FAST APPROXIMATE FACTOR ANALYSIS

Mladen Victor Wickerhauser

Department of Mathematics, Campus Box 1146
Washington University in St. Louis, Missouri 63130
Email: victor@jezebel.wustl.edu

ABSTRACT

The principal orthogonal factor analysis or Karhunen–Loève algorithm may be sped up by a low-complexity preprocessing step. A fast transform is selected from a large library of wavelet-like orthonormal bases, so as to maximize transform coding gain for an ensemble of vectors. Only the top few coefficients in the new basis, in order of variance across the ensemble, are then decorrelated by diagonalizing the autocovariance matrix. The method has computational complexity $O(d^2 \log d + d'^3)$ and $O(d \log d + d'^2)$ respectively for training and classifying a d-dimensional system, where $d' \ll d$. One application is described, the reduction of an ensemble of 16,384-pixel face images to a 10-parameter space using a desktop computer, retaining 90% of the variance of the ensemble.

1. INTRODUCTION

Let $X = \{X_n : n = 1, \ldots, N\} \in \mathbf{R}^d$ be an ensemble of vectors. The Karhunen–Loève basis for the ensemble consists of the eigenvectors of the symmetric positive definite autocovariance matrix $M = E(X \otimes X)$, or

$$M_{ij} = \frac{1}{N} \sum_{n=1}^{N} X_n(i) X_n(j). \tag{1}$$

Since the autocovariance matrix for the Karhunen–Loève eigenvectors is diagonal, it follows that coefficients with respect to the Karhunen–Loève basis are uncorrelated random variables. Taking another perspective, they achieve the maximum linear transform coding gain, or equivalently the minimum entropy, of any linear transform code used to transmit X. Unfortunately, finding these eigenvectors requires diagonalizing a matrix of order d, which has complexity $O(d^3)$. In the cases of interest, d is very large and X spans $\mathbf{R}^d$, implying $N \geq d$. For large d, of the order of 10^4, such a complex problem is intractable with desktop computing equipment. For d of the order of 10^6, such as for images scanned at 1024×1024 pixels, the Karhunen–Loève method is out of reach of even the most powerful supercomputers.

The fast approximate factor analysis algorithm searches a library of orthonormal bases for the one closest to the Karhunen–Loève basis. Distance is measured in the natural way: we compute the transform coding gain of each basis in the library, to find the basis which maximizes it.

Consider the picture below, a schematic of all the orthonormal bases of $\mathbf{R}^d$. The points marked by "x" represent bases in some library of fast transform bases. The point marked "o" represents the optimal, or Karhunen–Loève basis, for a given ensemble of vectors. The point marked "xx" represents the fast expansion which has the greatest transform coding gain:

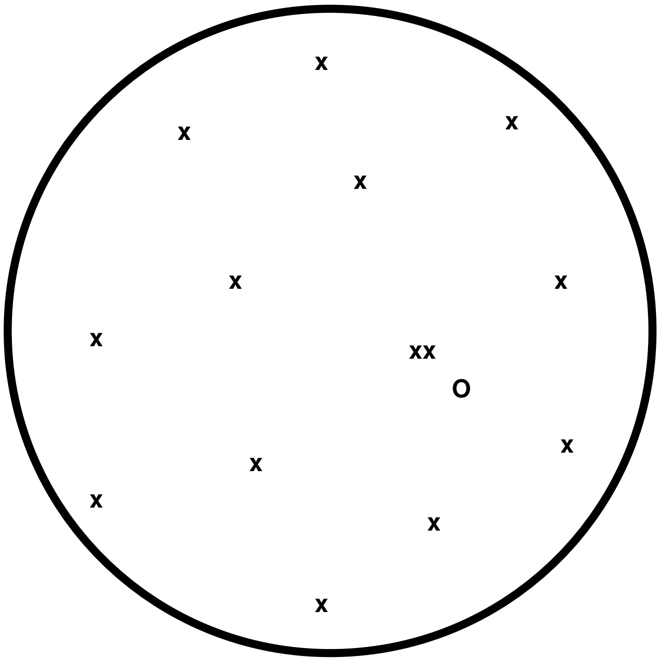

Figure 1. Orthonormal bases for $\mathbf{R}^d$.
x = fast transform bases; o = Karhunen–Loève basis; xx = best fast basis.

