

# A Mean-Removed Variation of Weighted Universal Vector Quantization for Image Coding \*

Barry D. Andrews  
Information Systems Laboratory  
Stanford University  
Stanford, CA 94305-4055  
email: andrews@isl.stanford.edu

Michelle Effros  
Information Systems Laboratory  
Stanford University  
Stanford, CA 94305-4055  
email: effros@isl.stanford.edu

Philip A. Chou  
Xerox Palo Alto Research Center  
3333 Coyote Hill Road  
Palo Alto, CA 94304  
email: chou@parc.xerox.com

Robert M. Gray  
Information Systems Laboratory  
Stanford University  
Stanford, CA 94305-4055  
email: gray@isl.stanford.edu

*Abstract* - Approaching a multi-codebook system as a *codebook* of codebooks, weighted universal vector quantization (WUVQ) uses traditional codeword design techniques to design locally optimal multi-codebook systems. Application of this technique to a sequence of medical images produces a 10.3 dB improvement over standard full search vector quantization followed by entropy coding at the cost of increased complexity. In this paper we propose a mean-removed variation of WUVQ. Each codebook in the system is given a "mean" or "prediction" value which is subtracted from all supervectors that map to the given codebook. The chosen codebook's codewords are then used to encode the resulting residuals. Application of the mean-removed system to the medical data set achieves up to .5 dB improvement over WUVQ at no rate expense.

---

\*This material is based upon work supported under a Natural Sciences and Engineering Research Council of Canada Scholarship and a Sony Corporation Fellowship, a National Science Foundation Graduate Fellowship, and National Science Foundation Grant MIP-9016974.

## 1 Introduction

In multi-codebook vector quantization systems, blocks or vectors of incoming data samples are characterized and sent to one of some number of codebooks. One example of a multi-codebook vector quantization system is a finite-state vector quantizer, in which the state, or the index of the current codebook, is determined from indices of previously quantized vectors. Thus the state can be tracked by the decoder without side information. Another example of a multi-codebook vector quantization system is classified vector quantization, in which the “state” or index of the current codebook is determined from characteristics of the current vector and its context. In this case the state must be transmitted as side information. Weighted universal vector quantization (WUVQ) is a multi-codebook vector quantization system in which the “state” or codebook index is periodically transmitted as side information, and held constant over the period. Thus all vectors within the supervector constituting the period are quantized with the same codebook. Moreover, the codebook is selected to reproduce the supervector with the lowest distortion. This increases the delay associated with encoding any single vector (the entire supervector must be encoded with each codebook before a codebook can be chosen), but decreases the amount of side information that needs to be sent. WUVQ can be regarded as a two stage system: the encoder sends the state or index of the codebook with the lowest distortion on the supervector, and then encodes each vector in the supervector with the given codebook.

In the proposed mean-removed variation of WUVQ, we attempt to exploit properties common to those supervectors that map to a single codebook. We therefore include a “mean” or “prediction” supervector for each codebook, and subtract that value from each incoming data supervector. The residual supervector is then broken into its corresponding vectors and encoded using the given codebook. The inclusion of a single “prediction” for each codebook costs nothing in terms of rate and the mean-removed variation includes as a special case (with zero-valued “predictions”) the WUVQ system.

## 2 Weighted Universal Vector Quantization

Given a class  $\Lambda$  of source distributions, a weighted universal quantizer is a series of block codes whose expected performance on any single source  $\theta \in \Lambda$  approaches, as the block size increases to infinity, the asymptotic performance of a system designed explicitly for  $\theta$ -distributed processes. To be more exact, let  $\{X_i\}$  be a random process over alphabet  $\mathcal{X}$  with process measure  $P_\theta$ ,  $\theta \in \Lambda$ , and let  $\beta^N \circ \alpha^N$  be a length- $N$  block code with encoder  $\alpha^N : \mathcal{X}^N \rightarrow \mathcal{C}$  and decoder  $\beta^N : \mathcal{C} \rightarrow \mathcal{Y}^N$ , where  $\mathcal{C} = \{c_1, \dots, c_M\} \subseteq \{0, 1\}^*$  is some binary prefix code and  $\mathcal{Y}$  is the reproduction alphabet, typically equal to  $\mathcal{X}$ . A weighted universal source code with respect to a fidelity criterion  $d(X^N, Y^N) = \sum_i d(X_i, Y_i)$  is a sequence of block codes  $\{(\alpha^N, \beta^N)\}$  such that for each  $\theta \in \Lambda$  there exists a corresponding sequence of points  $\{(R_{N,\theta}, D_{N,\theta})\}$  on the graph of the  $N$ th order operational distortion-rate function for

