M.I.T. Media Laboratory Perceptual Computing Group, Technical Report #253
Also to appear in *Proc. ICASSP,* Adelaide, Australia, April 1994

CLUSTER-BASED PROBABILITY MODEL APPLIED TO IMAGE RESTORATION AND COMPRESSION

Kris Popat and Rosalind W. Picard

Media Laboratory, E15-391
Massachusetts Institute of Technology, Cambridge, MA 02139
popat@media.mit.edu, picard@media.mit.edu

ABSTRACT

The performance of a statistical signal processing system is determined in large part by the accuracy of the probabilistic model it employs. Accurate modeling often requires working in several dimensions, but doing so can introduce dimensionality-related difficulties. A recently introduced model circumvents some of these difficulties while maintaining accuracy sufficient to account for much of the high-order, nonlinear statistical interdependence of samples. Properties of this model are reviewed, and its power demonstrated by application to image restoration and compression. Also described is a vector quantization (VQ) scheme which employs the model in entropy coding a Z^N-lattice. The scheme has the advantage over standard VQ of bounding maximum instantaneous errors.

1. INTRODUCTION

Many signal processing techniques employ probabilistic models, explicitly or implicitly. The performance of such techniques depends in large part on the accuracy of the model — that is, on how well it predicts or accounts for signal behavior. In image processing, the samples (pixels) exhibit significant statistical interdependence; this must be exploited if the model is to be accurate. The model should describe the joint behavior of pixels that are statistically related. Unfortunately, as the number of pixels (dimension) increases, two fundamental problems arise: the available training data becomes relatively sparse, and the effective alphabet becomes so large as to be unmanageable.

Kernel estimation [1] alleviates the sparse data problem by making a smoothness assumption about the underlying probability law. Traditional kernel estimates *contain* the training data, which can make the estimates unwieldy. Also, with traditional kernel estimates, the problem of the large alphabet remains. In principle, the joint statistical description provided by the kernel estimate could be used to obtain a conditional description for the individual pixels (using the relationship between joint and conditional probability); however, in practice this computation would be infeasible because of the dynamic range of the quantities involved.

A modeling technique for random vectors which alleviates these problems has recently been proposed, and its application to textured data considered [2]. The model combines kernel estimation with clustering, yielding a semiparametric probability mass function (PMF) estimate which summarizes — rather than contains — the training data. Because the model is cluster based, it is inferable from a limited set of training data, despite the high dimensionality. In addition, the technique is amenable to a computational procedure which circumvents the problem of exponential growth in alphabet size. The next section reviews this model; the remaining sections consider application of the model to image restoration and both lossy and lossless compression.

This work was supported in part by HP Labs and NEC Corp.

2. CLUSTER-BASED PROBABILITY MODEL

The essential details of the model are now summarized. Let $\mathbf{x} = (x_1, \ldots, x_N)$ be a discrete random vector that obeys an unknown probability mass function (PMF) $p(\mathbf{x})$. This PMF is approximated by a function

$$q(\mathbf{x}) = \sum_{m=1}^{M} w_m \prod_{n=1}^{N} f_{m,n}(x_n), \tag{1}$$

where M is a parameter that determines the complexity of the model, the w_m's are positive weights which sum to one, and the $f_{m,n}(x_n)$'s are individual 1-dimensional PMF's. The weights and 1-d PMF's are obtained via cluster analysis, taking the w_m's to be the normalized cluster populations, and fitting discretized separable Gaussians to each cluster to obtain the $f_{m,n}(x_n)$'s.

The advantage of restricting $q(\mathbf{x})$ to be of the form (1) is computational: the vector components can be processed sequentially rather than jointly, with no loss of accuracy. In particular, let $\mathbf{X} = (X_1, \ldots, X_N)$ denote a particular realization of $\mathbf{x}$. Then $q(x_n|X_1, \ldots, X_{n-1})$ can be computed recursively for $1 \leq n \leq N$ as

$$q(x_n|X_1, \ldots, X_{n-1}) = \sum_{m=1}^{M} r_{m,n-1} f_{m,n}(x_n), \tag{2}$$

where

$$r_{m,n} = \begin{cases} w_m & \text{if } n = 0; \\ C_n r_{m,n-1} f_{m,n}(X_n) & \text{if } 1 \leq n \leq N, \end{cases} \tag{3}$$

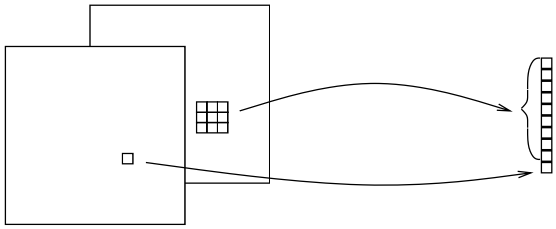

Figure 1: Formation of data vector $\mathbf{x}$ from degraded and original pixels, for use in the training phase. In the restoration phase, the final component of the vector is unknown; its value is estimated by maximizing its PMF conditioned on the degraded pixels.

and where C_n is a normalizing constant chosen such that $\sum_{m=1}^{M} r_{m,n} = 1$ [2].