Using wavelet packets [CMW], we can build a library of more than 2^d bases for $\mathbf{R}^d$, yet still search it for the maximum in $O(d \log d)$ operations.

2. NOTATION

Write $\overline{X} = E(X) = \frac{1}{N}\sum_{n=1}^{N} X_n$, and let $\sigma(X) \in \mathbf{R}^d$ be the vector of variances of the coefficients of X. Namely,

$$\sigma(X)(i) = \frac{1}{N}\sum_{n=1}^{N}[X_n(i) - \overline{X}(i)]^2. \tag{2}$$

We may assume without loss that $\overline{X} = 0$. Write $\mathrm{Var}(X)$ for the sum of the coefficients of $\sigma(X)$, which is the total variance of the ensemble X.

Let $U : \mathbf{R}^d \to \mathbf{R}^d$ be orthogonal and write $Y = UX$ for the map $Y_n = UX_n$. Since U is linear, $\overline{Y} = \overline{UX} = U\overline{X} = 0$, and since U is orthogonal, $\mathrm{Var}(X) = \mathrm{Var}(Y)$. Define the *transform coding gain* as in [JN] by the formula $G_{TC}(U) = \mathrm{Var}(UX)/\exp H(UX)$, where

$$H(X) = \frac{1}{d}\sum_{i=0}^{d-1} \log \sigma(X)(i). \tag{3}$$

$G_{TC}(UX)$ is maximized when $H(UX)$ is minimized, and H is the entropy of the direct sum of d independent Gaussian random variables with variances $\sigma(X)(i)$, $i = 0, \ldots, d-1$. The Karhunen–Loève transform is a global minimum for H, and we will say that the best approximation to the Karhunen–Loève transformation from a library $\mathcal{U}$ of orthogonal transformations is the minimum of $H(UX)$ with the constraint $U \in \mathcal{U}$.

3. ALGORITHM

Notice that H is an information cost function in the sense of [CW]. We may create a large library of orthogonal bases by recursive quadrature mirror filter convolution-decimation, and use the best-basis search algorithm with H to find the best approximation to the Karhunen–Loève basis.

Let S and D be low-pass and high-pass quadrature mirror filters (QMF's), respectively determined by the coefficients $h(n)$ and $g(n)$ as follows:

$$\begin{aligned} s_i = Sx(i) &= \sum_n h(n)x(2i+n) \\ d_i = Dx(i) &= \sum_n g(n)x(2i+n) \end{aligned} \tag{4}$$

These are perfect reconstruction filters if $S^*D = 0$, $SS^* = DD^* = I$, and $S^*S + D^*D = I$, where S^* denotes the adjoint of S, etc. The simplest nontrivial example is given by the Haar QMF's $h(0) = h(1) = g(0) = 1/\sqrt{2}$, $g(1) = -1/\sqrt{2}$. Of course these may be recursively applied, giving for instance $sd_i = SDx(i)$.

In the case $d = 8$ the complete 1-dimensional wavelet packet expansion may be organized as follows:

$X_i(0)$	$X_i(1)$	$X_i(2)$	$X_i(3)$	$X_i(4)$	$X_i(5)$	$X_i(6)$	$X_i(7)$
s_0	s_1	s_2	s_3	d_0	d_1	d_2	d_3
ss_0	ss_1	ds_0	ds_1	sd_0	sd_1	dd_0	dd_1
sss_0	dss_0	sds_0	dds_0	ssd_0	dsd_0	sdd_0	ddd_0

Figure 2. Complete wavelet packet expansion of X_i.

