FUZZY TRELLIS VECTOR QUANTIZATION OF IMAGES

Tariq Haddad and Abbas Yongaçoğlu

School of Information Technology and Engineering, University of Ottawa
Ottawa, Ontario, K1N 6N5, Canada
Fax: +1 (613) 562 5175
thaddad@site.uottawa.ca yongacog@site.uottawa.ca

ABSTRACT

This paper introduces a new codebook search algorithm for trellis vector quantization systems (TVQ). The development of the new algorithm is based on the symbol-MAP channel decoding algorithm, which is modified for data compression to deliver soft distortion-related reliability information. Following a rate-distortion theoretic approach, the soft information is used to derive a codebook search algorithm that is capable of solving the problems associated with the LBG algorithm. The derived algorithm is fuzzy in the sense that it follows a soft association rule, however, it is deterministic in the descent towards the global minimum distortion point. Although the derivation is general, the algorithm is tested using gray-scale images, which provide a nonconvex square-error distortion surface. As shown in the simulation section, the new algorithm provides lower energy codebooks (~ 0.8 dB gain), while being significantly less sensitive to initial codebooks using short training image sequences.

1. INTRODUCTION

It has been shown that trellis waveform coding provides near theoretical limit distortion-rate performance [1], [2]. A trellis vector quantization (TVQ) system is characterized by a finite-state machine (FSM) decoder and a trellis search encoding algorithm. The channel output sequence is fed into the FSM decoder, which produces the corresponding index of a reproduction codeword. The FSM decoder is used to construct the associated encoding trellis, whose branches are colored by the reproduction codewords. The encoding process involves a search algorithm that delivers the proper channel input sequence, which represents the branch indices of the minimum distortion path in the trellis. The authors in [3] describe an efficient approach for TVQ system design using an LBG-based algorithm to search for an optimal minimum distortion partition. However, the LBG algorithm is extremely sensitive to the initialization phase, which controls the descent to a locally optimal configuration. Consequently, it is very likely that the nonconvex distortion surface, which is associated with images, tricks a poorly initialized LBG algorithm into settling at a locally optimal solution. The limitations of the LBG-algorithm called for other global codebook search techniques, such as deterministic annealing [4].

One of the algorithms used with TVQ is the Viterbi algorithm. The function of the Viterbi algorithm in searching for the minimum distortion path is analogous to its function in delivering the most probable transmitted sequence in channel decoding. In anticipation of producing an uncorrelated and uniformly distributed sequence of symbols; i. e., a redundancy free sequence, the Viterbi algorithm is applied with TVQ (VTVQ) considering independent and identically distributed trellis branch indices. Consequently, given a source output sequence of vectors, using an algorithm that produces a sequence of minimum distortion symbols is equivalent to using the Viterbi algorithm, which produces the minimum distortion sequence. Based on this fact, we introduced in [5] and [6] a new TVQ search algorithm, which represented the "natural dual" to symbol-MAP trellis channel decoding. The new search algorithm in [5] and [6] is based on the early BCJR algorithm introduced for channel decoding in [7]. The algorithm performs decision-delayed processing and delivers for each time instant a set of reliability (soft) values that are used to decide on the minimum distortion source encoder output symbol. Since the algorithm delivers soft information, in the form of transition probabilities, this encoding approach is called "soft trellis vector quantization" (STVQ).

In this paper, we extend STVQ by using the reliability information, which is a product of this compression approach, to provide a robust fuzzy, yet deterministic, procedure for the design of trellis compression systems [5]. The fuzzy design approach alleviates the problems associated with the LBG algorithm when applied to TVQ codebook design. The derivation of the new fuzzy trellis search algorithm is conceptually based on the development of the rate distortion theory, and is similar to deterministic annealing for data clustering, in particular vector quantization. We will show through several simulation examples using gray images that the new fuzzy design approach produces lower distortion configurations, while being significantly less sensitive to initial codebooks.

2. THE STVQ ALGORITHM

In this section we briefly describe the soft compression algorithm (STVQ), which is derived in [6, 5] based on an analogy drawn between a source and a communication channel. Consider using a labeled-transition trellis associated with a FSM-decoder of constraint length ν and k-bit in-

0-7803-6293-4/00/$10.00 ©2000 IEEE.

puts. Accordingly, the TVQ is characterized by a trellis that has $M = 2^{k(\nu-1)}$ states, and 2^k branches stemming out of each state. The branches are colored or labeled by the indices that address a reproduction codebook $\mathcal{C} = \{C_0, C_2, ..., C_{K-1}\}$, where $K = 2^{k\nu}$.

