Open access • Proceedings Article • DOI:10.1109/DCC.1994.305907

Variable dimension weighted universal vector quantization and noiseless coding — Source link

Michelle Effros, Philip A. Chou, Robert M. Gray

Institutions: Stanford University

Published on: 29 Mar 1994 - Data Compression Conference

Topics: Vector quantization, Linde–Buzo–Gray algorithm, Signal-to-quantization-noise ratio and Codebook

Variable Dimension Weighted Universal Vector Quantization and Noiseless Coding *

M. Effros,* P. A. Chou,[†] and R. M. Gray[‡]

* *Information Systems Laboratory, Stanford, CA 94305-4055,*
phone: (415) 723-2675, fax: (415) 723-8473, email: effros@isl.stanford.edu

[†] *Xerox Palo Alto Research Center, 3333 Coyote Road, Palo Alto, CA 94304*
phone: (415) 812-4496, fax: (415) 812-4241, email: chou@parc.xerox.com

[‡] *Information Systems Laboratory, Stanford, CA 94305-4055,*
phone: (415) 723-4001, fax: (415) 723-8473, email: gray@isl.stanford.edu

Abstract

A new algorithm for variable dimension weighted universal coding is introduced. Combining the multi-codebook system of weighted universal vector quantization (WUVQ), the partitioning technique of variable dimension vector quantization, and the optimal design strategy common to both, variable dimension WUVQ allows mixture sources to be effectively carved into their component subsources, each of which can then be encoded with the codebook best matched to that source. Application of variable dimension WUVQ to a sequence of medical images provides up to 4.8 dB improvement in signal to quantization noise ratio over WUVQ and up to 11 dB improvement over a standard full-search vector quantizer followed by an entropy code. The optimal partitioning technique can likewise be applied with a collection of noiseless codes, as found in weighted universal noiseless coding (WUNC). The resulting algorithm for variable dimension WUNC is also described.

1 Introduction

Weighted universal vector quantizers and weighted universal noiseless codes [1, 2, 3] are two-stage codes. In each, an incoming data string is broken into supervectors, which are blocks of vectors. All of the vectors in a particular supervector are encoded with a single code. Given some collection of codes, the encoder chooses the code that

*This material is based upon work partially supported by an AT&T Bell Laboratories Ph.D. Scholarship and by a grant from the Center for Telecommunications at Stanford. The authors would also like to thank Professor Eve Riskin for the use of an HP 720 donated by Hewlett Packard Laboratories.

most closely matches the statistics of the vectors in the supervector at hand. The first stage of a two-stage code describes the chosen quantizer or noiseless code, and the second stage describes the data using that code. By keeping a variety of codes and allowing the code in use to change from supervector to supervector, WUVQs and WUNCs can roughly track local variation in source statistics. The expense of tracking this information is the side information necessary to describe the code used on each supervector.

While WUVQ and WUNC use fixed supervector lengths, variable dimension WUVQ (VDWUVQ) and variable dimension WUNC (VDWUNC) employ the optimal partitioning algorithm of variable dimension VQ (VDVQ) [4, 5] to break the image into variable sized supervectors in an optimal fashion. The basic algorithm proceeds as follows. An incoming data stream is partitioned into variable length supervectors. The encoder then describes these supervectors, one by one, to the decoder. To describe a supervector, the encoder first sends the length of the supervector, then describes the index of the code with which that supervector's component vectors will be encoded, and finally describes the component vectors using the chosen code. The optimal partition is the partition that minimizes the overall distortion subject to a constraint on the rate, which now includes the rate associated with describing the length of each supervector. The employment of optimally chosen supervectors of variable sizes allows the VDWUVQ to better carve the data into its component subsources, each of which is then encoded separately by an encoder matched to that source.

Section 2 reviews the basic WUVQ and WUNC algorithms. The VDWUVQ and VDWUNC algorithms are then described in section 3. The final section presents results achieved by applying the VDWUVQ algorithm to a set of medical images. While both fixed- and variable-rate versions of VDWUVQ and VDWUNC are possible, this paper focuses on the variable-rate case.