Then the sums of the squares are accumulated in an array of variances:

$\sum_i X_i^2(0)$	$\sum_i X_i^2(1)$	$\sum_i X_i^2(2)$	$\sum_i X_i^2(3)$	$\sum_i X_i^2(4)$	$\sum_i X_i^2(5)$	$\sum_i X_i^2(6)$	$\sum_i X_i^2(7)$
$\sum_i s_0^2$	$\sum_i s_1^2$	$\sum_i s_2^2$	$\sum_i s_3^2$	$\sum_i d_0^2$	$\sum_i d_1^2$	$\sum_i d_2^2$	$\sum_i d_3^2$
$\sum_i ss_0^2$	$\sum_i ss_1^2$	$\sum_i ds_0^2$	$\sum_i ds_1^2$	$\sum_i sd_0^2$	$\sum_i sd_1^2$	$\sum_i dd_0^2$	$\sum_i dd_1^2$
$\sum_i sss_0^2$	$\sum_i dss_0^2$	$\sum_i sds_0^2$	$\sum_i dds_0^2$	$\sum_i ssd_0^2$	$\sum_i dsd_0^2$	$\sum_i sdd_0^2$	$\sum_i ddd_0^2$

Figure 3. Variances of wavelet packet coordinates of X.

This tree of variances may be searched for the orthogonal basis which minimizes H. Let U be this basis, which will be called the *joint best basis for the ensemble X, in the library.* We write $\{U_i \in \mathbf{R}^d : i = 1, \ldots, d\}$ for the rows of U. We may suppose that these rows are numbered so that $\sigma(UX)$ is in decreasing order. Then we fix $\epsilon > 0$ and let d' be the smallest integer such that $\sum_{n=1}^{d'} \sigma(UX)(n) \geq (1-\epsilon)\mathrm{Var}(X)$. Then the projection of X onto the span of $U' = \{U_1 \ldots U_{d'}\}$ contains $1 - \epsilon$ of the variance.

The search for U involves computing H for each subblock of the variance tree, then performing a recursive binary comparison between that quantity and the sum $H(Y_1) + H(Y_2)$, where Y_1 and Y_2 are the optimal bases of the subblock's two immediate descendents. The better of the values is passed up the tree for the next comparison, and the winning basis is marked for later extraction. The search terminates after the root, or original signal, is compared with the the best bases of its immediate D and S descendents.

Already these d' vectors U' are a good basis for the ensemble X, but they may be further decorrelated by Karhunen–Loève factor analysis. Suppose that U'_i, for $1 \leq i \leq d'$, are the row vectors of U'. Then the autocovariance matrix for this new collection is given by

$$M'_{ij} = \frac{1}{N}\sum_{n=1}^{N} U'_i X_n U'_j X_n. \tag{5}$$

Thus M' is order d', and can be diagonalized in $O(d'^3)$ operations. Let K' be the matrix of eigenvectors of M'. Then we may call the composition $K'U'$ (associated to ϵ) the *approximate Karhunen–Loève transform with relative variance error* ϵ.

4. COMPLEXITY

The algorithm is fast because we expect that even for small ϵ we will obtain $d' \ll d$. To count operations in the worst case, assume:

N random vectors

d-dimensional parameter space $\mathbf{R}^d$

Full-rank autocovariance matrix ($N > d$)

4.1 Training complexity.

1. Expanding N vectors $X_n \in \mathbf{R}^d$ into wavelet packet coefficients: $O(Nd\log d)$;
2. Summing squares into the variance tree: $O(d\log d)$;
3. Searching the variance tree for a best basis: $O(d)$;
4. Sorting the best basis vectors into decreasing order of importance: $O(d\log d)$;
5. Diagonalizing the autocovariance matrix of the top d' best-basis vectors: $O(d'^3)$.

