In International Joint Conference on Neural Networks, IEEE, 1999.

Optimal Dimension Reduction and Transform Coding with Mixture Principal Components

Cynthia Archer and Todd K. Leen
archer@cse.ogi.edu *tleen@cse.ogi.edu*

Department of Computer Science and Engineering
Oregon Graduate Institute of Science & Technology
20000 N.W. Walker Rd, Beaverton, OR 97006-1000

Abstract

This paper addresses the problem of resource allocation in local linear models for non-linear principal component analysis (PCA). In the local PCA model, the data space is partitioned into regions and PCA is performed in each region. Our primary result is that the advantage of these models over conventional PCA has been significantly underestimated in previous work.

We apply local PCA models to the problems of image dimension reduction and transform coding. Our results show that by allocating representation or coding resources to the different image regions, instead of using a fixed arbitrary dimension everywhere, substantial increases in dimension reduced or compressed image quality can be achieved.

Introduction

This paper addresses resource allocation in local linear models for non-linear principal component analysis (PCA). These models, previously discussed in [1, 2, 3], are an alternative to non-linear PCA models such as the five-layer, non-linear autoassociators discussed in [4, 5]. The latter construct smooth, curved manifolds that are close to the data. In the local PCA model, the data space is partitioned into regions, and PCA is performed in each region. Geometrically, such models approximate the data manifold by a set of local PCA hyperplanes. When used for dimension reduction and for transform coding, local PCA models exhibit a clear performance advantage over simple PCA. They are faster to fit than five-layer, nonlinear autoassociators, and often outperform them.

Despite their success, previous studies under-utilize the potential of these models. For dimension reduction, previous authors choose a *global* target dimension, and hence neglect the variability in intrinsic data dimension from region to region in the data space. Here we construct a Langrangian-based algorithm that allows the model's dimension to be adjusted locally in order to decrease distortion, while the *average* dimension is constrained to a particular value.

In applications to transform coding, previous authors [3, 6] choose a dimension for the transform coefficients, project the data to this reduced dimension, and then distribute coding bits among the retained coordinates. This approach constrains the bit distribution process to find the best allocation of coding bits in a subset of the full search space. If the chosen dimension is small, that subset will not include the distribution of coding bits that minimizes compression distortion. In order to optimize transform coding, we apply Riskin's bit allocation scheme [7]. This is a simple iterative greedy search that distributes coding bits among the coefficients so as to minimize compression distortion subject to a constraint on the overall bit rate. For this greedy search method to be effective, the rate-distortion functions for the coefficient quantizers must be convex, which means that the quantizers must be well-tuned to the input data. We use a library of quantizers optimized for a set of probability densities that reflect the coefficient value distributions seen in our image data.

Local PCA algorithms cluster the input data into regions and perform PCA on the data that falls within each region. For the work described here, the clustering is done with a standard vector quantizer using a mean square error (MSE) distortion function (Euclidean distance clustering [1]).

The experimental results we describe here were obtained on a database of images consisting of 50 frames each from two video sequences of city street intersections [1] The frames from the sequences measure 512×512 and 512×480 pixels respectively. The training set consists of eight frames from the first half of each sequence. Frames from the second half of each sequence were used for testing.

[1] The images in our database are available for research use from the Institute für Algorithmen und Kognitive Systeme, Universität Karlsruhe website, i21www.ira.uka.de/image_sequences.

Dimension Reduction

While transform coding depends on both the adaptive transform and the coding of the transform coefficients, dimension reduction dispenses with coding. It is therefore more purely a window into the performance of the transform. Dimension reduction can be used to preprocess data for other signal processing tasks, such as compression, classification and detection, or density estimation. It is vital for visualization of high-dimensional data.

Previous work on dimension reduction operates almost exclusively with reduction to a *fixed, globally-defined* dimension. The dimension may be chosen by a fidelity requirement that places an upper bound on the allowed average distortion. Alternatively, several authors [8, 9] have applied minimum description length (MDL) criteria to PCA for estimating signal dimension. We argue that there is no compelling argument to require a single, global dimension[2] In fact, a simple exploration of the local correlation eigenstructure of images shows that different regions of the data space have different dimension. Figure 1 gives an example showing differences in the number of relatively high variance eigendirections in different image regions.

Here we propose a *resource allocation* approach to local dimension assignment. Our algorithm assigns dimension m_i to the i^{th} region of a local PCA model so as to *minimize the expected distortion*, subject to the constraint that the *average dimensionality is fixed*. The algorithm can potentially assign a different dimension to each region of the image signal space.