2 The WUVQ and WUNC Algorithms

Given a class, Λ, of process measures and a distribution over that class, a weighted universal code is a sequence of block codes for which the expected performance over all $\theta \in \Lambda$ approaches, with increasing block size, the theoretically optimal expected performance. To be more exact, let $\{X_i\}$ be a random process with alphabet $\mathcal{X}$ and process measure P_θ, $\theta \in \Lambda$, and let $\beta^n \circ \alpha^n$ be a length-n block code with encoder $\alpha^n : \mathcal{X}^n \rightarrow \mathcal{S}$ and decoder $\beta^n : \mathcal{S} \rightarrow \hat{\mathcal{X}}^n$, where $\mathcal{S} \subset \{0,1\}^*$ is a binary prefix code, $|\alpha^n(x^n)|$ is the length of the code for x^n, and $\hat{\mathcal{X}}$ is the reproduction alphabet. In weighted universal quantization, $\hat{\mathcal{X}}$ is not required to equal $\mathcal{X}$, $\mathcal{X}$ may be uncountably infinite, and $\beta^n(\alpha^n(x^n)) = \hat{x}^n$ with distortion $d_n(x^n, \hat{x}^n) = \sum_i d_1(x_i, \hat{x}_i)$, where, $d_1(x_i, \hat{x}_i)$ is an additive fidelity criterion. In noiseless coding, $\hat{\mathcal{X}} = \mathcal{X}$, $\mathcal{X}$ is finite or countably infinite, and $\beta^n(\alpha^n(x^n)) = x^n$ for all $x^n \in \mathcal{X}$. A weighted universal source code 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 nth order

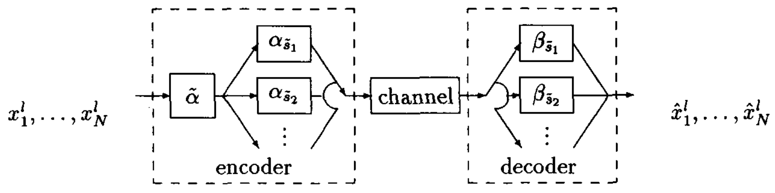

Figure 1: Weighted Universal VQ

operational distortion-rate function for P_θ for which the per letter "rate" redundancy

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

and the per-letter "distortion" redundancy

$$\frac{1}{n}E_\theta d_n(X^n, \beta^n(\alpha^n(X^n))) - D_{n,\theta}$$

each go to zero in expectation with respect to a probability measure (weighting) on θ. Notice that the distortion redundancy requirement is trivially satisfied for all noiseless codes.

In [6], Neuhoff, Gray, and Davisson show that a weighted universal source code can be constructed from a sequence of multi-codebook vector quantizers; in [1], 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 [7], Chou and Effros provide bounds on the rates at which the expected rate and distortion redundancies go to zero for WUVQ. Two-stage universal noiseless codes were earlier developed in [8] and [9]. In this section we describe WUVQs and WUNCs and the algorithm by which they can be optimally designed.

The WUVQ and WUNC algorithms [1, 2, 3] treat the optimal design of two-stage weighted universal vector quantizers and noiseless coders. In both cases, the first stage of the system is a "quantizer" which maps the input space of possible data vectors to the output space of possible block codes; the chosen second-stage block code then maps the data vector to the output space of possible reproductions. In universal variable-rate quantization the block codes are variable-rate quantizers, while in universal noiseless coding, these block codes are block noiseless codes.