Hence the total complexity of constructing the approximate Karhunen–Loève transform $K'U'$, is $O(Nd\log d + d'^3)$. This compares favorably with the complexity $O(Nd^2 + d^3)$ of the full Karhunen–Loève expansion, since we expect $d' \ll d$.

Depending on circumstances, the last step $U' \mapsto K'U'$ may not be necessary, since a large reduction in the number of parameters is already achieved by transforming into the orthonormal basis determined by U'. This reduces the complexity to $O(Nd\log d)$, with the penalty being less decorrelation of the factors.

4.2 Classifier complexity.

1. Expanding 1 vector into wavelet packet coefficients: $O(d\log d)$.
2. Projecting onto the eigenvectors of the $d' \times d'$ autocovariance matrix: $O(d'^2)$

Since $d' \ll d$, this estimate compares favorably with the complexity of the full Karhunen–Loève expansion, which is $O(d^2)$. Further savings are possible, notably because only a small fraction $d'' \ll d'$ of the Karhunen–Loève eigenvectors are needed to capture almost all of the original ensemble variance. Hence the total classification complexity is bounded by $O(d \log d + d''d')$.

5. APPLICATION TO THE MUG'S GALLERY PROBLEM

The ensemble of vectors contained 143 digitized 128x128x8bit pictures of Brown University students. These were already normalized with the pupils impaled on two fixed points near the center of the image. We first transformed the data to floating point numbers, computed average values for the pixels, and subtracted the average from each pixel to obtain "caricatures," or deviations from the average. Figure 4 is one of these caricatures:

Figure 4. Face, minus the average face, yields a caricature.

Figure 5 shows how the variance accumulates pixel by pixel, with the pixels sorted into decreasing order of variance.

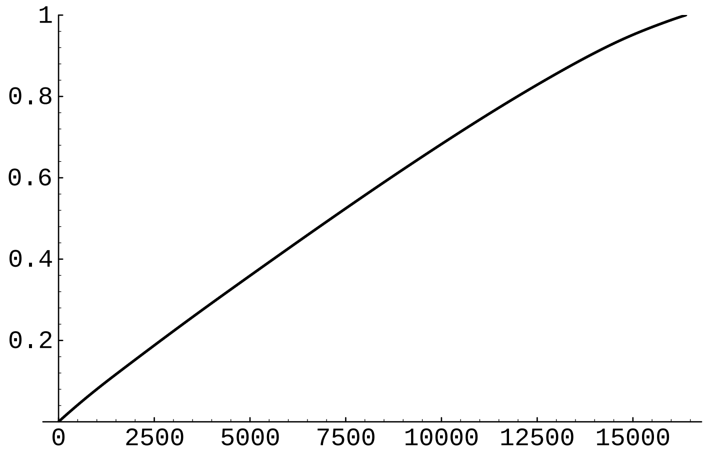

Figure 5. Accumulation of variance in the original basis.

Each caricature was treated as a picture and expanded into 2 dimensional wavelet packets as described in [W]. The QMF's used were the so called "C 12" filters, whose coefficients are listed below:

$$\begin{aligned} h = \{&0.0163873364631797850, -0.0414649367819664850, -0.0673725547222998740, \\ &0.386110066823092900, 0.812723635449606130, 0.417005184423777600, \\ &-0.0764885990782645940, -0.0594344186464712400, 0.0236801719468767500, \\ &0.00561143481936598850, -0.00182320887091009920, -0.000720549445368115120\}; \end{aligned} \tag{5}$$

$$\begin{aligned} g = \{&-0.000720549445368115120, 0.00182320887091009920, 0.00561143481936598850, \\ &-0.0236801719468767500, -0.0594344186464712400, 0.0764885990782645940, \\ &0.417005184423777600, -0.812723635449606130, 0.386110066823092900, \\ &0.0673725547222998740, -0.0414649367819664850, -0.0163873364631797850\}; \end{aligned} \tag{6}$$

These are perfect reconstruction filters, as may be easily verified. They produce wavelets which are regular and very close to symmetric, making them suitable for the analysis of smooth isotropic pictures.