Our algorithm is motivated by a Lagrangian formulation in which one minimizes the expected distortion, subject to a constraint on the average dimension used. The Lagrangian is

$$L = \sum_i p_i\, d(m_i) + \Lambda\,(m_0 - \sum_i p_i\, m_i), \qquad (1)$$

where p_i is the a priori probability of picking data from the i^{th} region, $d(m_i)$ is the expected distortion for data in the i^{th} region when reduced to an m_i dimensional representation, m_0 is the desired average dimension, and Λ is an undetermined multiplier.

In general, realizing the theoretical minimization of this Lagrangian is not possible. However the following simple heuristic does allow one to reach an empirical minimum at approximately the desired average dimension:

1. For all regions i, set $m_i = 0$.
2. Find the largest λ_{i,m_i+1}.
3. Allocate 1 additional dimension to the corresponding region, i.e. $m_i = m_i + 1$.
4. Calculate the average dimension $\bar{m} = \sum_i p_i\, m_i$. If the average dimension reaches m_0, then stop. Otherwise loop to (2).

The variable local dimension allowed by this algorithm, compared to a fixed global dimension, results in substantial increases in the Signal-to-noise ratio (SNR) of dimension reduced signals. Figure 2 shows results for reduction of 64-dimensional blocks of image pixel values to both eight average and fixed dimensions. When using 32 PCA regions, the local dimension assigned by our

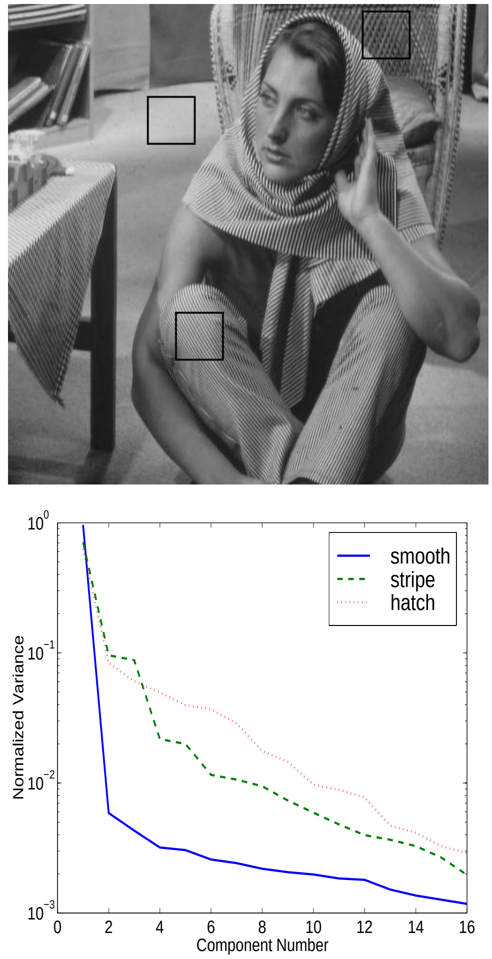

Figure 1: The three regions in the example image delineated by squares have vastly different correlation eigenspectra. Assigning dimension by a fidelity requirement that places an upper bound on distortion would find a different dimensionality for each region.

[2]Admittedly, this might be more convenient for data visualization.

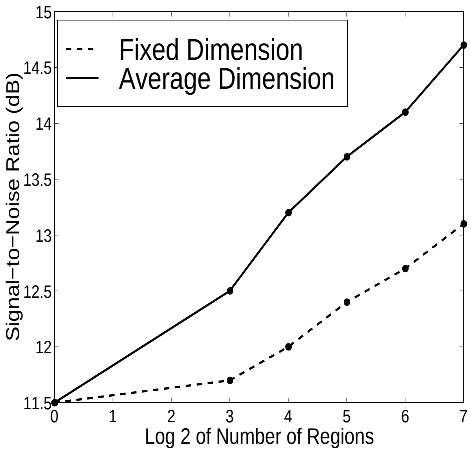

Figure 2: Dimension reduction of image blocks from 64 to 8 dimensions with fixed local dimension (dashed curve) and variable local dimension (solid curve).

algorithm varies between 3 and 23 among the different regions. The corresponding image SNR improvement is 1.3 dB relative to assigning all regions dimension eight.

