MULTI-STAGE VECTOR QUANTIZER DESIGN FOR IMAGE TRANSMISSION OVER PACKET NETWORKS

Hosam Khalil and Kenneth Rose

Signal Compression Laboratory
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106, USA
hkhalil@scl.ece.ucsb.edu rose@ece.ucsb.edu

ABSTRACT

The design of a source-channel system based on the multi-stage vector quantizer (MSVQ) is optimized for packet networks. Resilience to packet loss is further enhanced by a proposed interleaving scheme that ensures that a single lost packet only eliminates a subset of the vector stages. The design is optimized while taking into account compression efficiency, packet loss rate, and the interleaving technique in use. Simulation results demonstrate that the packet network-optimized MSVQ outperforms traditional MSVQ for a variety of images and channel conditions, with gains of up to 2.0 dB in PSNR. Although the formulation is given in the context of packet networks, the work is directly extendible to the broader category of erasure channels.

1. INTRODUCTION

Packet networks have dramatically gained in importance and popularity in recent years, especially due to the widespread use of the Internet. Naturally, much research effort is currently focused on robustness to the type of errors that are encountered in such channels, namely, packet loss due to congestion and delays. Packet networks, hence, represent a special case of erasure channels. This paper is concerned with the design of source-channel coding systems for packet networks in particular, and erasure channels in general. Another potential application is in wireless communications over deep fading channels which may also be viewed as erasure channels.

Most existing image and video robustness-to-channel techniques heavily depend on decoder-driven concealment techniques (see e.g., [1].) Currently, more research is shifting towards encoder-driven robustness techniques. A known approach to combat packet loss is the design of diversity systems which employ multiple description coding. This direction of research was stimulated by Vaishampayan's work on scalar quantizers [2] and was later pursued in the context of transform coding [3] and vector quantization [4]. The basic idea is to encode and transmit more than one description of the source over different channels or packets. In the event of packet loss or channel failure, the decoder reconstructs the data from the remaining received descriptions. The encoder is thus designed such that there is sufficient redundancy in the descriptions.

MSVQ is a very efficient signal coding technique. It has met with considerable success in audio coding, and has recently been found to be applicable also to image and video coding [5]. The low complexity and memory requirements were among the most attractive features of MSVQ. In the case of images, blocks of uniform size are input to the quantizer which in turn finds the best set of MSVQ-parameters.

MSVQ decomposes the source vector into the sum of code vectors, one per stage. Historically, MSVQ was conceived as a sequential quantization operation where each stage simply quantizes the residual of the previous stage. More recently, the greedy nature of simple sequential encoding was recognized, and efficient techniques were proposed to seek better approximation of the source vector as combination of stage-vectors [6]. In [7], an unequally protected MSVQ design was proposed where the receiver estimates the channel conditions and decodes as many stages of the quantized signal as can be reliably decoded. In that work, whose motivation is wireless applications, the multi-stage coder is viewed as a tool for successive refinement, *i.e.,* to decode stage n, all stages up to $n-1$ need to be correctly decoded. In the current work, which addresses packet network applications, we directly optimize the MSVQ for general information loss patterns. It complements our preliminary work in speech coding where MSVQ is a standard tool [8].

This paper is organized as follows. In Section 2, we introduce basic means to exploit the robustness potential of MSVQ, via appropriate interleaving techniques. In Section 3, we propose a design technique to optimize the MSVQ for the given packet loss statistics. We then present simulation results and conclusions in Sections 4 and 5, respectively.

2. INTERLEAVING FOR ROBUST MSVQ

In this section we demonstrate the importance of appropriate interleaving to the robustness of MSVQ, and propose such an interleaving scheme. An L-stage MSVQ quantizes an input vector by searching for a set of indices $(a_0, a_1, a_2, \ldots, a_{L-1})$ pertaining to the (hopefully) best choice of stage-vectors for its representation. The process of finding a

This work is supported in part by the NSF under grant MIP-9707764, the University of California MICRO Program, Conexant Systems, Inc., Fujitsu Laboratories of America, Inc., Lernout & Hauspie Speech Products, Lucent Technologies, Inc., and Qualcomm, Inc.