Two-dimensional data is filtered by separating dimensions, as depicted in the Figure 6. The resulting quadtree depicts a nominal decomposition of 2-dimensional frequency space into finer and finer subsquares, which also correspond to larger and larger 2-dimensional wavelet packets.

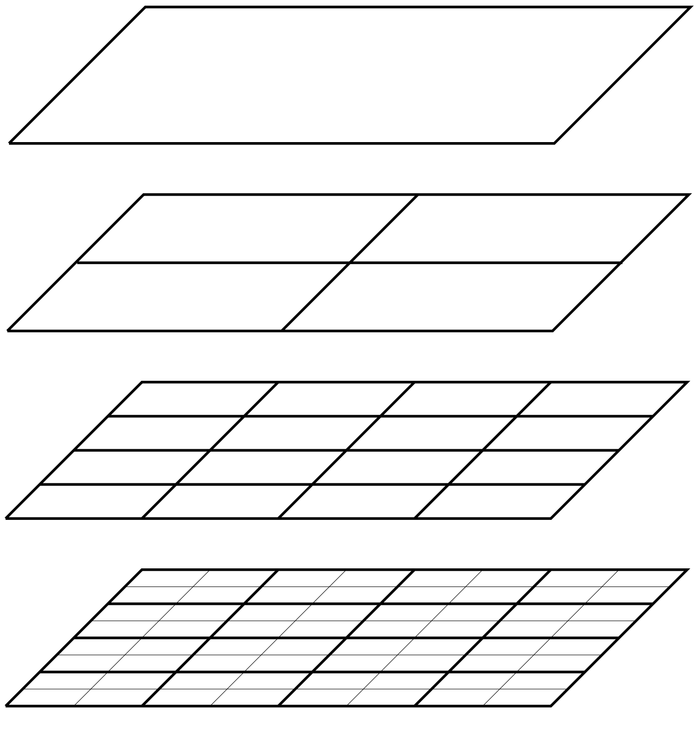

Figure 6. Separable 2-dimensional wavelet packet analysis.

The squares of the amplitudes were summed into a tree of variances, which was searched via the best-basis search procedure. Call this most-concentrated basis the *joint best basis* for the ensemble. In the joint best basis, 400 coordinates (of 16,384) contained more than 90% of the variance of the ensemble.

Figure 7 shows the total variance on the first d' coordinates in the joint best basis, sorted in decreasing order, as a fraction of the total variance of the ensemble, for $1 \leq d' \leq 1024$. Using 1024 parameters captures more than 95% of the ensemble variance, but requires somewhat more computer power than is readily available on a desktop. A 400 parameter system can be analyzed on a desktop workstation in minutes.

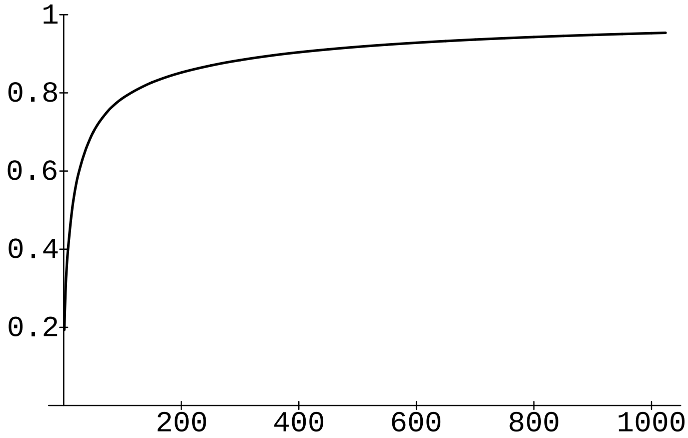

Figure 7. Accumulation of variance in the joint best basis.