Coding and Compression

Optimal coding of a transformed (e.g. by global or local PCA) signal requires accurate quantizer design and smart allocation of coding bits between the quantizers for different transform coefficients. Our primary contributions in this area are twofold. First, we have constructed a library of density models that accurately model the distributions of local PCA transform coefficient values. This library allows one to *quickly* build nearly optimal quantizers. Second, we have developed a procedure that uses these rapidly-fit, accurate quantizers to find the optimal bit allocation between transform coefficients[3]

Quantizer Models

Our density models must be accurate, otherwise the bit allocation procedure will size the quantizers incorrectly. The typical density assumptions of a uniform or Gaussian density for the leading coefficient and a Laplace density for all the other coefficients are often inaccurate [10, 11]. These assumptions are especially inaccurate for local PCA transform coefficients. Therefore, unlike typical transform coding systems [12, 13, 14, 3], we do not select a priori which densities will be used to model which coefficients. Instead, we find the density from our library that best reflects the distribution of coefficient values along each local eigendirection.

[3]Bit allocation is a *resource* allocation problem, much like the local dimension assignment discussed in the previous section.

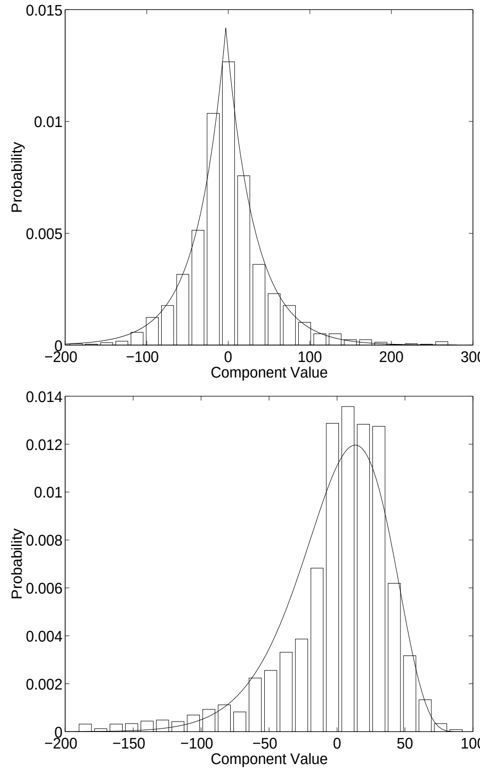

Figure 3: Coefficient value histograms and model densities: Laplace (top) and Pearson (bottom)

To determine a reasonable family of probability densities, we examined histograms of transform coefficient values from several images. Coefficient distributions are typically unimodal, but not always symmetric. We use three generalized Gaussian densities

$$p(x) = \frac{pK(p)}{2\beta\Gamma(1/p)} \exp(-K(p)\frac{|x-\alpha|}{\beta}^p), \qquad (2)$$

with $p = 1/2$, 1, and 2 and $K(p) = \frac{\Gamma(3/p)}{\Gamma(1/p)}^{1/2}$, to fit the symmetric distributions. Pearson densities

$$p(x) = \frac{1}{\beta\Gamma(p)} \frac{(x-\alpha)}{\beta}^{p-1} \exp(-\frac{x-\alpha}{\beta}), \qquad (3)$$

with $p = 2$, 3, and 7, are used to model asymmetric densities. We also include the uniform distribution. Figure 3 shows two representative coefficient histograms overlaid with fitted densities.

In order to identify which model distribution to apply to a particular transform coefficient, we fit *all* the candidates from our library to the coefficient values by maximum data likelihood. The fitting amounts to estimating shift and width scaling parameters of each density model. From among the candidates, we then choose the fitted density with the highest data likelihood.

Our library of model densities defines a library of optimal quantizers. We fit Lloyd [15] quantizers to each of the model densities for a range of quantizer bit-rates. The quantizer reproduction values are then automatically scaled and shifted when the densities are fit to the data.

Bit Allocation

Having established the above quantizer models, we apply Risken's bit allocation algorithm [7] to select the size for each coefficient quantizer. This method assigns (non-negative) integer numbers of quantization bits among the transform coefficients so as to minimize the average distortion subject to a constraint on the total number of coding bits. Let b_{ij} denote the number of bits allocated to the j^{th} transform coefficient in the i^{th} region. Two constraints are possible: fix the *average* bit-rate $\sum_{i=1}^{M} p_i \left(\sum_{j=1}^{d} b_{ij}\right)$, or fix the bit-rate in each region $\sum_{j=1}^{d} b_{ij}$. The former constraint leads to a variable rate code; different numbers of bits are assigned to each region, hence the number of bits used to encode a block will vary. The later constraint leads to a fixed rate code; each block is encoded with the same number of bits. We have explored both fixed and variable rate codes, although for lack of space, we will only give results here for fixed rate coding.

