Bit-Plane Golomb Coding for Sources with Laplacian Distributions

[1]R. Yu, [2]C. C. Ko, [1]S. Rahardja, and [1]X. Lin
[1]Laboratories for Information Technology, A*STAR,
21 Heng Mui Keng Terrace, Singapore 119613, email: {rsyu,rsusanto,linxiao}@lit.org.sg
[2]Department of Electrical and Computer Engineering,
National University of Singapore, Singapore 119260, email: elekocc@nus.edu.sg

ABSTRACT

This paper presents a bit-plane coding algorithm for Laplacian distributed sources that are commonly encountered in signal compression applications. By exploiting the statistical characteristics of the sources, the proposed algorithm achieves a rate-distortion performance that is essentially comparable to an optimal non-scalable scalar quantizer, while at the same time operates at a complexity level suitable for most practical implementations.

1. INTRODUCTION

Recently, compression algorithms that provide rate/fidelity scalability have attracted more and more attention. This is especially so for applications where a range of coding rates are required, or where channel bandwidth fluctuates. Among the coding strategies that provide such a scalability, the embedding coding algorithm has been shown to be very attractive, which has been widely implemented in image [1], video [2], speech [3] and audio [4] coding systems. The most distinguished feature of an embedded coding system lies in its ability to generate a compression bit stream that can be randomly truncated to fit certain rate, fidelity, or complexity constraints without loss of optimality. This decoupling property of the rate/fidelity adjustment to the encoding procedure actually bring much flexibility to a scalable coding system, where the rate/fidelity adjustment can be performed locally at the encoder, the decoder or the transmission network by simply truncating the compressed bit-stream.

Since data are commonly stored in a binary format in most electronic computing devices, one natural approach to implement an embedded coding system is through sequential *bit-plane coding* (BPC) [1][2][4][12], where the input data are sequentially scanned and coded by bit-planes, usually from the most significant to the least, to generate the compressed bit stream. In addition to its structural simplicity, sequential bit-plane coding (by a careful selection of the coding order) also conforms very well to the condition for optimal embedded coding, or the so-called embedded principle [12,13], in most situations.

Despite the obvious advantages, designing an embedded coding scheme that works at the *rate-distortion* (R-D) curve of a optimal non-scalable quantizer is generally a formidable task, as there are often intricate interaction between the different quality layers embedded in the final bit streams. In many embedded systems, compromises have to be made to maintain the complexity to an acceptable level at the price of reduced quality. However, as will be shown in this paper, optimal BPC is actually achievable at a complexity level that is suitable for most practical systems if we constrain the source to be *independent and identically distributed* (i.i.d) with a Laplacian distribution. Although it seems to be rather over-restrictive, such a condition is actually commonly encountered in many signal compression applications. For example, it was found that image, video, speech and audio signals are approximately Laplacian distributed [5-7]. Furthermore, the i.i.d constraint is also easily satisfied for more general sources by incorporating de-correlation and context modeling techniques.

We will also introduce in this paper a particular sub-class of BPC, namely, *bit-plane Golomb code* (BPGC), which is constructed by a discretization of the optimal BPC for Laplacian sources. It is found that the BPGC gives an identical compression performance to that of the Golomb code [8] for non-negative integer sources with geometrical distributions. The Golomb code has been shown to be the optimal Huffman code for such sources [9] and has been widely adopted in signal compression applications such as the JPEG-LS [10] and lossless audio coding schemes [11] due to its very simple coding/decoding rules and good compression performance. Further R-D analysis for BPGC shows that it gives a compression performance that closely approximates an optimal fixed-rate *entropy-constrained scalar quantizer* (ECSQ) [5] over a wide range of rates for source with Laplacian distribution.

2. BPC FOR LAPLACIAN SOURCES

Consider an input k-dimensional data vector $x = \{x_1,\ldots,x_k\}$ where x_i is extracted from an i.i.d. random source of some alphabet $A \subset \Re$. In a BPC scheme, bit-planes of x is formed by a binary representation of x_i

$$x_i = (2s_i - 1)\cdot \sum_{j=\infty}^{-\infty} b_{i,j}\cdot 2^j, i = 1,\ldots,k \quad (1)$$

that comprises of a sign symbol

$$s_i \overset{\Delta}{=} \begin{cases} 1 & x_i \ge 0 \\ 0 & x_i < 0 \end{cases}, \; i = 1,\ldots,k \quad (2)$$

and bit-plane symbols $b_{i,j} \in \{0,1\}$, $i = 1,...,k$. The bit-planes are then scanned and coded, in an order that can be reproduced at the decoder, to produce the compressed bit-stream. Note that in practice, the bit-plane scanning procedure may start from the most significant bit-plane $M-1$ given by $2^{M-1} \le \max\{|x_i|\} < 2^M, i = 1,...,k$.