The top 400 coordinates were recomputed for each caricature and their autocovariance matrix over the ensemble was diagonalized using the LINPACK singular value decomposition routine.

Figure 8 shows the total variance on the first d' coordinates in the Karhunen–Loève basis, sorted in decreasing order, as a fraction of the total variance of the 400 joint best basis coefficients, for $1 \leq d' \leq 143$. The Karhunen–Loève post-processing for this small ensemble concentrates 98% of the retained variance from the top 400 joint best-basis parameters into 10 coefficients.

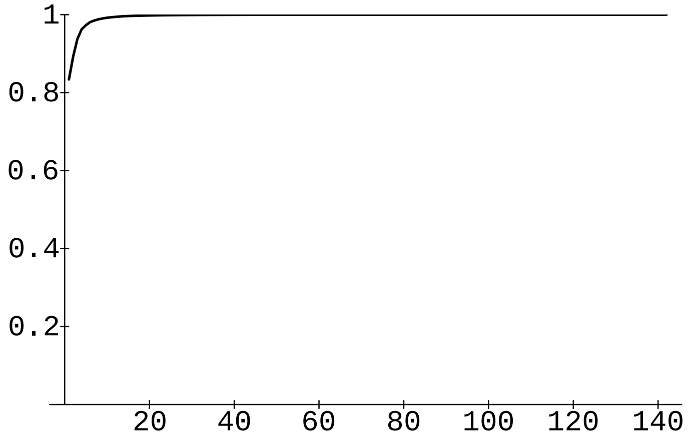

Figure 8. Accumulation of variance in the approximate Karhunen–Loève basis.

Figure 9 shows the top 6 Karhunen–Loève "eigenfaces" with respect to the top 400 joint best-basis parameters. These have been normalized to fill the dynamic range of the printing device.