Coding Results

Previous applications of local PCA to image compression first reduce the dimension of the transformed image blocks to an arbitrary, fixed target dimension and then quantize the retained coefficients [6, 3]. Tipping and Bishop [6] reduce an image block to four transform coefficients and than uniformly quantize these coefficients to eight bits each, which yields a compressed bit rate of 0.5 bits per pixel (bpp). Dony and Haykin [3] reduce an image block to two, four, or eight coefficients and then allocate bits, using estimates of quantization distortion, among the retained coefficients. These methods constrain the bit allocation process to distribute coding resources to only a subset of the available transform coefficients. Consequently, image information essential to maintaining high compressed image quality may be discarded.

Instead of arbitrary target dimension selection, our compression method uses bit allocation to determine the number of retained coefficients and quantizer sizes that give the best image quality for a particular level of compression. The plots in Figure 4 illustrate how compressed image quality is improved when all of the transform coefficients are considered for coding. Signal-

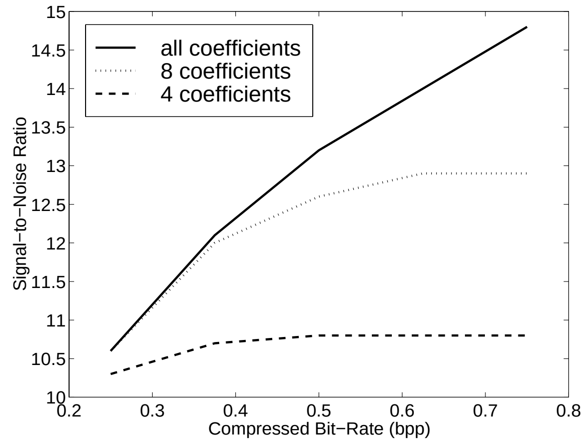

Figure 4: Quality of constrained vs. unconstrained bit allocation for compression using 128 region partition.

to-noise ratios are shown for a test image that has been compressed using local (128 region) PCA transforms, when bit allocation was performed (from bottom to top) among four coefficients, among eight coefficients, and among all transform coefficients. For compression to 0.5 bpp, the bit allocation process retains a minimum of ten to a maximum of fourteen coefficients per region. The 0.5 bpp compressed image SNR is 2.4 dB higher[4] when unrestricted bit allocation is used instead of restricting coding to only four coefficients.

The enhanced image quality that results when coder design is not restricted to a subset of the transform coefficients can be seen in the reconstructed images. Figure 5 shows three versions of an enlarged section from a test image: original image (left), the image compressed to 0.5 bpp using four 8-bit coefficients (center), and the image compressed to 0.5 bpp using bit allocation to select dimension and size quantizers (right). Blocking effects and gray-level quantization are more pronounced in the four coefficient image. In addition, the fine details in the properly compressed image are less blurred than in the four coefficient image. Both our SNR results and visual examination of the compressed images shows that proper coding enhances the compression performance of local PCA-based transform coding.

Discussion

We have shown that the application of local PCA algorithms to dimension reduction are significantly enhanced by careful resource allocation. The fidelity gains achieved are comparable to the advantage of us-

[4] Hypothesis tests performed on the mean MSE of twenty compressed test images indicate that an SNR change on the order of 0.1 dB is meaningful at a 0.005 significance level.

