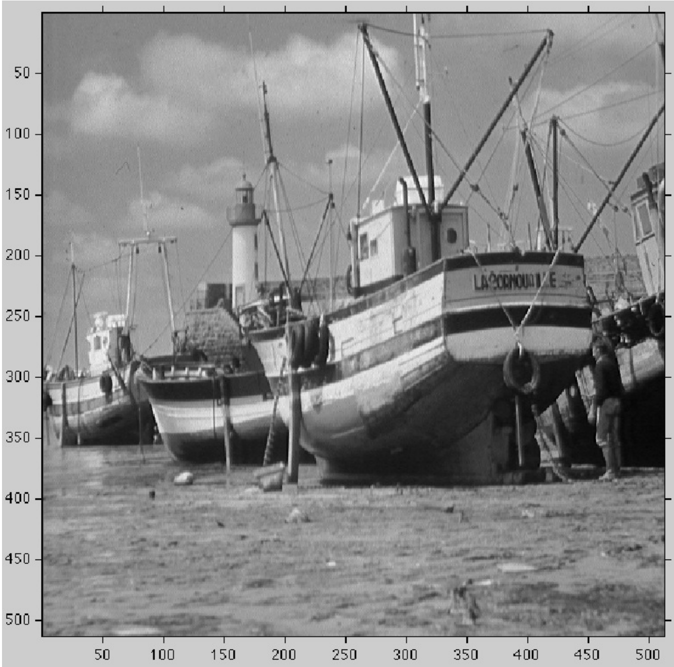
ELEG 467/667 - Imaging and Audio Signal Processing

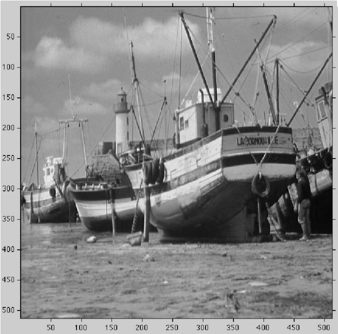
Gonzalo R. Arce

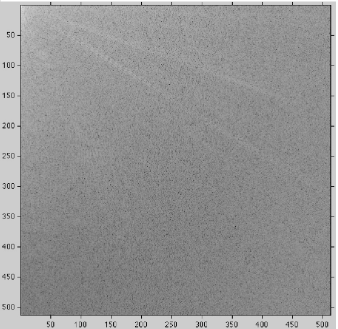
Chapter IV(e)

Department of Electrical and Computer Engineering
University of Delaware
Newark, DE, 19716