Figure 1 shows a two-stage coder. The vector sequence $x_1^l, \ldots, x_N^l \in \mathbb{R}^l$ is partitioned evenly into supervectors $y_1, \ldots, y_K \in \mathbb{R}^{ln}$, where each supervector contains n l-vectors. For simplicity it is assumed that n evenly divides N, so that $K = N/n$ is an integer. The first-stage quantizer contains an encoder $\tilde{\alpha} : \mathcal{X}^{ln} \to \tilde{\mathcal{S}}$ which maps the input space of possible supervectors to the output space of code descriptors. The first-stage description of the supervector y_i, $\tilde{s} = \tilde{\alpha}(y_i) \in \tilde{\mathcal{S}}$, indexes the chosen length-l block code used to code the l-vectors of y_i. We use $|\tilde{s}|$ to denote the first-stage description length. The second-stage block code contains an encoder and a decoder. Let $\{\alpha_{\tilde{s}} : \tilde{s} \in \tilde{\mathcal{S}}\}$ and $\{\beta_{\tilde{s}} : \tilde{s} \in \tilde{\mathcal{S}}\}$ be the collections of block encoders and decoders respectively. For any $\tilde{s} \in \tilde{\mathcal{S}}$, $\alpha_{\tilde{s}} : \mathcal{X}^l \to \mathcal{S}_{\tilde{s}}$ and $\beta_{\tilde{s}} : \mathcal{S}_{\tilde{s}} \to \hat{\mathcal{X}}^l$. Denote the component vectors of supervector y_i by $y_{i,1}, \ldots, y_{i,n}$. If $\tilde{s}$ is the first-stage description of y_i, then $s_{\tilde{s}} = \alpha_{\tilde{s}}(y_{i,j}) \in \mathcal{S}_{\tilde{s}}$ is the second-stage description of component vector $y_{i,j}$, $|s_{\tilde{s}}|$ is the length of the second-stage description of $y_{i,j}$, and $d_l(y_{i,j}, \beta_{\tilde{s}}(\alpha_{\tilde{s}}(y_{i,j})))$ is the distortion incurred by using the block code indexed by $\tilde{s}$ on $y_{i,j}$. The distortion is always zero in noiseless coding. The total distortion and rate associated with encoding supervector y_i (a block of n l-vectors) with codebook $\tilde{s}$ are

$$d(y_i, \tilde{s}) = \sum_{j=1}^{n} d_l(y_{i,j}, \beta_{\tilde{s}}(\alpha_{\tilde{s}}(y_{i,j})))$$

and

$$r(y_i, \tilde{s}) = |\tilde{s}| + \sum_{j=1}^{n} |\alpha_{\tilde{s}}(y_{i,j})|.$$

For two-stage noiseless codes and two-stage variable-rate quantizers, $\tilde{\mathcal{S}}$ and $\mathcal{S}_{\tilde{s}}$ are Huffman codes matched to the probabilities of the block codes and codewords, respectively.

Two-stage source codes are designed using an extension of the generalized Lloyd algorithm to minimize

$$\sum_{i=1}^{K} \sum_{\tilde{s} \in \tilde{\mathcal{S}}} J_\lambda(y_i, \tilde{s}) 1(\tilde{\alpha}(y_i) = \tilde{s})$$

on a training sequence, where $1(A)$ is 1 if statement A is true and 0 otherwise, the Lagrangian functional J_λ is defined as

$$J_\lambda(y_i, \tilde{s}) = d(y_i, \tilde{s}) + \lambda r(y_i, \tilde{s}),$$

and λ is the Lagrange multiplier. In noiseless coding, where the distortion is always equal to zero, λ is chosen to equal one, so that J_λ equals the total rate. In variable-rate vector quantization, nonnegative λ values are chosen based on the desired final rate of the system. The optimization process consists of a nested descent algorithm in which the Lloyd algorithm's nearest neighbor and centroid conditions are employed on both stages of the system. In the first-stage quantizer, satisfying the nearest neighbor condition means mapping each supervector to the codebook that will best encode that supervector's component vectors. In the second stage, satisfying the nearest neighbor condition means mapping each vector to the best codeword. In both cases, "best" is measured by the functional J_λ. By satisfying the centroid condition, we optimize the decoder for a given encoder. The first-stage quantizer meets this condition by redesigning each codebook in its collection of codebooks to match the statistics of the data mapped to that codebook. Each second-stage decoder likewise meets this condition by redesigning each of its codewords to match the data mapped to it.