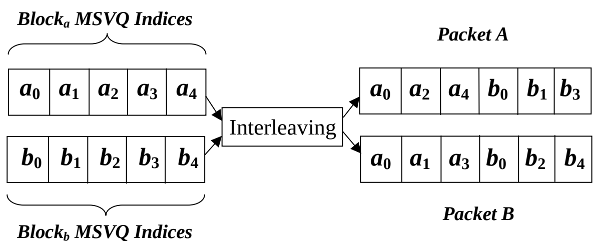

Figure 1: MSVQ parameter interleaving example with two data packets. Note that a_0 and b_0 are transmitted on both packets.

good combination of representative stage-vectors typically involves an "M-search" [6]: Instead of greedily selecting the best vector at the current stage, the decision is postponed, and M "survivors" are temporarily stored. In other words, the delayed decision allows the search procedure to seek the best combination of stage-vectors rather than to locally fit each stage vector. This fact coupled with the observation that, at high dimensionality, individual stage-vectors exhibit a high degree of mutual orthogonality, supports the potential benefits of decoding and utilizing all received stage-vectors even when some preceding indices are missing.

Let us first consider the performance of an MSVQ under various index loss patterns. Some typical experimental results are shown in Table 1. The reconstruction error is averaged over the test set. For a given L-stage MSVQ, the pattern of index losses can be described using an "index transmission vector" $T = (T_0, T_1, \ldots, T_{L-1})$, where

$$T_i = \begin{cases} 1 & \text{if index } i \text{ is correctly received} \\ 0 & \text{if index } i \text{ is lost.} \end{cases} \quad (1)$$

In the first column of Table 1, we list the index transmission vectors. The second and third columns list, for traditional MSVQ design and source-channel optimized design, respectively, the mean squared error (MSE) results for image block reconstruction subject to the corresponding index transmission vector, averaged over the test set. The last row in Table 1 shows the expected MSE that would result if the concealment method of [9] is used. The concealment method would be used when all block information is lost or discarded due to losses in transmission. It can be seen that a lower MSE is achieved by reconstructing using as many indices as received than having to resort to the concealment method. Table 1 also demonstrates further improvements of channel-optimized MSVQ over traditional MSVQ. In the next section, we will explain the design steps needed to achieve such an improvement.

Note that all transmission patterns in Table 1 indicate successful transmission of the first stage-index. The requirement that a_0 always be received is due to the fact that a_0 contains the most significant information, and the loss of a_0 often renders the remaining indices useless. For this reason, in some audio standards, a_0 is more heavily protected using FEC codes.

Shown in the second column of Table 1 is the performance for a traditional source-optimized MSVQ under various loss scenarios. The important fact we can deduce is that in the situation where individual indices may become lost, using any remaining received indices contributes to lower average distortion than discarding those indices. In the third column of Table 1, we show the performance for a source-channel optimized MSVQ where a stage index loss rate of 30% is assumed. It can be seen that MSE for the more severe losses is significantly decreased. On the other hand, as expected, MSE for the less severe losses is increased. When the whole transmission problem is cast in probabilistic terms, we find that the overall expected MSE for a source-channel optimization is much lower than that of a traditional source optimization.

Table 1: Average MSE in image reconstruction from partial received information.

Transmission Vector	Source-Optimized MSVQ	Source-Channel-Optimized MSVQ
	MSE	MSE
11111	22.82	29.50
11110	32.47	36.45
11101	39.74	39.77
11100	47.75	51.55
11011	57.12	44.29
11010	65.85	58.85
11001	72.15	64.15
11000	79.25	83.54
10111	148.17	47.66
10110	156.92	64.34
10101	163.20	70.94
10100	170.32	92.45
10011	178.03	78.02
10010	185.86	102.32
10001	191.18	110.89
10000	197.38	140.02
Concealment	196.29	196.29

We propose a framework for encoding the image blocks into data packets such that the decoding is more robust to channel losses. Table 1 strongly suggests that it is better to have a loss distributed over several image blocks than to have a complete loss of an image block, which would then require concealment as mentioned above. An interleaving method is the natural solution to this problem. See Fig. 1 for an example of interleaving of the MSVQ-parameters for two image blocks over two packets. Notice that a_0 and b_0 are transmitted on both packets. Although this redundancy increases the bit rate, it was found to contribute significantly to robustness.

Using the interleaving scheme of Fig. 1, and given packet loss rate of p, we have four different transmission vectors: 00000, 10101, 11010, 11111, with probabilities p^2, $p(1-p)$, $(1-p)p$, and $(1-p)^2$, respectively. The transmission vector 00000 occurs when both packets are lost, and in this case a concealment technique is applied. Note that in this case, a tailored source-channel optimization, where only the above four scenarios may occur, will achieve even better MSE than that shown in the third column of Table 1.