Consider that the output vector of a stationary discrete-time source at time n is denoted by $Y_n = [y_{n1}, y_{n2}, \ldots, y_{nL}]$. The derivation of the STVQ algorithm is based on the fact that we may consider the TVQ process as decoding at the output of the channel defined by [6, 5]

$$p(Y_n|x_n, s_n) = \exp\{-\eta d(Y_n, x_n|s_n)\} \tag{1}$$

where η is a Lagrange multiplier, and $d(Y_n, x_n|s_n)$ is the distortion between the source output vector Y_n and the codeword associated with state s_n and branch index x_n. Then, given a source output sequence Y_1^N, we choose an encoder output symbol according to

$$x_n = \arg\max_{i\in\mathcal{A}} \Pr\{x_n = i|Y_1^N\} \tag{2}$$

The forward-backward procedure described in [7] is used to solve for the set of probabilities $\Pr\{x_n = i|Y_1^N\}$, $i = 0, 1, \ldots, I-1$ and $I = 2^k$. We proceed by defining the joint probability variable

$$\lambda_n^i(m) = \Pr\{x_n = i, s_n = m, Y_1^N\} \tag{3}$$

where $s_n = m$ is the encoder state at time n. It follows that

$$\Pr\{x_n = i|Y_1^N\} = \frac{\sum_{m=0}^{M-1} \lambda_n^i(m)}{\sum_{m=0}^{M-1}\sum_{i=0}^{I-1} \lambda_n^i(m)} \tag{4}$$

It is shown in [6, 5] that the joint probability variable defined in Eq. (3) can be written as

$$\lambda_n^i(m) = \alpha_n^i(m) \cdot \beta_n^i(m) \tag{5}$$

where

$$\left.\begin{array}{rcl} \alpha_n^i(m) & = & \Pr\{x_n = i, s_n = m, Y_1^n\} \\ \beta_n^i(m) & = & \Pr\{Y_{n+1}^N|x_n = i, s_n = m\} \end{array}\right\} \tag{6}$$

It can be shown that the forward variable is evaluated using

$$\alpha_n^i(m) = \gamma_n^i(m)\sum_{j=0}^{I-1} \alpha_{n-1}^j(S_b^j(m)) \quad n = 2, 3, \ldots, N \tag{7}$$

where $S_b^j(m)$ is the state at which we arrive if we go backwards from state $s_n = m$ along the branch $x_{n-1} = j$. Besides that,

$$\gamma_n^i(m) = \exp\{-\eta d(Y_n, x_n = i|s_n = m)\} \tag{8}$$

On the other hand, we can easily show that the recursion of the backward variable is given by

$$\beta_{n-1}^j(\acute{m}) = \sum_{i=0}^{I-1} \beta_n^i(S_f^j(\acute{m}))\cdot\gamma_n^i(S_f^j(\acute{m})) \quad n = N, N-1, \ldots, 2 \tag{9}$$

where $S_f^j(\acute{m})$ is the state at which we arrive if we go forward from state $s_{n-1} = \acute{m}$ along the branch $x_{n-1} = j$. The implementation approach and the impact of the Lagrange multiplier η, used in Eq. (8), on the algorithm's performance is discussed in [5, 6].

3. FUZZY SOFT TRELLIS VECTOR QUANTIZATION

In this section we use the development of the rate distortion theory to extend the STVQ algorithm to provide an efficient and robust fuzzy and deterministic trellis codebook search approach. We will demonstrate in this section that the developed approach is capable of descending to lower distortion points while being highly insensitive to the initialization of the search process. In [8], theorem 6.3.3 in particular, Blahut parameterizes the rate distortion function $R(D)$ in an equality, rather than an inequality, using a Lagrange multiplier κ. In that theorem, the author shows that the optimal conditional distribution that achieves the minimum value of $R(D)$ is [8]

$$Q_{l|j} = \frac{q_l\, e^{\kappa d_{jl}}}{\sum_l q_l\, e^{\kappa d_{jl}}} \tag{10}$$

with an average distortion D

$$D = \sum_j \sum_i p_j Q_{l|j}^* d_{jl} \tag{11}$$

where j and l are used to index the input and reproduction alphabets, respectively. $Q_{l|j}$ is the conditional probability function between l and j, and $Q_{l|j}^*$ is the distribution that achieves the point on the $R(D)$ curve. Besides that, d_{jl} is the distortion between symbols l and j, p_j is the input distribution, while q_l is the output distribution.

