complexity of the stage

IMAGE	HYBRID CODER	DPCM	[3]	[4]
LENA	4.27	4.80	4.42	4.20
BRIDGE	4.30	4.82	4.30	4.32

Table 1: Performance comparison of the hybrid lossless coder with DPCM, [3], and [4].

statistical models can still be high. Moreover, contextual information is usually located in a relatively small region of the state space. In other words, many states do not occur, and corresponding tables of conditional probabilities are not populated. Thus, the third task is to reduce the number of states while sacrificing a minimal loss in performance. The PNN algorithm [14] was shown to be successful in reducing the size of the stage statistical model by one order of magnitude while still limiting the increase in entropy to about 1%. The same approach used to locate the best stage statistical model orders is used here, where the PNN algorithm is applied to each of the stage statistical models with just-determined orders such that a new complexity-entropy pair is obtained every time two conditioning states are merged into a new one. The BFOS algorithm is again applied to identify the best numbers of conditioning states subject to a limit T_2 ($T_2 \ll T_1$) on the total number of conditional probabilities.

IV. EXPERIMENTAL RESULTS

Several images of size 512×512 taken from the USC database were used to design a CEC-RVQ codebook as described in the previous section. In all cases, test images were excluded from the training set. The CEC-RVQ codebook contains 12 stage codebooks with four 4×4 code vectors in each codebook. It is searched using the dynamic M -search algorithm, leading to approximately 60 vector Lagrangian calculations per input vector. The conditioning scheme we use is the one illustrated in Figure 3.

To locate the best orders for the stage models for a fixed maximum number of 4096 conditional probabilities, a balanced tree with depth 6 is constructed where the best 1, 2, ..., 6 conditioning stage symbols are used. After the BFOS algorithm is employed, the number of conditioning states is further reduced by the PNN algorithm, whose outputs are used to populate yet another tree. Finally, the BFOS algorithm is used again to generate the FSM where the number of conditional probabilities is limited to 512.

The CEC-RVQ that yields the minimum overall entropy is determined as described previously using the training sequence. The corresponding set of stage codebooks, mapping tables generated by the PNN algorithm, and tables of conditional probabilities are used for encoding. Table 1 shows the entropy performance of the proposed hybrid coder, DPCM, and that of two of the best loss-

less compression techniques [3, 4] on the test images LENA and BOAT. The entropy is used as a measure so that the comparison is fair. An actual adaptive arithmetic coder was used to encode both the output of the stage RVQs and the residual image, and the compression ratios were slightly larger. Obviously, the proposed coder compares very favorably. Even better compression performance may be attained by using larger vector sizes and/or exploiting any statistical dependencies between the multistage images and the residual one. Preliminary experimental results are encouraging further study.

V. ACKNOWLEDGEMENT

The first author wishes to acknowledge W. Chung for valuable discussions.

REFERENCES

- [1] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Trans. Comm.*, vol. 29, no. 6, pp. 858-867, 1981.
- [2] S. M. Lei, T. C. Chen, and K. H. Tzou, "Subband HDTV coding using high-order conditional statistics," *IEEE JSAC*, vol. 11, pp. 65-76, Jan. 1993.
- [3] X. Ginesta and S. P. Kim, "Semi-adaptive context-tree based lossless image compression," in *ICIP*, (Austin, TX), Nov. 1994.
- [4] K. Papat and R. W. Picard, "Exaggerated consensus in lossless image compression," in *ICIP*, (Austin, TX), Nov. 1994.
- [5] S. Yu, M. N. Wernick, and N. P. Galatsanos, "Lossless compression of multidimensional medical image data using binary-decomposed high-order entropy coding," in *ICIP*, (Austin, TX), Nov. 1994.
- [6] F. Kossentini, W. Chung, and M. Smith, "Image coding using high-order conditional entropy-constrained residual VQ," in *ICIP*, (Austin, TX), Nov. 1994.
- [7] F. Kossentini, W. Chung, and M. Smith, "Conditional entropy-constrained residual VQ with application to image coding," *Submitted to Trans. on Image Processing*, July 1994.
- [8] J. Dozier and J. Tilton, "Data compression for data archival, browse or quick-look," in *Space & Earth Sci. Data Comp. Workshop*, (Snowbird, UT), Mar. 1991.
- [9] J. T. D. Dovik and M. Manohar, "Compression through decomposition into browse and residual images," in *Space & Earth Sci. Data Comp. Workshop*, (Snowbird, UT), Mar. 1993.
- [10] W. Abbott III, R. Kay, and R. Pieper, "Performance considerations for the application of the lossless browse and residual model," in *Space & Earth Sci. Data Comp. Workshop*, (Salt Lake City, UT), Mar. 1994.
- [11] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," in *ICASSP*, (Detroit, MI), Apr. 1995.
- [12] F. Kossentini and M. Smith, "A fast searching technique for residual vector quantizers," *Signal Processing Letters*, vol. 1, pp. 114-116, July 1994.
- [13] E. A. Riskin, "Optimal bit allocation via the generated BFOS algorithm," *IEEE Trans. on Information Theory*, vol. 37, pp. 400-402, Mar. 1991.
- [14] W. H. Equitz, "New vector quantization clustering algorithm," *IEEE Trans. on ASSP*, vol. 37, pp. 1568-1575, Oct. 1989.