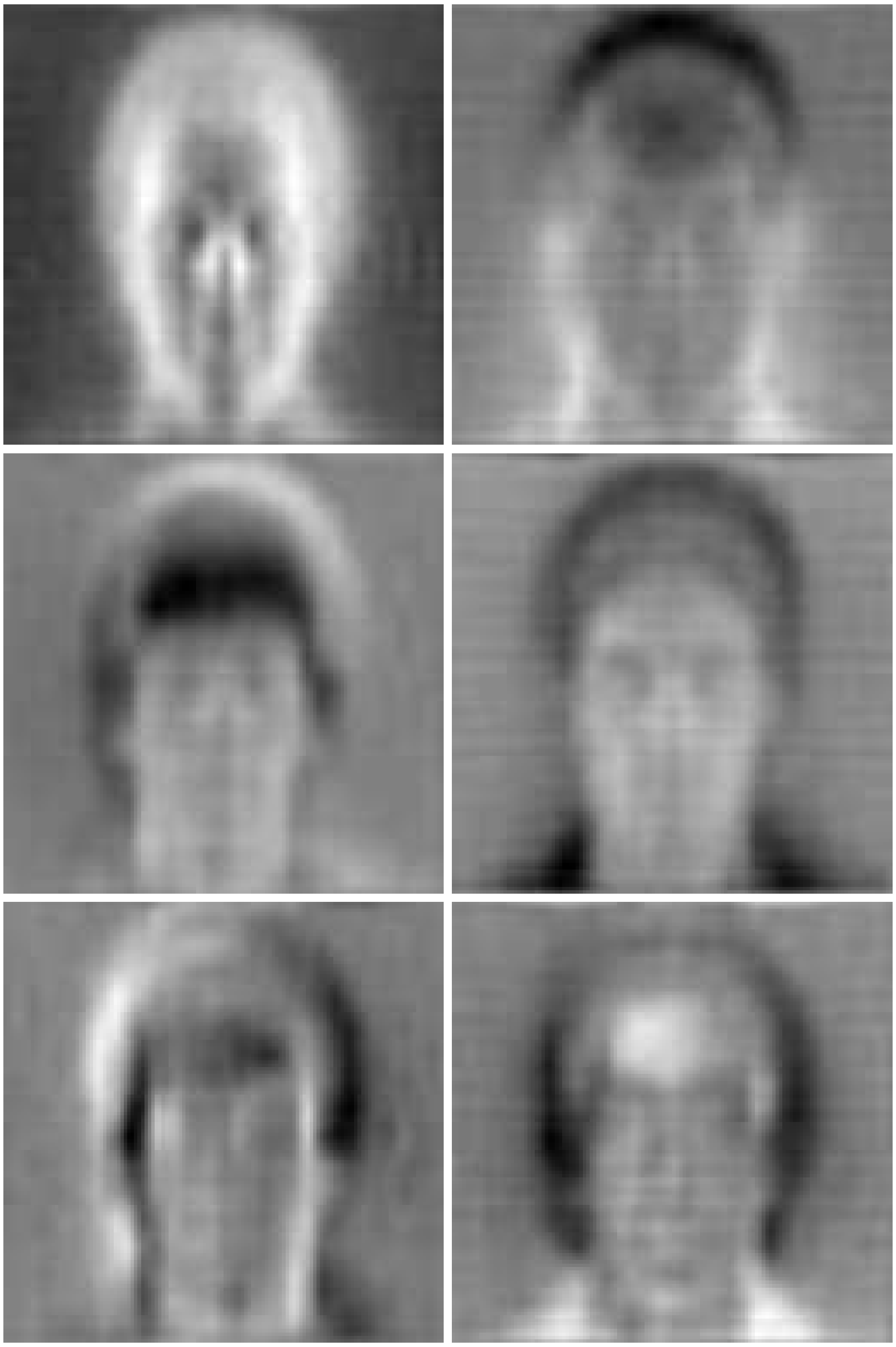

Figure 9. Top six Karhunen–Loève eigenfaces.

Figure 10 shows the top 6 joint best basis wavelet packets. Whereas the Karhunen–Loève basis functions look more or less like faces (or at least heads), the wavelet packets are abstract blobs which can be better localized at specific facial features.