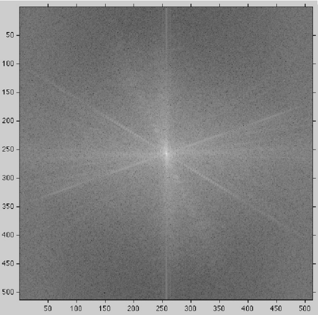
Spring 2013


50
100
150
200
250
300
350
400
450
500
LA CORNOUAILLE

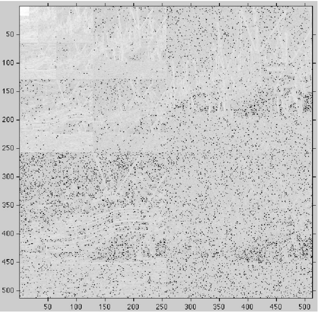

50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500
LA CORNOUAILLE


50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

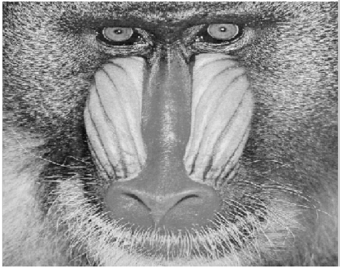
DCT

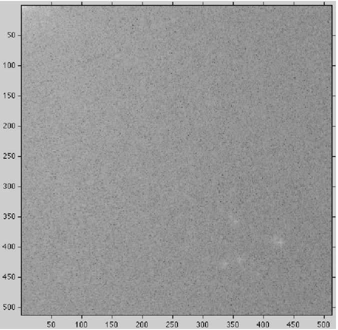

50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

FFT

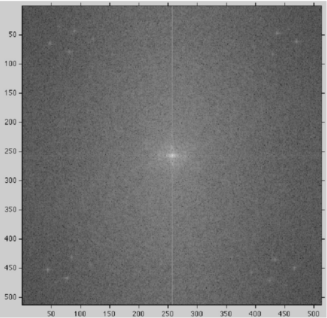

50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

HAAR

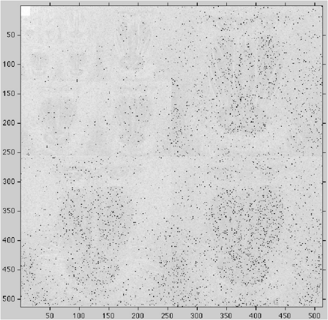



50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

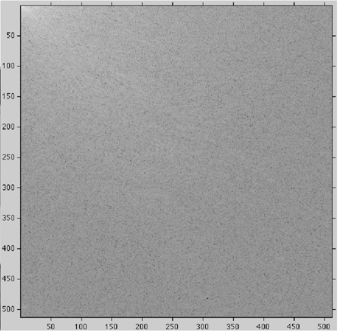
DCT


50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

FFT

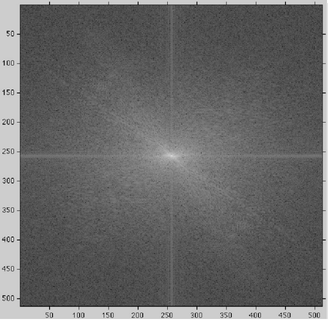

50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

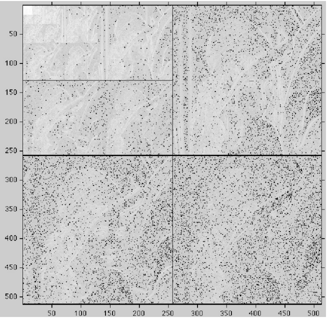
HAAR



DCT

FFT





HAAR

The Karhunen-Loeve Transform

A unitary transform represents a rotation of the signal vector in a particular N-dimensional space. Is one transform better than others and is there a best basis space?

The principal component analysis (PCA) and the associated Karhunen-Loeve Transform (KLT) will shed light to these questions.

Samples from a random signal $x(t)$, $x_m \ (m = 0, \cdots, N-1)$, form the vector

$$\mathbf{x} = [x_0, \cdots, x_{N-1}]^T,$$

with *mean vector* and covariance matrix given by

$$\mathbf{m}_x \triangleq E(\mathbf{x}) = [E(x_0), \cdots, E(x_{N-1})]^T = [\mu_0, \cdots, \mu_{N-1}]^T$$

$$\Sigma_x \triangleq E[(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T] = E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}_x \mathbf{m}_\mathbf{x}{}^T = \begin{bmatrix} .. & .. & .. \\ .. & \sigma_{ij}^2 & .. \\ .. & .. & .. \end{bmatrix}$$

with $\sigma_{ij}^2 \triangleq E(x_i x_j) - \mu_i \mu_j$ being the covariance of x_i and x_j. For $i = j$, we have $\sigma_i^2 \triangleq E(x_i - \mu_i)^2 = E(x_i^2) - \mu_i^2$.

If the signal is

- Complex, the covariance matrix is Hermitian, i.e., $\Sigma_x^{*T} = \Sigma_x$.
- Real, then $\Sigma_x^* = \Sigma_x$ is real and symmetric $\Sigma_x^T = \Sigma_x$.

The *correlation matrix* of X is defined as

$$\mathbf{R}_x \triangleq E(\mathbf{x}\mathbf{x}^T) = \begin{bmatrix} .. & .. & .. \\ .. & r_{ij} & .. \\ .. & .. & .. \end{bmatrix}$$

where $r_{ij} = E(x_i x_j) = \sigma_{ij}^2 + \mu_i \mu_j$.
Since $\sigma_{ij} = \sigma_{ji}$ and $r_{ij} = r_{ji}$, both $\Sigma_X = \Sigma_X^T$ and $\mathbf{R}_x = \mathbf{R}_x^T$ are symmetric.