Theorem 6.3.3 in [8] is significant in the sense that it culminated into the famous Blahut's algorithm, which computes certain number of points on the rate distortion curve parameterized by the corresponding values of the Lagrange multiplier κ. For a particular value of $\kappa \in]-\infty, 0]$, Blahut's algorithm iteratively finds the best distribution of the reproduction alphabets q_l, which along with $Q_{l|j}$ defined in Eq. (10) achieves a point on the rate distortion curve. In other words, Blahut algorithm characterizes the distribution of the optimal reproduction alphabets (q_l) as well as the transition probabilities ($Q_{l|j}$), which control the encoding rule, for a given value of κ.

We demonstrated in the previous section that the derivation of the STVQ algorithm is based on the association distribution given by Eq. (1) and parameterized by the Lagrange multiplier η. This distribution is clearly identical to the conditional distribution $Q_{l|j}$ shown in Eq. (10), which leads to an optimal rate-distortion point parameterized by κ. The two distributions are different in their consideration of the distribution of the output alphabet (q_l). While it is the goal to search for an optimal output alphabet distribution (q_l) in the development of the rate-distortion theory, practical compression systems, and the STVQ algorithm in particular, adopt a uniform distribution to describe the reproduction alphabets. This consideration reduces the degrees of freedom in practical systems, and allows for the exact computation of the output alphabets.

The other point of interest is the similarity between the parameters η and κ, where η in Eq. (8) is equivalent to $-\kappa$ in Eq. (10). It is more perceptive to interpret the independent Lagrange multiplier κ as the slope of the tangent at the corresponding point on the rate-distortion curve. Accordingly, as $\kappa \to -\infty$ we approach the low-distortion high-rate points of the $R(D)$ curve, while $\kappa \to 0$ corresponds to the flat high-distortion low-rate points of the same curve. Unlike Blahut algorithm, which finds a rate-distortion point given the independent parameter κ, the TVQ problem involves encoding given a prespecified compression rate value. Therefore, the critical issue with the STVQ algorithm is to find the optimal η given a compression rate and an anticipated average distortion value. Experimental results showed that the values of η control the computational rate-distortion curve just like the values of κ control the theoretical rate-distortion curve [6].

Now, consider the general problem of data clustering. Irrespective of whether clustering is regular or performed over a trellis, the system design criterion is to minimize the average distortion

$$D = \sum_j \sum_l p_j \pi_{l|j} d_{jl} \tag{12}$$

since we are practically interested in delivering a hard decision, the conditional distribution $\pi_{l|j}$ is defined as

$$\pi_{l|j} = \begin{cases} 1, & j \in S_l \\ 0, & \text{Otherwise} \end{cases} \tag{13}$$

where S_l represents the lth encoding set. Furthermore, since we do not usually have access to the source distribution p_j, stochastic averaging in Eq. (12) is usually replaced by time averaging. Unlike the distortion measure defined in Eq. (11), which includes the soft conditional encoding rule $(Q_{l|j})$, the practical measure of Eq. (12) is based on a hard association rule $(\pi_{l|j})$.

In order to fit the notion of STVQ, we redefine the distribution $\pi_{l|j}$ as follows

$$\pi_n(l|Y_1^N) = \begin{cases} 1, & \psi_n(l) > \psi_n(j) \\ 0, & \text{Otherwise} \end{cases} \tag{14}$$

where $l, j = 0, 1, \ldots, 2^{k\nu} - 1$, $\psi_n(l) = \Pr\{C_l|Y_1^N\}$, and C_l is a codeword associated with a trellis state $s_n = m$ and a branch $x_n = i$. Accordingly,

$$\psi_n(l) = \frac{\lambda_n^i(m)}{\sum_{m=0}^{M-1} \sum_{i=0}^{I-1} \lambda_n^i(m)} \tag{15}$$