Regarding quality of the estimate, work on radial basis functions [3] suggests that $q(\mathbf{x})$ can well approximate any $p(\mathbf{x})$ that satisfies certain smoothness constraints. Questions remain about convergence of $q(\mathbf{x})$ to $p(\mathbf{x})$ as the number of clusters increases; also, more research is required to determine precise approximation properties of the model with respect to criteria such as information divergence [4].

3. APPLICATION TO MAXIMUM-LIKELIHOOD IMAGE RESTORATION

Suppose that an image has been degraded and we wish to restore it. It is assumed that a large set of (original, degraded) image pairs is available for the purpose of training. In our example, we consider degradation by additive white noise, but the procedure is well suited to many other types of distortion — even ones which cannot be expressed simply in mathematical form.

First, vectors are formed for each original pixel in the training set, as shown in Figure 1. Next, the necessary quantities in (1) are estimated by cluster analysis, as described in the previous section. For the examples in this paper, the clustering was carried out using the LBG algorithm [5].

For each pixel location in a given degraded image, let $X_1, \ldots, X_{N-1}$ denote the observed values of the neighborhood pixels, and let x_N be the original (unknown) pixel value. To restore the pixel, we take $q(x_N|X_1, \ldots, X_{N-1})$ as the likelihood function and maximize it with respect to x_N. Figure 2 shows experimental results for this type of restoration in the case of additive white Gaussian noise. The technique achieves significant noise reduction while maintaining considerable sharpness. For a set of 5 test images degraded with additive white Gaussian noise (variance = 100), the technique increased SNR by an average of 4.3 dB.

This technique requires only weak assumptions about the statistics of the degradation — stationarity and spatial locality. Therefore, it is expected to perform reasonably in restoration problems where the degradation process is local but highly nonlinear and/or difficult to express mathematically, such as de-halftoning and film grain reduction.

Figure 2: Image restoration example. An original 128×128 8-bit image (top left) is degraded by additive white Gaussian noise with a variance of 100 (top right). Using a 3×3 neighborhood ($N = 9 + 1 = 10$) and $M = 256$, the cluster-based model was trained on a set of 20 (original, degraded) pairs of natural images, which did not include the test image. In the restored image (bottom left) the noise has been reduced at the cost of a small amount of softening and patchiness. For comparison, the result of applying a separable 11-tap Wiener filter is also shown (bottom right) — note that the proposed technique achieves less blur than the filtering approach, with a comparable degree of noise reduction.

This method of restoration is similar in spirit to an interpolative vector quantization technique proposed by Gersho [6]. Both have the potential to "learn" nonlinear statistical relationships from training data and to use those relationships to fill in missing values. However, the techniques differ in one important respect. In the proposed scheme several clusters interact in determining the restored value, rather than the value being determined by a single codebook entry. As a consequence, restored values are not limited to only those appearing explicitly in a codebook. The technique is more than a lookup table; it uses the available information to *synthesize* the missing value.

4. APPLICATION TO LOSSLESS COMPRESSION OF GRAYSCALE IMAGES

Langdon and Rissanen [7] have described an efficient reversible compression scheme for binary images. In their system, each pixel is arithmetically encoded using a PMF that is conditioned on a nearby set of previously encoded pixels, i.e., on a neighborhood of pixels that precede it in

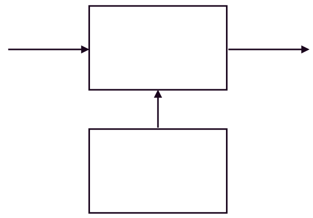

Figure 3: Block diagram of lossless compression system for grayscale images.

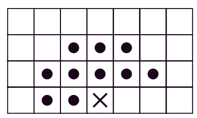

Figure 4: Semicausal conditioning neighborhood for lossless compression. The pixel marked '$\times$' corresponds to x_n, and the pixels marked '$\bullet$' correspond to $X_1, \ldots, X_{N-1}$.

raster order. For binary images and typical neighborhood sizes of $N-1 \approx 10$, it is feasible to estimate the conditional PMF's from occurrence counts, since the number of possible conditioning states (2^{N-1}) remains manageable. In the case of grayscale images however, the number of possible conditioning states ($2^{8(N-1)}$) becomes astronomically large, making count-based probability estimation infeasible.