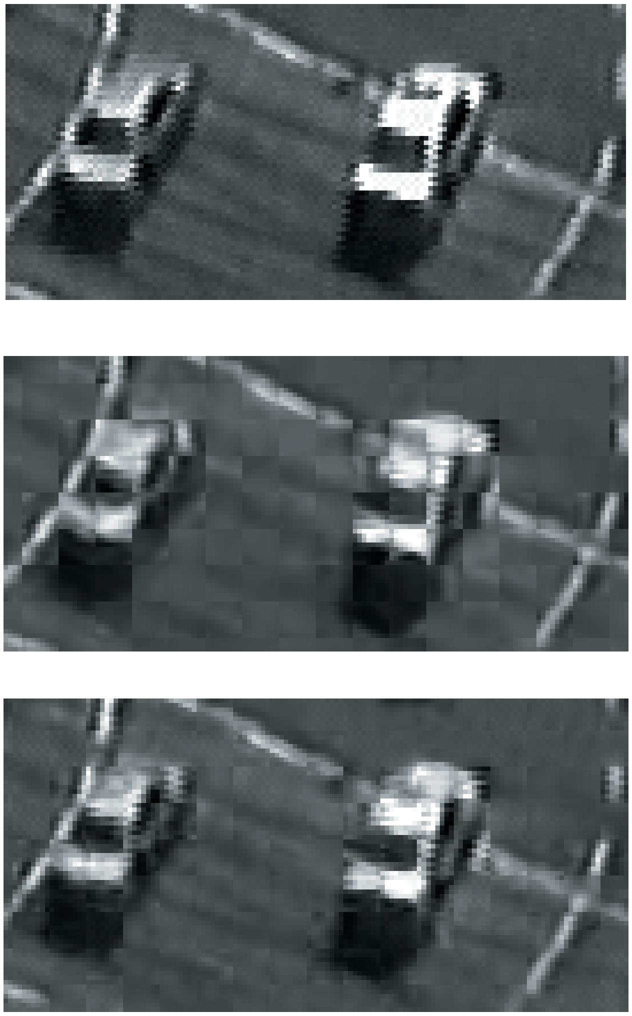

Figure 5: Example 0.5 bpp compressed image segments for 128 region local PCA. Top to Bottom: Original, Four Coefficients, All Coefficients

ing locally-tuned PCA transforms over standard, global PCA transforms. For example, look back at the image dimension reduction examples in figure 2. For reduction to an eight dimensional representation everywhere, the SNR increases about 1.5 dB in going from the global PCA (1 region) to the 128 region PCA. However, for the 128 region PCA, changing to a local dimension allocation increases the SNR about another 1.5 dB.

For applications of local PCA to transform coding, we have shown that careful allocation of coding bits and accurate quantizer design significantly improves compressed image quality. To see this, look back at the coding results in figure 4, where the image is compressed from 8 bits to 0.75 bits per pixel and smaller. The typical choice of coding only four coefficients, substantially degrades compressed image quality, especially at useful compression levels. For instance, at 0.75 bpp, using proper bit allocation improves the compressed image SNR by 3.0 dB over retaining and coding only four coefficients.

Acknowledgements – This work was supported by the NSF through grant ECS-9704094.

References

[1] Nanda Kambhatla and Todd K. Leen. Fast non-linear dimension reduction. In Cowan, Tesauro, and Alspector, editors, *Advances in Neural Information Processing Systems 6*, pages 152,159. Morgan Kauffmann, Feb 1994.

[2] Nandakishore Kambhatla and Todd K. Leen. Optimal dimension reduction by local PCA. *Neural Computation*, 9:1493–1516, 1997.

[3] Robert D. Dony and Simon Haykin. Optimally adaptive transform coding. *IEEE Transactions on Image Processing*, 4:1358–1370, 1995.

[4] Mark A. Kramer. Nonlinear prinipal component analysis using autoassociative neural networks. *AIChE journal*, 37(2):233–243, February 1991.

[5] David DeMers and Garrison Cottrell. Non-linear dimensionality reduction. In Giles, Hanson, and Cowan, editors, *Advances in Neural Information Processing Systems, vol. 5*, San Mateo, CA, 1993. Morgan Kaufmann.

[6] M. Tipping and C. Bishop. Mixture of probabilistic principal component analyzers. *Neural Computation*, 11:443–483, 1999.

[7] Eve A. Riskin. Optimal bit allocation via the generalized BFOS algorithm. *IEEE Transactions on information Theory*, 37:400–402, 1991.

[8] Mati Wax and Thomas Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-33(2):387–392, 1985.

[9] T. Hediger, A. Passamante, and M.E. Farrell. Characterizing attractors using local intrinsic dimensions calculated by singular-value decomposition and information-theoretic criteria. *Physical Review*, A 41:5325–5332, 1990.

[10] R. Reiniger and J. Gibson. Distributions of the two-dimension dct coefficient for images. *IEEE Transaction on Communications*, COM-31(6):835–839, 1983.

[11] R.J. Clarke. *Transform Coding of Images*. Academic Press, 1985.

[12] W. Chen and C. Smith. Adaptive coding of monochrome and color images. *IEEE Transactions on Communications*, 25(11):1285–1292, 1977.

[13] W. Chen and W. Pratt. Scene adaptive coder. *IEEE Transactions on Communications*, COM-32(3):225–232, 1984.

[14] A. Hung and T. Meng. Optimal quantizer step sizes for transform coders. In *Proceedings of ICASSP*, volume 4, 1991.

[15] S. Lloyd. Least square optimization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.