with $\lambda_n^i(m)$ defined in Eq. (3). We show in [6] that for a particular compression rate and expected average distortion there is a proper value η^* in Eq. (8), below which the STVQ-codebook search process with the LBG algorithm converges to higher average distortion values. Therefore, considering that we are descending on the computational distortion rate curve from high distortion values (small η) to the distortion corresponding to the operational rate (at η^*), we can perform some sort of deterministic relaxation by minimizing a soft average distortion while incrementing the value of η from small values to η^*. This intuitively suggests using the soft information $\psi_n(l)$, which is delivered by the STVQ algorithm and defined in Eq. (15), in place of the hard association rule $\pi_n(l|Y_1^N)$ of Eq. (14). Accordingly, instead of minimizing the average distortion given by Eq. (12) with Eq. (14), we propose minimizing the following distortion measure in the process of searching for the optimal codebook

$$D = \frac{1}{N_T} \sum_n \sum_{l=0}^{2^{k\nu}-1} \psi_n(l)\, d(Y_n, C_l) \tag{16}$$

where N_T is the length of the training sequence. It follows that to minimize the distortion given by Eq. (16) we set the gradient of D with respect to $\{C_l\}$ to zero, which for the squared error distortion produces the centroid update equations

$$C_l = \frac{\sum_n \psi_n(l) Y_n}{\sum_n \psi_n(l)} \quad \forall C_l \tag{17}$$

According to Eq. (17), a source output vector Y_n contributes a $\psi_n(j)$-weight of its value to the update of each reproduction codeword C_j. Clearly this update equation provides the deterministic relaxation sought for to help the LBG algorithm finds the global minimum configuration. Eq. (17) with the STVQ algorithm defines a new fuzzy search process we refer to as fuzzy-STVQ or FSTVQ. The search process with the FSTVQ algorithm is identical to that of the LBG algorithm with the exception of using Eq. (17) and the proper η-schedule instead of hard clustering. The process starts with relatively small values of η that correspond to very fuzzy association. The level of randomness is reduced each time we increment η towards the optimal η^*. It is worth mentioning that an update equation, similar to the one shown by Eq. (17), is derived in [9] following a DA approach for the special case of a labeled-state trellis coding system. In that brief paper, the authors use a forward-backward algorithm to solve for a set of probabilities similar to $\psi_n(l)$ in association with a dynamic programming algorithm to search for the minimum distortion path.

4. PERFORMANCE EVALUATION

This section shows the advantage of the FSTVQ algorithm in the area of image compression. The new algorithm is compared with the LBG algorithm using both STVQ and Viterbi-based TVQ (VTVQ). It is worth mentioning that both STVQ and VTVQ exhibit similar performance when applied to different sources [6, 5]. Several cases are considered with different trellis structures with a compression rate of 0.0625 bit/pixel (bpp). To show the advantage of FSTVQ, we use a relatively short training sequence of 2 images with an external test image [10]. The training set consists of 256×256 gray images represented by 8-bit pixels. The test image, on the other hand, is the 512×512 gray-scale Lenna image. The search for the optimal codebook is independently repeated for 15 times, each time with a different initial codebook chosen uniformly at random from [0 255].

Compression at 0.0625 bpp is achieved using a trellis with $k = 1$ and 4×4 image blocks producing a vector dimension $L = 16$. The performance measure is the peak signal to noise ratio (PSNR) defined as

$$\text{PSNR} = 10 \log_{10} \left[\frac{Y_p^2}{E\{d(Y, \hat{Y})\}} \right]$$

Table 1: The rate 0.0625 bpp PSNR coding performance in dB for the Lenna image using 2 training images.

ν	STVQ	VTVQ	FSTVQ
3	20.992	21.002	21.070
4	21.230	21.296	21.739
5	21.680	21.677	22.011
6	21.909	21.844	22.160

where Y_p is the highest intensity value used to represent a pixel, and $E\{d(Y,\hat{Y})\}$ is the MSE between the original and the reconstructed image. The results shown in Figs. 1 and 2 clearly emphasize the advantage of FSTVQ in providing lower energy configurations while being significantly less sensitive to the initialization process. The sensitivity to initial codebooks becomes more pronounced as the complexity of the trellis structure increases, as shown in Fig. 2. Nonetheless, the fluctuation in performance occurs within smaller range at higher levels of PSNR as compared to the LBG-based algorithms. The lowest distortion codebooks for each value of ν, which represent the best-case scenarios amongst the 15 trials, are used to encode the Lenna image. The PSNR results are shown in Table 1, where we demonstrate the advantage of using FSTVQ for the different constraint lengths. Nonetheless, the performance of the codebooks generated by the LBG algorithm would have been even worse had we considered higher distortion configurations. For example the highest distortion codebooks from Fig. 1(b) are those associated with initial codebooks 7, 3 and 3 for FSTVQ, VTVQ and STVQ, respectively. However, the lowest distortion codebooks for VTVQ and STVQ in Fig. 1(b) are those associated with initial codebooks 12 and 5, respectively.