The proposed model can replace the count-based model, making direct arithmetic coding of grayscale pixels feasible. The system is shown in Figure 3. The pixels are arithmetically encoded in raster order, the PMF used for each pixel being conditioned on a set of previously-encoded pixels so that the decoder can have the same conditioning information. (At the top and left boundaries, unavailable conditioning pixels are arbitrarily set to 128; the resulting local inefficiency has little effect on the overall bit rate.)

The compression system was applied to several 8-bit monochrome images of natural scenes, using a 16-bit K-ary arithmetic coder [8] and the 10-pixel conditioning neighborhood shown in Figure 4. In each case, the cluster-based model was trained on a set of 20 images that did not include the test image. The resulting rates were between approximately 4.5 and 5.5 bits/pixel. This performance range is similar to that reported recently for other lossless compression approaches [9]. We believe that we can improve upon these results. For example, a hierarchical approach (similar to the one we used in texture synthesis [2]) is likely to result in better compression in the larger homogeneous regions of an image.

A popular method of lossless grayscale image compression is to apply the Langdon-Rissanen binary scheme (or a variant) to bit-planes of the image, rather than to the original image [9]. Although a K-ary source can always be reversibly decomposed into $\lceil \log_2 K \rceil$ binary sources, there is an important difference between direct and bit-plane encoding. The difference is in ease of modeling. While a smoothness assumption for the PMF is justified in the original pixel domain, it is definitely not justified in the bit-plane domain. Since smoothness of the PMF is lost, it cannot be exploited by the modeling apparatus. Each conditioning state must then have occurred explicitly in the training data in order for the corresponding conditional PMF to exist. In contrast, by working in the original pixel domain, the cluster-based model can exploit PMF smoothness to effectively *interpolate* the PMF's corresponding to missing conditioning states as necessary.

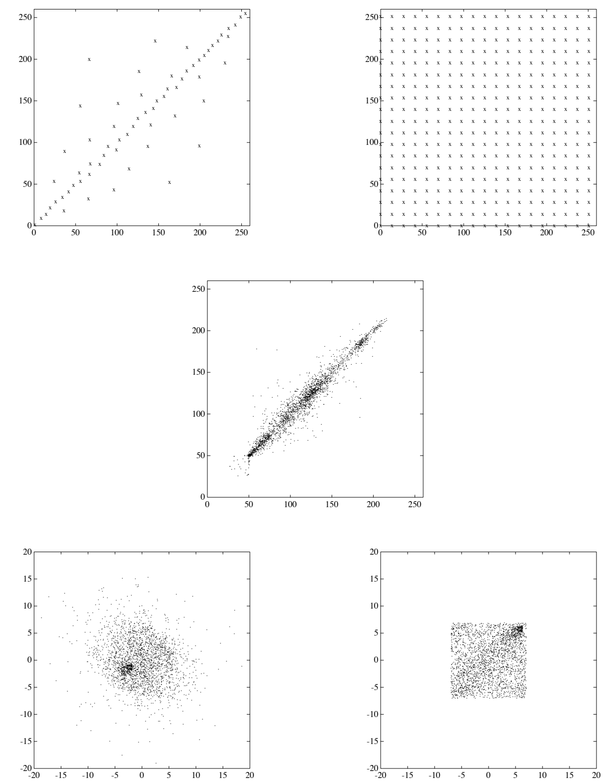

Figure 5: A vector quantization scheme that limits maximum error. Top left: standard VQ codebook with 64 reproductions, designed using the LBG algorithm for a particular 2-D source. Top right: reproductions corresponding to N-fold uniform scalar quantization with joint entropy coding, with the spacing chosen to give the same mean-square error as the 64-point standard VQ. Middle: a set of test points outside the training set, quantized using each of the two codebooks. Bottom: resulting error vectors for standard VQ (left) and N-fold scalar quantization (right).

5. LOSSY COMPRESSION: A VECTOR QUANTIZATION SCHEME WITH LIMITED MAXIMUM ERROR

In recent years, lattice quantization has received much attention as a computationally simpler alternative to full-search vector quantization (VQ). The usual approach is to choose an appropriate lattice, then scale and truncate it in such a way that rate and mean-square-error (MSE) performance is comparable to that of full-search VQ [10, 11, 12, 13].

Lattice VQ also has the potential of limiting maximum errors, provided that an untruncated lattice is used [14]. The idea is illustrated in Figure 5. For untruncated-lattice VQ to perform well for nonuniformly distributed input (the case of interest in image coding), it is essential that the lattice points be efficiently entropy coded. The difficulty lies in the astronomically large alphabet of lattice points that must be handled. Senoo and Girod [14] attack the problem by efficiently entropy encoding only those points that occur frequently in the training data, using a simple fixed-length code for the remaining points. The success of this approach depends on the "frequent set" being small enough to be likely to have been well represented in the training data. In high dimensions (say $N > 5$) and at medium to high rates (> 1 bit/pixel), this may not be true.