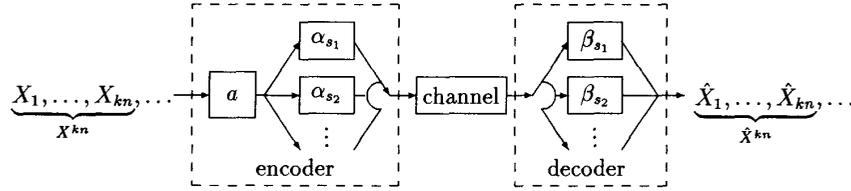


Figure 1: Weighted Universal VQ

$P_\theta$  for which the per letter “rate” redundancy

$$\frac{1}{N} E_\theta |\alpha^N(X^N)| - R_{N,\theta} \quad (1)$$

and the per-letter “distortion” redundancy

$$\frac{1}{N} E_\theta d(X^N, \beta^N(\alpha^N(X^N))) - D_{N,\theta} \quad (2)$$

each go to zero in expectation with respect to a probability measure (weighting) on  $\theta$ . Here  $|\alpha^N(X^N)|$  is used to denote the length of the binary code  $\alpha^N(X^N)$ .

In [1], Neuhoff, Gray, and Davisson show that a weighted universal source code can be constructed from a sequence of multi-codebook vector quantizers; in [2], Chou develops an algorithm for the design of weighted universal vector quantizers, which are locally optimal multi-codebook vector quantizers for weighted universal source coding; and in [3], Chou and Effros provide bounds on the rates at which the expected rate and distortion redundancies (1) and (2) go to zero for WUVQ. In this section we describe WUVQ and its design algorithm.

Consider the set of codebooks  $\{\beta_s \mid s \in \mathcal{S}\}$  as a *codebook* of codebooks, where  $\mathcal{S} = \{s_1, \dots, s_M\} \subseteq \{0, 1\}^*$  is a binary prefix code. In the WUVQ system, shown in Figure 1, the encoder breaks the source sequence into supervectors  $X^{kn}$ , consisting of  $n$  vectors  $X_1^k, \dots, X_n^k$ , each of dimension  $k$ . The encoder  $a$  maps each  $X^{kn}$  to a codebook  $\beta_s$ , and sends the binary code  $s$  to the decoder. Each  $k$ -dimensional vector in the supervector is then encoded with the encoder  $\alpha_s$  associated with the chosen codebook  $\beta_s$ . The decoder receives  $s$  and then decodes each  $k$ -dimensional vector using  $\beta_s$ .

Given a codebook  $\beta_s$ , the distortion associated with coding the supervector  $X^{kn}$  is

$$d(X^{kn}, \beta_s) = \sum_{i=1}^n d(X_i^k, \beta_s(\alpha_s(X_i^k)))$$

and the corresponding rate is

$$r(X^{kn}, \beta_s) = \sum_{i=1}^n |\alpha_s(X_i^k)|.$$

The function  $a$  maps the supervector  $X^{kn}$  to a codebook index  $s \in \mathcal{S}$ , which is sent to the decoder along with the encoded data. The per symbol average distortion in

this system is

$$D_{kn} = \frac{1}{kn} E[d(X^{kn}, \beta_{a(X^{kn})})],$$

and the per symbol average rate is

$$R_{kn} = \frac{1}{kn} E[|a(X^{kn})| + r(X^{kn}, \beta_{a(X^{kn})})],$$

i.e., the length of the codebook description plus the length of the data description given the codebook.

Locally optimal design for both fixed and variable rate WUVQ systems is achieved using the generalized Lloyd algorithm. For a fixed rate WUVQ, with fixed length  $s$  and  $\beta_s$ , the optimal encoder satisfies the nearest neighbor condition

$$a(X^{kn}) = \arg \min_s d(X^{kn}, \beta_s)$$

and the optimal decoder satisfies the centroid condition

$$\beta_s = \arg \min_{\beta} E[d(X^{kn}, \beta) | a(X^{kn}) = s].$$

The generalized Lloyd algorithm iteratively optimizes the encoder for a given decoder and then the decoder for the resulting encoder until the system converges to an optimal solution.