3 The VDWUVQ and VDWUNC Algorithms

VDWUVQ and VDWUNC are two-stage coding systems where the length of the supervectors is allowed to vary. An incoming data sequence, $x_1^l, \ldots, x_N^l$, is broken into supervectors of varying length, say $y_1, \ldots, y_K$, of lengths $|y_1|, \ldots, |y_K|$ vectors. The number of supervectors and their lengths are subject only to the constraint that the sum of the lengths must equal the initial data length ($\sum_i |y_i| = N$). Using x_1^{lN} to denote the sequence $x_1^l, \ldots x_N^l$, let $\{y_i\} \prec x_1^{lN}$ mean that $\{y_i\}$ partitions x_1^{lN} and satisfies the above constraint. Since the supervector length is allowed to vary, it must be described to the decoder along with the first- and second-stage code information. The encoder thus uses an entropy code γ to describe that length to the decoder, where $|\gamma(n)|$ is the rate associated with describing a supervector length of n. Notice that if $|\gamma(\cdot)|$ is zero for length n and infinity otherwise, VDWUVQ and VDWUNC will behave exactly like WUVQ and WUNC. Thus VDWUVQ and VDWUNC can only exceed their predecessors in performance, but they do so at the expense of the additional complexity necessary for obtaining the optimal partition.

The optimal design algorithm for VDWUVQs and VDWUNCs is again a variation on the generalized Lloyd algorithm, which iteratively achieves descent on the weighted sum of the total distortion and rate

$$\sum_{i=1}^{K} \sum_{\tilde{s} \in \tilde{\mathcal{S}}} J_\lambda(y_i, \tilde{s}) 1(\tilde{\alpha}(y_i) = \tilde{s}),$$

where the Lagrangian is now

$$J_\lambda(y_i, \tilde{s}) = d(y_i, \tilde{s}) + \lambda(r(y_i, \tilde{s}) + |\gamma(|y_i|)|).$$

Let $|\{y_i\}|$ denote the number of elements in the partition, then the training algorithm proceeds as follows:

1. Optimally parse the data for fixed γ, $\tilde{\alpha}$, $\tilde{\mathcal{S}}$, and $\{(\alpha_{\tilde{s}}, \beta_{\tilde{s}}, \mathcal{S}_{\tilde{s}})\}$ and then update γ to match the new partition. The optimal partition satisfies

   $$\{y_i\} = \arg \min_{\{z_i\} \prec x_1^{lN}} \sum_i \sum_{\tilde{s} \in \tilde{\mathcal{S}}} J_\lambda(z_i, \tilde{s}) 1(\tilde{\alpha}(z_i) = \tilde{s}).$$

   Thus update γ to satisfy $|\gamma(n)| = \lceil -\log_2 [\sum_i 1(|y_i| = n) / |\{y_i\}|] \rceil$.

2. Optimize the first-stage encoder $\tilde{\alpha}$ for fixed $\{y_i\}$, γ, $\tilde{\mathcal{S}}$, and $\{(\alpha_{\tilde{s}}, \beta_{\tilde{s}}, \mathcal{S}_{\tilde{s}})\}$ and then redesign the prefix code. The resulting encoder is

   $$\tilde{\alpha}(y_i) = \arg \min_{\tilde{s} \in \tilde{\mathcal{S}}} [d(y_i, \tilde{s}) + \lambda r(y_i, \tilde{s})],$$

   and for each $\tilde{s} \in \tilde{\mathcal{S}}$ its new prefix code satisfies $|\tilde{s}| = \lceil -\log_2 [\sum_i 1(\tilde{\alpha}(y_i) = \tilde{s}) / |\{y_i\}|] \rceil$.