If a simple hypercubic (Z^N) lattice is used, then the cluster-based model can be employed along with K-ary arithmetic coding to achieve the desired efficient entropy coding. This amounts to N-fold uniform scalar quantization followed by joint entropy coding of the quantized scalars. While it is true that a Z^N lattice is suboptimal in terms of space-filling, the performance penalty for using it instead of a more efficient lattice is known to be quite small. This has been established theoretically in the case of asymptotically high rate [10, 15], and experimentally in the low- and medium-rate regions for some sources [14]. In the example illustrated in Figure 5, the rate–MSE performance was examined and found to be comparable to that of standard VQ. If this performance can be shown to be comparable in general, then the proposed technique would be a good alternative to standard vector quantization in situations where it is desired that maximum errors be bounded.

As an area of future research, this approach can be applied to lossy image compression in a number of different ways. For example, in any image subband coding system, nonlinear statistical interdependence exists both within and across subbands, despite critical sampling. This interdependence might be handled by the proposed technique.

Pentland and Horowitz [16] describe a hierarchical technique in which statistical interdependence is exploited by conditioning the entropy coding of a VQ output on the corresponding VQ output at a coarser resolution level. Our approach could fit directly into their framework, with the conditioning occurring automatically by assembling vectors both spatially and across resolution levels.

6. SUMMARY

We reviewed some properties of a recently proposed cluster-based probability model, and discussed the model's application to image restoration and compression. The model was found to work well in a maximum-likelihood image restoration system. It performed moderately well in a lossless image compression system; its performance is expected to improve if hierarchical processing is employed. Finally, we described an entropy-coded vector quantization technique based on the model. In contrast to standard vector quantization, the proposed technique bounds maximum error.

7. REFERENCES

[1] Richard A. Tapia and James R. Thompson. *Nonparametric Probability Density Estimation.* The Johns Hopkins University Press, Baltimore, 1978.

[2] Kris Popat and R.W. Picard. A novel cluster-based probability model for texture synthesis, classification, and compression. In *Proc. SPIE Visual Communications '93*, Cambridge, Mass., 1993.

[3] M.J.D. Powell. Radial basis functions for multivariate interpolation: A review. In J.C. Mason and M.G. Cox (eds.), editors, *Algorithms for Approximation.* Clarendon Press, Oxford, U.K., 1987.

[4] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley, 1991.

[5] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Trans. Comm.*, COM-28:84–95, Jan. 1980.

[6] Allen Gersho. Optimal nonlinear interpolative vector quantization. *IEEE Trans. Comm.*, 38(9):1285–1287, Sept. 1990.

[7] Glen Langdon and Jorma Rissanen. Compression of black-white images with arithmetic coding. *IEEE Trans. Comm.*, COM-29:858–867, June 1981.

[8] Kris Popat. Scalar quantization with arithmetic coding. Master's thesis, Dept. of Elec. Eng. and Comp. Science, M.I.T., Cambridge, Mass., 1990.

[9] Majid Rabbani and Paul W. Jones. *Digital image compression techniques.* SPIE Optical Engineering Press, Bellingham, Washington, 1991.

[10] J.H. Conway and N.J.A. Sloane. *Sphere packings, lattices, and groups.* Springer-Verlag, 1988.

[11] J.H. Conway and N.J.A. Sloane. Fast quantizing and decoding algorithms for lattice quantizers and codes. *IEEE Trans. Inform. Theory*, IT-28:227–232, March 1982.

[12] D.G. Jeong and J.D. Gibson. Lattice vector quantization for image coding. In *ICASSP-89: 1989 International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, UK, May 1989. IEEE.

[13] M. Antonini, M. Barlaud, and P. Mathieu. Image coding using lattice vector quantization of wavelet coefficients. In *ICASSP-91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, April 1991. IEEE.

[14] Takanori Senoo and Bernd Girod. Vector quantization for entropy coding of image subbands. *IEEE Trans. Image Proc.*, 1:526–533, Oct. 1992.

[15] Tom D. Lookabaugh and Robert M. Gray. High-resolution quantization theory and the vector quantizer advantage. *IEEE Trans. Inform. Theory*, IT-35:1020–1033, 1989.

[16] Alex Pentland and Bradley Horowitz. A practical approach to fractal-based image compression. In *Proc. IEEE Data Comp. Conf.*, Utah, 1991.