3. SOURCE-CHANNEL-OPTIMIZED MSVQ

In this section, we propose an optimization algorithm to further strengthen MSVQ for use on packet network channels under interleaving schemes such as the one described in Section 2.

The traditional approach to the design of MSVQ is an iterative decent algorithm alternating between two optimization steps. Given the current quantizer codebooks, the training set is encoded, and statistics for the input vectors for each stage quantizer and its decision are accumulated. Next, codebook entries are recalculated from the current encoding partition statistics. These two steps guarantee monotone decrease in reconstruction error, and the procedure converges to an at least locally optimal solution.

The aim of the new algorithm is to extend the design algorithm to handle packet loss. In fact, we average over all possible transmission vectors with respect to probabilities $P(T)$, where $T \in \mathcal{T}$, the set of possible transmission vectors. For an L-stage MSVQ, an input vector $X \in \mathcal{X}^n$ is encoded into L indices $(i_0, i_1, \ldots, i_{L-1})$ that are concatenated into one overall codeword $i \in \mathcal{I}$, where $\mathcal{I}$ is the set of all possible codewords. (For simplicity of presentation we assume that no error correcting codes are used). The channel may cause some indices to be lost, and its effect is given by the function $f : \mathcal{I} \times \mathcal{T} \to \mathcal{I}'$, where $\mathcal{I}'$ is the set of all possible received words. The encoder and decoder are denoted by $\alpha : \mathcal{X}^n \to \mathcal{I}$ and $\beta : \mathcal{I}' \to \mathcal{Y}^n$, respectively. The function β, thus, maps codeword i' to a reproduction vector $Y \in \mathcal{Y}^n$.

Encoder Optimization

An optimal encoder α^* maps each source vector X to the codeword that would minimize the expected distortion. If transmitted codeword i is subject to channel losses in the pattern of the transmission vector T, then following our notation, the received codeword is $i' = f(i, T)$. Thus, the optimal encoder is given by

$$\alpha^*(X) = \arg\min_{i \in \mathcal{I}} \{ \sum_{T \in \mathcal{T}} P(T) d(X, \beta(f(i,T))) \} \qquad (2)$$

where $d(\cdot,\cdot)$ is the distortion measure. Note that for simplicity of description we have made abstraction of the stagewise encoding search.

Decoder Optimization

Given α, we find the optimal decoder β^* that minimizes the expected distortion given the received channel codeword index $i' \in \mathcal{I}'$ and the transmission vector $T \in \mathcal{T}$:

$$\beta^*(i') = \arg\min_{Y \in \mathcal{Y}^n} \{ E_X[d(X,Y) \mid f(\alpha(X), T) = i'] \} \qquad (3)$$

and the explicit solution for the case of squared error distortion is given by

$$\beta^*(i') = \{ E_X[X \mid f(\alpha(X), T) = i'] \}. \qquad (4)$$

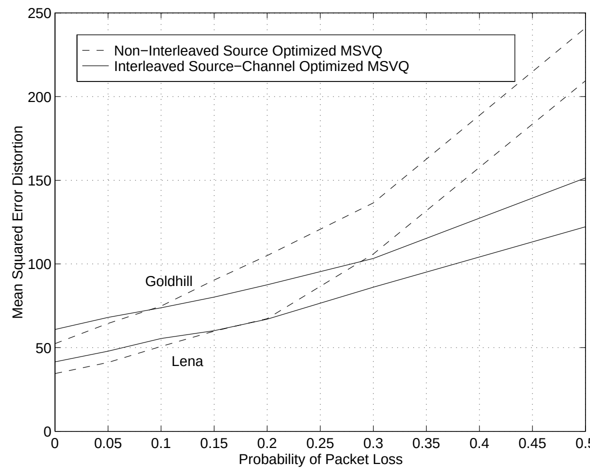

Figure 2: MSE performance versus packet loss rate on the *Lena* and *Goldhill* images

The iterative algorithm, which alternates between (2) and (4), must be initialized with some MSVQ. A reasonable choice of initialization is with an MSVQ that was optimized assuming a lossless channel (i.e., a source-optimized MSVQ). The actual algorithm we use here is implemented using a selective splitting procedure [10].