3. Optimize each block code for fixed $\{y_i\}$, γ, $\tilde{\alpha}$, and $\tilde{\mathcal{S}}$. This is accomplished for each $\tilde{s} \in \tilde{\mathcal{S}}$ by iteratively optimizing $\alpha_{\tilde{s}}, \beta_{\tilde{s}}$, and $\mathcal{S}_{\tilde{s}}$ until convergence, where the optimal encoder, decoder, and prefix code respectively satisfy the following three conditions. (Note that the change to the decoder has no effect in the case of noiseless coding.)

   $$\alpha_{\tilde{s}}(y_{i,j}) = \arg \min_{s_{\tilde{s}} \in \mathcal{S}_{\tilde{s}}} [d_l(y_{i,j}, \beta_{\tilde{s}}(s_{\tilde{s}})) + \lambda |s_{\tilde{s}}|] \qquad \forall y_{i,j} \text{ s.t. } \tilde{\alpha}(y_i) = \tilde{s}.$$

$$\beta_{\tilde{s}}(s_{\tilde{s}}) = \arg\min_{\hat{x}^l \in \hat{\mathcal{X}}^l} \sum_i \sum_j d_l(y_{i,j}, \hat{x}^l) 1(\alpha_{\tilde{s}}(y_{i,j}) = s_{\tilde{s}}) 1(\tilde{\alpha}(y_i) = \tilde{s}) \qquad \forall s_{\tilde{s}} \in \mathcal{S}_{\tilde{s}}$$

$$|s_{\tilde{s}}| = \left\lceil -\log_2 \left[\sum_i \sum_j 1(\alpha_{\tilde{s}}(y_{i,j}) = s_{\tilde{s}}) 1(\tilde{\alpha}(y_i) = \tilde{s}) / \sum_i \sum_j 1(\tilde{\alpha}(y_i) = \tilde{s}) \right] \right\rceil \qquad \forall s_{\tilde{s}} \in \mathcal{S}_{\tilde{s}}$$

4. Iterate steps 1-3 until convergence.

The optimal partition in step 1 of the above described algorithm is accomplished using the dynamic programming argument of VDVQ [5]. The argument proceeds as follows. Let $J_n = \min_{\{y_i\} \prec x_1^{ln}} \sum_i J_\lambda(y_i, \tilde{\alpha}(y_i))$ be the score of the best partitioning of the partial data sequence $x_1^l, \ldots, x_n^l$. If B is the maximum allowed supervector length, then for $1 \leq n \leq N$,

$$J_n = \min_{b \in \{1,\ldots,B\}} \left\{ J_{n-b} + \min_{\tilde{s} \in \tilde{\mathcal{S}}} J_\lambda(x_{n-b+1}^{lb}, \tilde{s}) \right\},$$

where $J_0 = 0$ and $J_n = \infty$ for $n < 0$. Now let b_n be the last supervector in the optimal partition of x_1^n, that is the supervector of length $\arg\min_{1 \leq b \leq B} \left\{ J_{n-b} + \min_{\tilde{s} \in \tilde{\mathcal{S}}} J_\lambda(x_{n-b+1}^{lb}, \tilde{s}) \right\}$ ending with the nth vector. Then the optimal partition of x_1^{lN} can be found by backtracking:

$$\{y_i\} = \{\ldots b_{N-|b_N|-|b_{N-|b_N|}|}, b_{N-|b_N|}, b_N\}.$$

While partitioning and encoding have been described as separate processes, they can be accomplished simultaneously (at the expense of greater storage requirements) by tracking not only the optimal performance and last supervector length at each time n but also the encoding information for that supervector.

4 Experimental Results

We here compare the performance of VDWUVQ with that of WUVQ and standard full-search VQ. The data sequence considered contains 25 magnetic resonance (MR) brain scans at an original 9 bpp. The first 20 images are used for training and the final 5 are used for testing. All of the vector quantizers use vectors of dimension 2×2 and the vectors are ordered according to a Peano scan (e.g., [10]). The supervectors in VDWUVQ are allowed to contain up to 16 vectors while the WUVQ has a constant supervector size of 4. All multi-codebook systems consisted of at most 256 codebooks, each with no more than 4 codewords.