5. REFERENCES

[1] A. J. Viterbi and J. K. Omura. "Trellis encoding of memoryless discrete-time sources with a fidelity criterion". *IEEE Transactions on Information Theory*, IT-20:325–332, May 1974.

[2] R. M. Gray. "Time-invariant trellis encoding of ergodic discrete-time sources with a fidelity criterion". *IEEE Transactions on Information Theory*, IT-23(1):71–83, Jan. 1977.

[3] L. C. Stewart, R. M. Gray, and Y. Linde. "The design of trellis waveform coders". *IEEE Transactions on Communications*, COM-30(4):702–710, April 1982.

[4] Kenneth Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". *Proceedings of the IEEE*, 86(11):2210–2239, Nov. 1998.

[5] Tariq Haddad and Abbas Yongaçoğlu. "Fuzzy soft trellis vector quantization". *Submitted for publication to IEEE Trans. on Communications*, Oct. 1999.

[6] Tariq Haddad and Abbas Yongaçoğlu. "Symbol-MAP-based trellis vector quantization". In *Proceedings of GLOBECOM'99*, Rio de Janeiro, Brasil, December 5-9 1999.

[7] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. "Optimal decoding of linear codes for minimizing symbol error rate". *IEEE Transactions on Information Theory*, IT-20(2):284–287, Mar. 1974.

[8] R. E. Blahut. *Principles and practice of information theory.* Addison-Wesley, Massachusetts, 1987.

[9] D. Miller, K. Rose, and P. Chou. "Deterministic annealing for trellis quantizater and HMM design using Baum-Welch re-estimation". In *ICASSP'94*, pages V261–V264, Adelaide, Australia, April 1994.

[10] P. Westerink, D. Bockee, J. Biemond, and J. Woods. "Subband coding of image using vector quantization". *IEEE Transactions on Communications*, 36(6):713–719, June 1988.

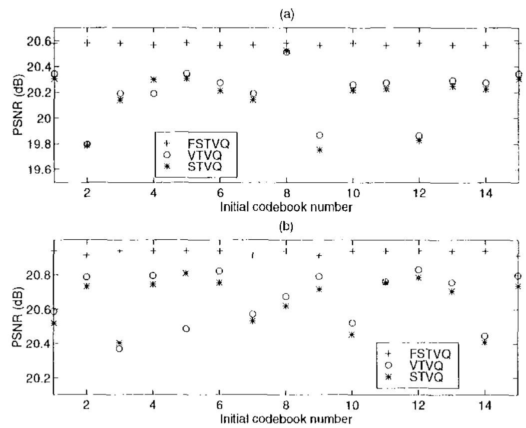

Figure 1: The rate 0.0625 bpp PSNR coding performance in (dB) using 2 training images: (a) $\nu = 3$, (b) $\nu = 4$.

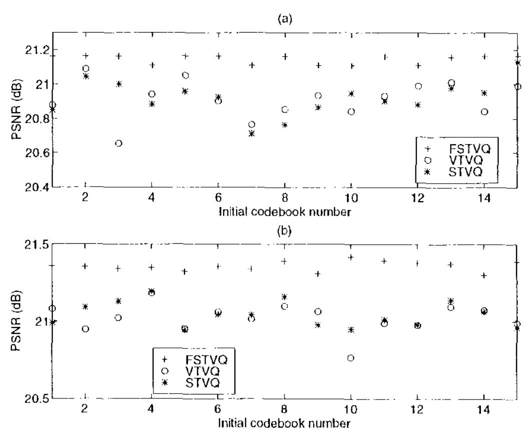

Figure 2: The rate 0.0625 bpp PSNR coding performance in (dB) using 2 training images: (a) $\nu = 5$, (b) $\nu = 6$.