4. SIMULATION RESULTS

Block-based image coding techniques involve dividing up an image into blocks of uniform size. Then the blocks are subject to quantization (or transformation followed by quantization). A competitive image coder would most likely incorporate a variable rate quantizer. In order to simplify the experiment and avoid complicating factors, the quantizer is assumed to be fixed rate. (For a variable rate coder, synchronizing techniques can be used.)

For the experiment, we assume a packet-based communication channel. In traditional interleaving techniques, interleaving is performed on the scale of whole image blocks (or MSVQ-parameters thereof), which we call block-interleaving. In this work, we propose an interleaving scheme that operates on the much finer level of MSVQ-indices, which we refer to as MSVQ-index interleaving. The MSVQ-index interleaving scheme in Fig. 1 will be used.

The source-channel optimized MSVQ interleaving scheme was tested as follows: A group of eight images were used for training a 5-stage MSVQ according (2) and (4). Blocks of size 8 × 8 are rearranged into vectors. The performance of the proposed technique is evaluated on an independent test set of two independent images, *Lena* and *Goldhill*, and compared to traditional source-optimized non-interleaved MSVQ. In the proposed method, the MSVQ indices are obtained according to (2), and are interleaved and sent in data packets according to Fig. 1. For the traditional method, the MSVQ indices are obtained using source-optimized MSVQ and are non-interleaved. The interleaved method requires a repetition of the first index, so at a rate of 6 bits per stage, a total rate of 36 bits per block (0.56 bpp) is used. For a fair comparison with the interleaved method, an extra MSVQ stage is provided for the non-interleaved method to equate the bit rate. In the case of a complete loss of a block, a concealment method is used [9]. Performance is shown in Fig. 2 versus packet loss rates. A considerable gain of more than 2 dB is achieved at high packet loss rates. In addition, the subjective quality is significantly improved. A subjective comparison at a loss rate of 30% is given in Fig. 3. The subjective gains of our proposed method are maintained at lower packet loss rates (not shown here for lack of space). It can thus be concluded that the interleaved MSVQ method, when optimized for general loss patterns, can be very robust to packet loss transmission.

5. CONCLUSION

A new interleaving scheme is proposed for robust image transmission over lossy packet networks. Moreover, we developed an optimization algorithm for MSVQ design that explicitly accounts for packet losses. Both subjective and objective comparisons demonstrate the merit of the proposed approach.

6. REFERENCES

[1] Y. Wang and Q-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE,* vol. 86, no. 5, pp. 974-996, May 1998.

[2] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Information Theory,* vol. 39, no. 3, pp. 821-834, May 1993.

[3] Y. Wang, M. T. Orchard, and A. R. Reibman, "Optimal pairwise correlating transforms for multiple description coding," *ICIP'98,* Chicago, IL, vol. 1, pp. 679-683, Oct. 1998.

[4] M. Fleming and M. Effros, "Generalized multiple description vector quantization," *Proc. Data Compression Conference, DCC'99,* Snowbird, UT, pp.3-12, Mar. 1999.

[5] F. Kossentini, M. J. T. Smith, and C. F. Barnes, "Image coding using entropy-constrained residual vector quantization," *IEEE Trans. Image Processing,* vol. 4, no. 10, pp. 1349-1357, Oct. 1995.

[6] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech and Audio Proc.* vol. 1, no. 4, pp. 373-385, Oct. 1993.

[7] S. Gadkari and K. Rose, "Unequally protected multistage vector quantization for time-varying channels," *ICC'98,* Atlanta, GA, vol. 2, pp. 786-90, June, 1998.

[8] H. Khalil and K. Rose, "MSVQ design for packet networks with application to LSF quantization," *ICASSP'2000*, June 2000.

[9] Z. Wang, Y. Yu, and D. Zhang, "Best neighborhood matching: An information loss restoration technique for block-based image coding systems," *IEEE Trans. Image Proc.,* vol. 7, no. 7, pp. 1056-1061, July 1998.

[10] H. Khalil and K. Rose, "A selective splitting approach to entropy-constrained single/multi-stage vector quantization design," *Image and Video Communications and Processing 2000,* San Jose, CA, Jan. 2000.

(a)

(b)

Figure 3: Performance example at packet loss rate of 30%: a) Interleaved source-channel optimized MSVQ (28.78 dB), b) Non-interleaved source optimized MSVQ (27.88 dB).