Figure 2 shows the performance of variable-rate VDWUVQ and WUVQ with varying values of λ and the performance of standard full-search VQ. All rates are reported in terms of entropy. Signal to quantization noise ratio (SQNR) is calculated as $-10\log(D/D_0)$, where D is the current distortion, D_0 is the zero rate distortion of a standard full-search VQ, and all distortion is measured by the squared error fidelity criterion. VDWUVQ shows its greatest improvement over WUVQ at very low rates, below .1 bpp, with an SQNR up to 4.8 dB higher. Both VDWUVQ and WUVQ achieve their greatest gains over standard full-search VQ at slightly higher rates.

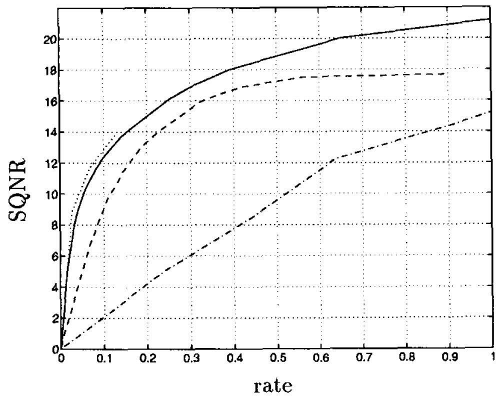

Figure 2: VDWUVQ, WUVQ, and standard full-search VQ on medical test sequence.
... variable-rate VDWUVQ with maximum supervector size 256,
— variable-rate VDWUVQ with maximum supervector size 16,
- - variable-rate WUVQ, -.- standard full-search VQ followed by an entropy code

VDWUVQ shows up to 11 dB improvement over standard VQ. Allowing supervector lengths up to 256 vectors improves performance by about another dB at the expense of increased computation.

Figure 3 compares the images resulting from quantizing the data at around .2 bpp, while figure 4 demonstrates the sizes and shapes of the supervectors used by the VDWUVQ algorithm in the previous figure.

References

[1] 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.

[2] B. D. Andrews, M. Effros, P. A. Chou, and R. M. Gray. A mean-removed variation of weighted universal vector quantization for image coding. In *Proceedings of the Data Compression Conference*, pages 302–309, Snow Bird, UT, April 1993. IEEE Computer Society.

[3] P.A. Chou, M. Effros, and R.M. Gray. A vector quantization approach to universal noiseless coding and quantization. *IEEE Transactions on Information Theory*. In preparation.

[4] P. A. Chou and T. Lookabaugh. Locally optimal variable-to-variable length source source coding with respect to a fidelity criterion. In *Proceedings of the IEEE International Symposium on Information Theory*, page 238, Budapest, Hungary, June 1991.