In an effort to pursue optimal truncation performance of the resulted embedded bit-stream, we may require that every prefix of the stream can be decoded to reconstruct the original data with a fidelity approaching that of an "optimal" compression algorithm at the same rate as that prefix. While a global optimal solution to this problem remains intractable at this moment, locally optimized, "greedy" heuristic solutions have been previously discussed in [12, 13], which lead to the so-called embedded principle [13] that requires an embedded bit-stream to carry information in the decreasing order of its importance. From the results of [12], it can be seen that for i.i.d data sources (except for those with very skew probability distribution), the embedded principle is satisfied well by a simple sequential bit-plane scanning procedure that comprises the following steps:

1. Starting from the most significant bit-plane $j = M-1$;
2. Code those bit-plane symbols $b_{i,j}$ with $b_{i,M-1} = b_{i,M-2} = ... = b_{i,j+1} = 0$ (Significance pass);
3. If $b_{i,j} = 1$ in the significance pass, code the sign symbol s_i;
4. Code those bit-plane symbols $b_{i,j}$ that are not coded in the significance pass (Refinement pass);
5. Progress to bit-plane $j-1$.

The above procedure is iterated until certain terminating criterion, which is usually a pre-defined rate/distortion constraint, is reached

We now turn our attention to entropy coding of bit-plane symbols, where the data compression is actually attained. It can be seen that for general data sources, there exists both intra and inter bit-plane dependencies among the probability distributions of bit-plane symbols. As a result, native probability table based approach to catch these statistical dependencies will generally results in a table with a large number of entries, which not only increases the complexity of coding algorithm, but also leads to large model cost [14] that eventually degrades the compression performance. For this reason, a simplified approach is generally adopted in practical applications [1][12][13] where only a limited set of bit-plane symbols that are likely to have very skew distribution, e.g., those scanned in significance pass, are entropy coded.

An alternative solution to this dilemma is by adopting a parametric approach that assumes a known function form for the pdf of the data source so that the number of the free parameters can be dramatically reduced. Specifically, if we have the constrain that the input data vectors be drawn from an i.i.d source that has a Laplacian pdf given by,

$$f_X(x) = e^{-|x|\sqrt{2/\sigma^2}} \Big/ \sqrt{2\sigma^2}\,, \qquad (3)$$

the above statistical dependencies will vanish and the probably distributions of the bit-plane symbols are simply given by

$$P_j \stackrel{\Delta}{=} \Pr(b_{i,j} = 1) = 1 - \left(1 + \theta^{2^j}\right)^{-1}, \; i = 1,...,k\,, \qquad (4)$$

$$\Pr(b_{i,j} = 0) = 1 - P_j\,, \; i = 1,...,k\,, \qquad (5)$$

and

$$\Pr(s_i = 1) = \Pr(s_i = 0) = 0.5\,, \; i = 1,...,k\,, \qquad (6)$$

where $\theta \stackrel{\Delta}{=} e^{-\sqrt{2/\sigma^2}}$. Clearly, P_j follows the following updating rule:

$$P_j = \sqrt{P_{j+1}} \Big/ \left(\sqrt{1 - P_{j+1}} + \sqrt{P_{j+1}}\right) \qquad (7)$$

This statistical independency among bit-planes significantly simplify the problem of optimal coding of the bit-plane symbols, which is easily achieved by entropy coding each bit-plane independently using, e.g., arithmetic code with probability assignment (4) – (6), provided the distribution parameter θ is known a prior. In a universal coding system that deals with sources with unknown distribution parameter, θ may be learnt from the current or previously coded data vector. For example, the *maximum likelihood* (ML) estimation of θ is given by

$$\theta = e^{-N/A}\,, \qquad (8)$$

where N and A are the length and the absolute sum of the data vector, respectively.

3. BIT-PLANE GOLOMB CODE

The BPC scheme described above can be further simplified by limiting its possible probabilities assignment. Specifically, we consider the code family $C = \{G^L \mid L \in \mathrm{Z}\}$, where G^L is a BPC code whose probability assignment Q_j^L is given by

$$Q_j^L = \begin{cases} 1/\left(1 + 2^{2^{j-L}}\right) & j \ge L \\ 1/2 & j < L \end{cases}, \qquad (9)$$

for bit-plane j. Clearly, Q_j^L follows the probability updating rule of (7) for bit-planes $j \ge L$ and enters a "lazy mode" (since entropy coding of a symbol with probability $1/2$ can be achieved by directly outputting that symbol to the compressed bit stream) for bit-planes $j < L$ where the probability skew is very small.