For $\mathbf{y} = \mathbf{A}^T\mathbf{x}$, the mean and covariance are given by

$$\mathbf{m}_y = E(\mathbf{y}) = E(\mathbf{A}^T\mathbf{x}) = \mathbf{A}^T E(\mathbf{x}) = \mathbf{A}^T\mathbf{m}_x$$

$$\begin{aligned}
\Sigma_y &= E(\mathbf{y}\mathbf{y}^T) - \mathbf{m}_y\mathbf{m}_y^T \\
&= E[(\mathbf{A}^T\mathbf{x})(\mathbf{A}^T\mathbf{x})^T] - (\mathbf{A}^T\mathbf{m}_x)(\mathbf{A}^T\mathbf{m}_x)^T \\
&= E[\mathbf{A}^T(\mathbf{x}\mathbf{x}^T)\mathbf{A}] - \mathbf{A}^T\mathbf{m}_x\mathbf{m}_x^T\mathbf{A} \\
&= \mathbf{A}^T[E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}_x\mathbf{m}_x^T]\mathbf{A} && (1) \\
&= \mathbf{A}^T\Sigma_x\mathbf{A} && (2)
\end{aligned}$$

Karhunen-Loeve Transform (KLT)

Let ϕ_k be the eigenvector corresponding to the kth eigenvalue λ_k of the covariance matrix Σ_x, i.e.,

$$\Sigma_x \phi_k = \lambda_k \phi_k \qquad (k = 0, \cdots, N-1)$$

$$\begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \sigma_{ij} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \\ \phi_k \\ \\ \end{bmatrix} = \lambda_k \begin{bmatrix} \\ \phi_k \\ \\ \end{bmatrix} \qquad (k = 0, \cdots, N-1)$$

where the eigenvectors ϕ_i's are orthogonal:

$$< \phi_i, \phi_j >= \phi_i^T \phi_j^* = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

We then form an $N \times N$ unitary (orthogonal if $\mathbf{x}$ is real) matrix Φ

$$\Phi \triangleq [\phi_0, \cdots, \phi_{N-1}]$$

satisfying

$$\Phi^{*T}\Phi = \begin{bmatrix} \phi_0^{*T} \\ \phi_1^{*T} \\ \cdots \\ \phi_{N-1}^{*T} \end{bmatrix} \begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \end{bmatrix} = \mathbf{I}.$$

Hence,

$$\Phi^{-1} = \Phi^{*T}$$

The N eigenequations above can be combined into:

$$\Sigma_x \Phi = \Phi \Lambda$$

$$\begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \sigma_{ij} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} [\phi_0, \cdots, \phi_{N-1}] = [\phi_0, \cdots, \phi_{N-1}] \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_{N-1} \end{bmatrix}$$

where Λ is a diagonal matrix $\Lambda = diag(\lambda_0, \cdots, \lambda_{N-1})$.

Multiplying $\Phi^T = \Phi^{-1}$ on both sides, Σ_x can be diagonalized:

$$\Phi^{*T} \Sigma_x \Phi = \Phi^{-1} \Sigma_x \Phi = \Phi^{-1} \Phi \Lambda = \Lambda$$

Define a unitary Karhunen-Loeve Transform of **x** as:

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \Phi^{*T}\mathbf{x} = \begin{bmatrix} \phi_0^{*T} \\ \phi_1^{*T} \\ \vdots \\ \phi_{N-1}^{*T} \end{bmatrix} \mathbf{x}$$

where the ith component y_i of the transform vector is the projection of **x** onto ϕ_i:

$$y_i =< \phi_i, \mathbf{x} >= \phi_i^{*T}\mathbf{x}$$

Left multiplying $\Phi = (\Phi^{*T})^{-1}$ on both sides of the transform $\mathbf{y} = \Phi^{*T}\mathbf{x}$, we get the inverse transform:

$$\mathbf{x} = \Phi\mathbf{y} = [\phi_0, \phi_1, \cdots, \phi_{N-1}] \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \sum_{i=0}^{N-1} y_i \phi_i$$

The vector **x** is then represented in an N-dimensional space spanned by the N eigenvector basis set ϕ_i ($i = 0, \cdots, N-1$).