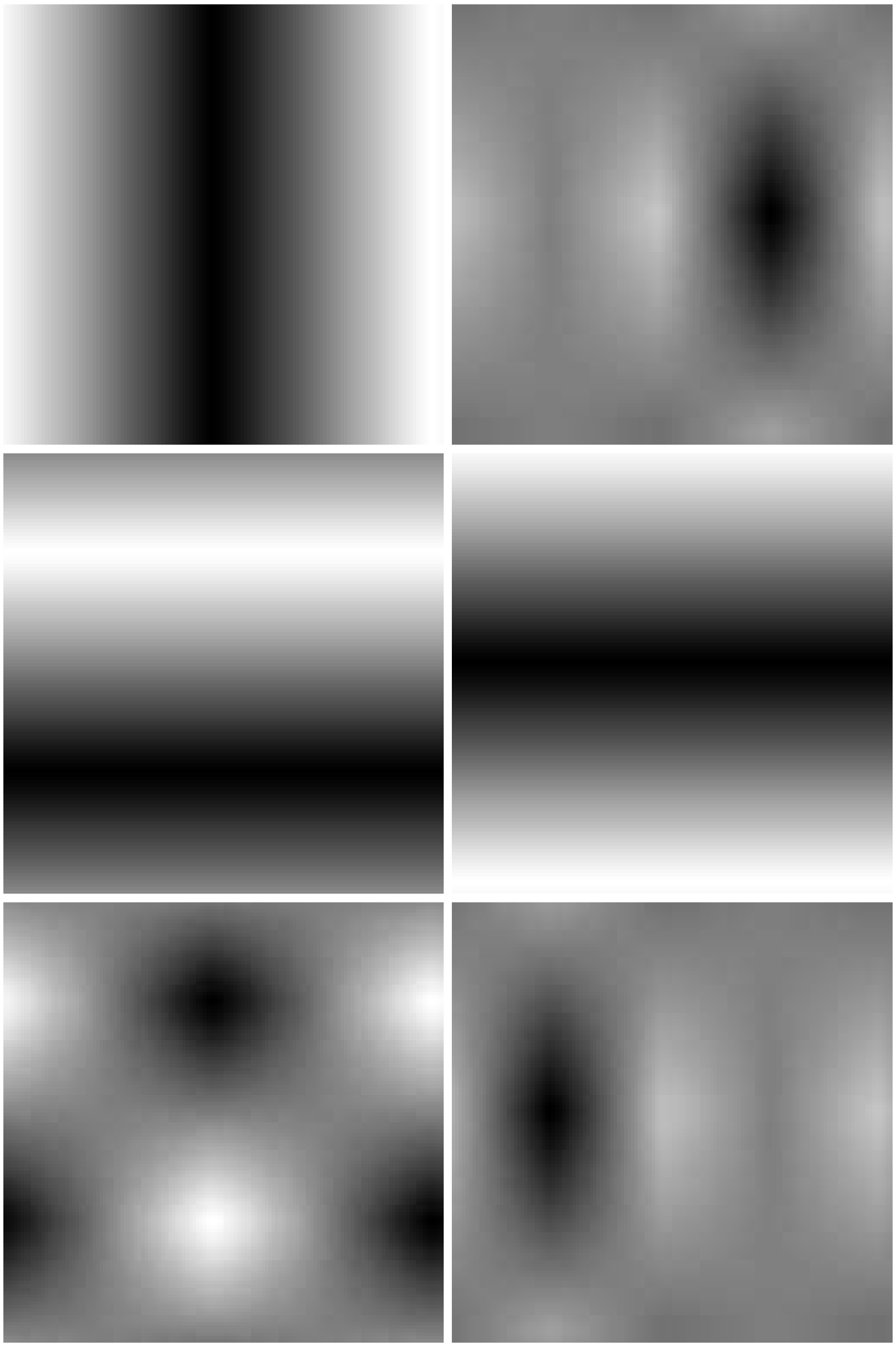

Figure 10. Top six joint best basis functions.

6. CONCLUSIONS

Wavelet packet analysis reduces the number of parameters needed to perform approximate Karhunen–Loève expansions. For a factor analysis "explaining" all but ϵ of the ensemble variance in a full-rank, d-dimensional collection of random vectors, the complexity will be $O(d^2 \log d + {d'}^3)$, where $d' \ll d$. Analysing a new vector in the approximate Karhunen–Loève basis has complexity $O(d \log d + d''d')$, where $d'' \ll d' \ll d$.

The analysis of one system of dimension 16,384, preserving 90% of the ensemble variance, can be performed on a desktop computer in minutes. The result is a reduction in dimension from $d = 16,384$ through $d' = 400$ to $d'' = 10$.

7. ACKNOWLEDGMENTS

Lawrence Sirovich of Brown University provided the normalized face images used in the "mug's gallery" application. Amir Auerbuch of Technion programmed the computation of Karhunen–Loève eigenvectors for the reduced parameter set. This research was supported in part by the Air Force Office of Scientific Research, the National Science Foundation, and by the Defense Advanced Research Projects Agency.

8. REFERENCES

[CMW] Ronald R. Coifman, Yves Meyer and M. Victor Wickerhauser, *Wavelet analysis and signal processing*, Wavelets and their Applications, ed. Beylkin *et al.*, Jones and Bartlett, Boston, 1991.

[CW] R. R. Coifman and M. V. Wickerhauser, *Entropy based methods for best basis selection*, IEEE Trans. Information Theory (1991) (to appear).

[JN] Nuggehally S. Jayant and Peter Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[W] M. V. Wickerhauser, *Picture compression by best-basis sub-band coding*, preprint, Yale University (1990).