Clearly, C actually gives a partitions rule that partitions the support of θ into disjoint regions, so that each region has a correspondent optimal G^L from C. Here the optimality is measured, reasonably, in terms of the expected code length for G^L if it is truncated at some bit-plane $T \le L$. We may further denote code G^L as G_T^L if it is truncated at bit-plane T to facilitate our sequent discussion. It can be shown that the expected code length $\overline{n}_L$ (the cost of the sign symbols is ignored since it does not affect the result of the optimal region) of G_L^L is given by

$$\overline{n}_L(\theta) = -\sum_{j=L}^{\infty} P_j \log_2 Q_j^L - \sum_{j=L}^{\infty} (1-P_j)\log_2(1-Q_j^L)$$
$$= \left(1-\theta^{2^L}\right)^{-1}. \qquad (10)$$

Thus the decision boundary for code G^L and G^{L+1} is given by:

$$\overline{n}_L(\theta) = \overline{n}_{L+1}(\theta) + 1, \qquad (11)$$

which simplifies to $\theta = \phi^{2^{-L}}$ by substituting (10) into (11). Here, $\phi \overset{\Delta}{=} (\sqrt{5}-1)/2$ is the inverse of the golden ratio. The decision rule for the optimal G^L is then given by:

$$\phi^{2^{-L+1}} \le \theta < \phi^{2^{-L}}, \qquad (12)$$

It is interesting to observe that G_0^L with $L \ge 0$ actually gives an identical expected code length to that of Golomb code with parameter 2^L for non-negative geometrically distributed integer [9]. For this reason, the code family of C can be referred to as *bit-plan Golomb code* (BPGC). In addition, note that since the Golomb code does not have counterparts for BPGC with $L < 0$, it does not perform equally well for sources with very low entropy contents.

In practical applications, we may wish to have a low-complexity adaptation rule for the optimal BPGC given the sufficient statistics N and A for θ. By taking the above equivalency of BPGC to the Golomb code, the approximation procedure described in [10] for optimal Golomb code is repeated here, which finally leads to the following simple adaptation rule:

$$L = \min\{L' \in Z \mid 2^{L'+1} N \ge A\}. \qquad (13)$$

A practical implementation of the above selection rule is given in the following C program, which is extended from the "one-liner" C program in [10] to support a wider range of the value of L:

```
if (N<=A) for (L=0;(N<<(L+1))<=A;L++);
else for (L=-1;(N>>(-L))>A;L--);
```

In the decoder, the bit-planes of x are reconstructed sequentially following the same order as in the encoder, resulting in a partial reconstruction up to, say, bit-plane T of the original bit-planes. This partial reconstruction actually specifies an interval for the value of x and the optimal (in terms of mean square error) reproduction $\hat{x} = \{\hat{x}_1, ..., \hat{x}_k\}$ of x is then given by the centroid of that interval, which may be calculated iteratively given the bit-plane wise probability assignment (9).

4. PERFORMANCE ANALYSIS

From the above coding/decoding procedures of BPGC, it can be seen that G_T^L is essentially equivalent to a *uniform threshold quantizer* (UTQ) [5] that has a step size $\Delta = 2^T$ and a central dead-zone 2Δ, which (with a slightly different size of the dead-zone) is known to be a nearly optimal ECSQ for Laplacian sources [14]. Specifically, if we consider the following signal-to-noise-ratio SNR performance:

$$SNR \overset{\Delta}{=} 10\log_{10}(\sigma^2/D) \text{ dB}, \qquad (14)$$

where σ^2 is the variance of the source and D is the normalized mean-square error distortion:

$$D \overset{\Delta}{=} \frac{1}{k} E\left\{\|x-\hat{x}\|^2\right\} = \frac{1}{k} E\left\{\sum_{i=1}^{k} |x_i - \hat{x}_i|^2\right\}, \qquad (15)$$

it can be shown that the SNR performance of the code G_T^L for an i.i.d Laplacian source with distribution parameter θ is given by

$$SNR_T^L = 10\log_{10}\left\{1 + \frac{\theta^\Delta}{2}\left[(\Delta\ln\theta)^2\theta^\Delta/(1-\theta^\Delta)^2 - (\Delta\ln\theta)^2 + 2\Delta\ln\theta - 1\right]\right\}^{-1} \qquad (16)$$

Furthermore, the normalized expected code length, or the rate of G_T^L for that source is given by

$$R_T^L = -\sum_{j=T}^{\infty} P_j \log_2 Q_j^L - \sum_{j=T}^{\infty} (1-P_j)\log_2(1-Q_j^L) + \theta^\Delta, \qquad (17)$$

where P_j and Q_j^L are defined as in (4) and (9), respectively. Note that the above equation can be further simplified as

$$R_T^L = (1-\theta^\Delta)^{-1} + L - T + \theta^\Delta, \qquad (18)$$

if $L \ge T$.