For a variable rate (entropy-constrained) WUVQ, the optimal encoder satisfies

$$a(X^{kn}) = \arg \min_s [d(X^{kn}, \beta_s) + \lambda(|s| + r(X^{kn}, \beta_s))]$$

and the optimal decoder satisfies

$$\beta_s = \arg \min_{\beta} E [d(X^{kn}, \beta) + \lambda(|s| + r(X^{kn}, \beta)) | a(X^{kn}) = s],$$

where  $\lambda$  is a Lagrangian constant chosen as a function of the desired final rate. Here too, we employ the generalized Lloyd algorithm to iteratively optimize the encoder and decoder in a manner parallel to that of [4].

### 3 A Mean-Removed Variation of WUVQ

In the mean-removed variation of WUVQ, we alter the WUVQ structure (Figure 1) by associating with each codebook  $\beta_s$ , a  $kn$ -dimensional supervector prediction  $P_s^{kn}$ . The encoder  $a$  maps each supervector  $X^{kn}$  to a codebook/prediction pair  $(\beta_s, P_s^{kn})$ . Given an index  $s$ , we encode a supervector  $X^{kn}$  by first subtracting  $P_s^{kn}$  and then encoding the component vectors of the residual supervector  $(X^{kn} - P_s^{kn})$  using  $\beta_s$ . Since only one supervector prediction is included with each codebook, the addition of predictions to the system requires no additional side information to be sent from encoder to decoder.

We now assume that  $d$  is the squared-error fidelity criterion. The distortion and rate associated with coding the supervector  $X^{kn}$  with the codebook/prediction pair  $(\beta_s, P_s^{kn})$  are

$$d(X^{kn}, (\beta_s, P_s^{kn})) = \sum_{i=1}^n d(X_i^k - P_{s,i}^k, \beta_s(\alpha_s(X_i^k - P_{s,i}^k)))$$

and

$$r(X^{kn}, (\beta_s, P_s^{kn})) = \sum_{i=1}^n |\alpha_s(X_i^k - P_{s,i}^k)|,$$

and the corresponding per symbol average rate and distortion are

$$D_{kn} = \frac{1}{kn} E[d(X^{kn}, (\beta_{a(X^{kn})}, P_{a(X^{kn})}^{kn}))],$$

and

$$R_{kn} = \frac{1}{kn} E[|a(X^{kn})| + r(X^{kn}, (\beta_{a(X^{kn})}, P_{a(X^{kn})}^{kn}))],$$

respectively.

To design a locally optimal fixed rate mean-removed system, we again use a variation of the generalized Lloyd algorithm, iteratively improving the encoder for an existing set of predictions and codebooks, each prediction for a given encoder and codebook, and then each codebook for an existing encoder and prediction. For a given number of codebooks and a given fixed rate per codebook, the optimal  $(a, \{\beta_s, P_s^{kn}\})$  minimizes  $E[d(X^{kn}, (\beta_{a(X^{kn})}, P_{a(X^{kn})}^{kn}))]$ . We therefore employ a descent algorithm on  $E[d(X^{kn}, (\beta_{a(X^{kn})}, P_{a(X^{kn})}^{kn}))]$  to achieve a locally optimal solution. Using  $\tilde{a}$ ,  $\tilde{P}_s^{kn}$ , and  $\tilde{\beta}_s$  to represent the updated versions of  $a$ ,  $P_s^{kn}$ , and  $\beta_s$ , the descent algorithm proceeds as follows:

1.  $\tilde{a}(X^{kn}) = \arg \min_s d(X^{kn}, (\beta_s, P_s^{kn}))$
2.  $\tilde{P}_{s,i}^k = E[X_i^k - \beta_s(X_i^k - P_{s,i}^k) | \tilde{a}(X^{kn}) = s]$
3.  $\tilde{\beta}_s = \arg \min_{\beta} E[d(X^{kn}, (\beta, \tilde{P}_s^{kn})) | \tilde{a}(X^{kn}) = s]$
4. Iterate until convergence.