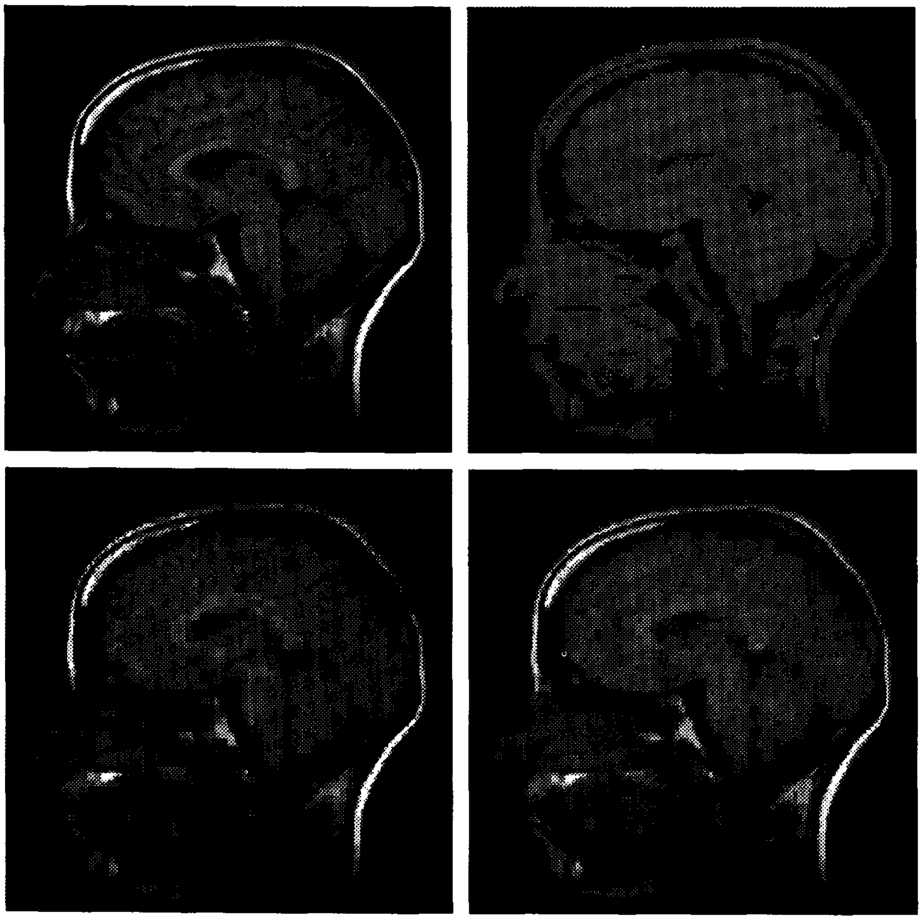

Figure 3: From top left to bottom right: The original 9 bpp image, standard VQ at .25 bpp and SQNR = 6.75 dB, WUVQ at .184 bpp and SQNR = 11.6 dB, and VDWUVQ at .166 bpp and SQNR = 13.2 dB.

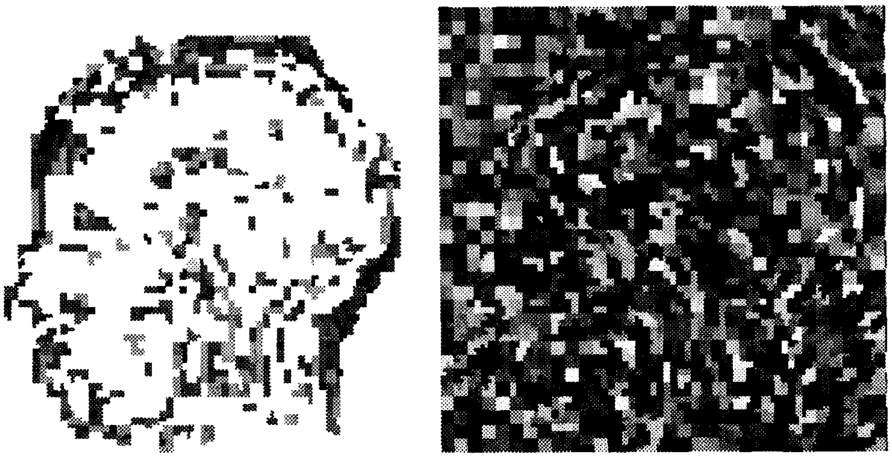

Figure 4: Supervector sizes and shapes for the .166 bpp image. The picture on the left shows the size of the supervector in which each pixel was coded, with supervector sizes from 1 to 16 vectors denoted in shades from black to white. The picture on the right gives a random shade to each supervector in order to demonstrate supervector shapes.

[5] P.A. Chou and T. Lookabaugh. Variable dimension vector quantization of linear predictive coefficients of speech. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Adelaide, Australia, April 1994.

[6] 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.

[7] 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*, page 53, San Antonio, Texas, January 1993.

[8] R. F. Rice and J. R. Plaunt. Adaptive variable-length coding for efficient compression of spacecraft television data. *IEEE Transactions on Communications*, 19(12):889–897, December 1971.

[9] L. D. Davisson. Universal noiseless coding. *IEEE Transactions on Information Theory*, 19(6):783–795, November 1973.

[10] R. J. Stevens, A. F. Lehar, and F. H. Preston. Manipulation and presentation of multidimensional image data using the peano scan. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(5):520–533, September 1983.