Fig. 1 gives an example of the performance of G_T^L with respect to an optimal ECSQ for a Laplacian source. The theoretical R-D bound for the Laplacian source computed by using the Blahut-Arimoto algorithm [16, 17] is also plotted in this figure for reference. It can be seen from Fig. 1 that the performance of G_T^L lags only slightly when compared with that of the optimal ECSQ, which is due to its sub-optimal dead-zone and cost resulting from the informational divergence between G_T^L and the underlying data source. At high rates, the penalty of the sub-optimal dead-zone dies out and the performance gap between these two schemes is mainly from the information divergence, which is found to be smaller than 0.1 bits in most cases by a redundancy analysis similar to that in [9].

In order to further verify our results, numerical experiments were also conducted where the R-D performance of an experimental BPGC coder is measured with source data vectors generated by a pseudo Laplacian random variable generator. The results are given in Fig. 1, which are obtained by averaging an ensemble of 1000 data vectors, each with a length of 512 samples. Clearly, these results strongly validate our previous analysis since the experimental BPGC coder achieves an R-D performance that is very close to its theoretical R-D curve.

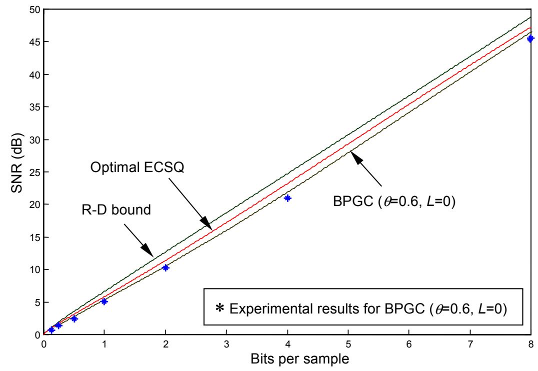

Figure 1. BPGC performance

5. CONCLUSION

We have shown that optimal BPC for Laplacian source is actually achievable at a complexity level that is suitable for a practical implementation. In addition, an important sub-class of BPC, namely, BPGC is also introduced in this paper. This gives the same lossless compression performance as the Golomb code for geometrically distributed integer sources. Furthermore, its excellent R-D performance for Laplacian sources is confirmed by an R-D analysis.

REFERENCE

[1] J. Shapiro, "Embedded image coding using zerotree of wavelet coefficients," *IEEE Tran. Signal Processing*, vol. 41, pp. 3445 – 3462, Dec. 1993

[2] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Tran. Image Processing*, vol. 3, pp. 572-588, Sep. 1994

[3] Ramprashad, S.A. "High quality embedded wideband speech coding using an inherently layered coding paradigm," *Proceedings of ICASSP '00,* pp 1145 – 1148.

[4] S. H. Park et al., "Multi-Layer Bit-Sliced Bit Rate Scalable Audio Coding," presented at the *103rd Convention of the Audio Engineering Society*, New York, Sep. 1997 (preprint 4520).

[5] G. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. Information Theory*, vol. 42, pp. 1365-1374, Sep. 1996.

[6] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Comm.*, vol. 31, pp 835 – 839, June 1983.

[7] M. D. Paez and T. H. Glisson, "Minimum mean-squared-error quantization in speech PCM and DPCM systems," *IEEE Trans. Comm.*, vol. 20, pp. 225 – 230, 1972.

[8] S. W. Golomb, "Run-length encodings," *IEEE Tran. Information Theory*, vol. 12, pp. 399-401, July 1966.

[9] R. Gallager and D. V. Voorhis, "Optimal source codes for geometrically distributed integer alphabets," *IEEE Trans. Information Theory*, vol. 21, pp. 228-230, Mar. 1975.

[10] M. J. Weinberger and et at, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Tran. Image Processing*, vol. 9, pp 1309-1324, Aug. 2000.

[11] M. Hans and R. W. Schafer, "Lossless compression of digital audio", *IEEE Signal Processing Magazine,* pp. 21-32, July. 2001.

[12] J. Li and S. Lie, "An embedded still image coder with rate-distortion optimization," *IEEE Trans. On Image Processing*, vol. 8, no. 7, pp. 913-924, July 1999

[13] E. Ordentlich et al, "A low-complexity modeling approach for embedded coding of wavelet coefficients," *HP Labs Tech. Reports*, HPL-97-150, 1997.

[14] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, London, U.K.: World Scientific, 1989.

[15] T. Berger, "Minimum entropy quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 149 – 157, Mar. 1982.

[16] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Tran. Information Theory*, vol. 18, pp. 14-20, Jan. 1972.

[17] R. E. Blahut, "Computation of channel capacity and rate-distortion functions,", *IEEE Tran. Information Theory*, vol. 18, pp. 460 – 473, July 1972.