Given a target rate, the optimal variable rate mean-removed system is the system that minimizes the distortion subject to a constraint on the rate. The generalized Lloyd algorithm is thus employed to minimize the Lagrangian

$$E[d(X^{kn}, (\beta_s, P_s^{kn})) + \lambda r(X^{kn}, (\beta_s, P_s^{kn}))],$$

where  $\lambda$  is the Lagrangian coefficient and rate is calculated in terms of entropy. Thus the iterative design strategy becomes

1.  $\tilde{a}(X^{kn}) = \arg \min_s [d(X^{kn}, (\beta_s, P_s^{kn})) + \lambda(|s| + r(X^{kn}, (\beta_s, P_s^{kn})))]$
2.  $\tilde{P}_{s,i}^k = E[X_i^k - \beta_s(X_i^k - P_{s,i}^k) | \tilde{a}(X^{kn}) = s]$
3.  $\tilde{\beta}_s = \arg \min_{\beta} E [d(X^{kn}, (\beta, \tilde{P}_s^{kn})) + \lambda(|s| + r(X^{kn}, (\beta, \tilde{P}_s^{kn}))) | \tilde{a}(X^{kn}) = s]$
4. Iterate until convergence.

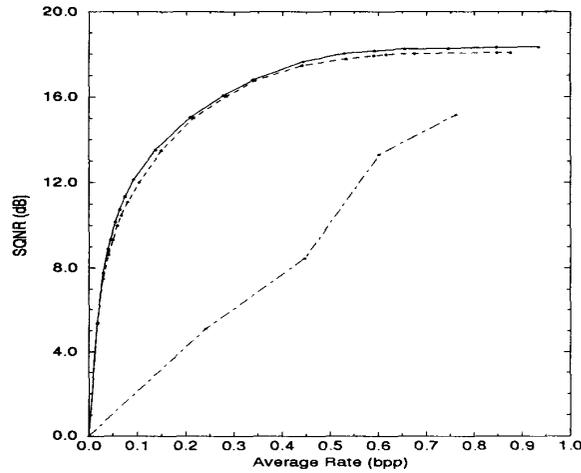


Figure 2: Entropy constrained mean-removed WUVQ, entropy constrained WUVQ, and standard full-search VQ on medical test sequence.

— variable rate mean-removed WUVQ, - - variable rate WUVQ,  
 -.- standard full search VQ

## 4 Experimental Results

WUVQ and its mean-removed variation were tested on a sequence of medical images and compared to standard full search VQ. The training sequence consisted of 20 magnetic resonance (MR) images and the test sequence consisted of 5 MR images outside of the training sequence. Performance results are reported in terms of signal-to-quantization-noise ratio (SQNR), where

$$\text{SQNR} = 10 \log_{10} \frac{\text{VAR}(X)}{D},$$

$X$  is the pixel intensity, and  $D$  is the average squared-error between the 9-bpp original and the quantized reproduction.

For both WUVQ and the mean-removed variation of WUVQ, supervectors consisted of 16  $2 \times 2$  vectors and thus the dimension of the predictions for the mean-removed case was 64 ( $8 \times 8$ ). All multi-codebook systems consisted of at most 32 codebooks, each with no more than 16 codewords.

Figure 2 shows the performance of variable rate WUVQ and the variable rate mean-removed variation of WUVQ with varying values of  $\lambda$ . Also in that figure is the performance of standard full search VQ with a  $2 \times 2$  vector dimension. All rates are reported in terms of entropy. At low rates WUVQ shows a 10.3 dB improvement over standard VQ. The mean-removed variation of WUVQ shows up to .5 dB improvement over WUVQ.

Figure 3 shows an original test image alongside the same image compressed to



Figure 3: Original and compressed using the mean-removed variation of WUVQ at .51 bpp, SQNR = 15.67 dB

.51 bpp using the mean-removed variation of WUVQ.

## 5 Conclusion

A mean-removed variation of WUVQ has been presented. The algorithm exploits the similarity between supervectors mapped to a specific codebook in a WUVQ system by subtracting a prediction from each supervector before encoding. A single index is used to specify both the codebook and the prediction, and thus no additional side information is required. The algorithm achieves at low rates a 10.5 dB improvement over standard full search VQ followed by entropy coding and up to .5 dB improvement over WUVQ. Both WUVQ and the mean-removed variation can be generalized to multi-level systems.

## References

- [1] D. L. Neuhoff, R. M. Gray, and L. D. Davisson. Fixed rate universal block source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 21(5):511–523, September 1975.
- [2] P. A. Chou. Code clustering for weighted universal VQ and other applications. In *Proceedings of the IEEE International Symposium on Information Theory*, page 253, Budapest, Hungary, June 1991.

- [3] P. A. Chou and M. Effros. Rate and distortion redundancies for source coding with respect to a fidelity criterion. In *IEEE International Symposium on Information Theory*, San Antonio, Texas, January 1993.
  - [4] P. A. Chou, T. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(1):31-42, January 1989.
-