KLT Completely Decorrelates the Signal

Among all possible orthogonal transforms, KLT is optimal in that

- The KLT completely decorrelates the signal
- The KLT maximally compacts the energy (information) contained in the signal.

To see the first property, consider the mean vector $\mathbf{m}_y$ and covariance matrix Σ_y of $\mathbf{y} = \Phi^T\mathbf{x}$:

$$\mathbf{m}_y = \Phi^T\mathbf{m}_x$$

$$\begin{aligned}
\Sigma_y &= E(\mathbf{y}\mathbf{y}^T) - \mathbf{m}_y\mathbf{m}_y^T && (3)\\
&= E[(\Phi^T\mathbf{x})(\Phi^T\mathbf{x})^T] - (\Phi^T\mathbf{m}_x)(\Phi^T\mathbf{m}_x)^T \\
&= E[\Phi^T(\mathbf{x}\mathbf{x}^T)\Phi] - \Phi^T\mathbf{m}_x\mathbf{m}_x^T\Phi \\
&= \Phi^T[E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}_x\mathbf{m}_x^T]\Phi \\
&= \Phi^T\Sigma_x\Phi = \Lambda \\
&= diag[\lambda_0, \lambda_1, \cdots, \lambda_{N-1}] && (4)
\end{aligned}$$

The above can be written in matrix form:

$$\Sigma_y = \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \sigma_{ij} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} = \Phi^T \Sigma_x \Phi = \Lambda = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_{N-1} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_{N-1}^2 \end{bmatrix}$$

Hence:

- After the KLT, the covariance matrix of the signal $\mathbf{y} = \Phi^T \mathbf{x}$ is diagonalized, i.e., the covariance $\sigma_{ij} = 0$ between y_i and y_j is always zero.
- The signal is completely decorrelated.
- The variance of y_i is the same as the ith eigenvalue of the covariance matrix of $\mathbf{x}$, i.e., $\sigma_i^2 = \lambda_i$.

Energy Compaction of the KLT

Here we show that the KLT maximally compacts the signal energy into a small number of components.

Let $\mathbf{A}$ be an arbitrary orthogonal matrix satisfying $\mathbf{A}^{-1} = \mathbf{A}^T$, and represent $\mathbf{A}$ in terms of its column vectors $\mathbf{a}_i$,

$$\mathbf{A} = [\mathbf{a}_0, \cdots, \mathbf{a}_{N-1}], \quad \text{or} \quad \mathbf{A}^T = \begin{bmatrix} \mathbf{a}_0^T \\ \vdots \\ \mathbf{a}_{N-1}^T \end{bmatrix}.$$

$\mathbf{A}$ then defines an orthogonal transform as

$$\mathbf{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix} = \mathbf{A}^T\mathbf{x} = \begin{bmatrix} \mathbf{a}_0^T \\ \vdots \\ \mathbf{a}_{N-1}^T \end{bmatrix} \mathbf{x}$$

where the ith component of $\mathbf{y}$ is $y_i = \mathbf{a}_i^T\,\mathbf{x}$. The inverse transform is:

$$\mathbf{x} = \mathbf{A}\mathbf{y} = [\mathbf{a}_0, \cdots, \mathbf{a}_{N-1}] \begin{bmatrix} y_0 \\ \cdots \\ y_{N-1} \end{bmatrix} = \sum_{i=0}^{N-1} y_i \mathbf{a}_i$$

The variances of the signal components before and after the KLT transform are:

$$\sigma_{x_i}^2 = E[(x_i - \mu_{x_i})^2] \stackrel{\triangle}{=} E(e_{x_i}), \quad \text{and} \quad \sigma_{y_i}^2 = E[(y_i - \mu_{y_i})^2] \stackrel{\triangle}{=} E(e_{y_i})$$

where $e_{x_i} \stackrel{\triangle}{=} (x_i - \mu_{x_i})^2$ is the energy in the ith component, and the trace matrix $tr\Sigma_x$ represents the total energy

$$\text{The total energy in } \mathbf{x} = tr\Sigma_x = \sum_{i=0}^{N-1} \sigma_{x_i}^2 = \sum_{i=0}^{N-1} E(e_{x_i})$$

Since $tr(\mathbf{AB}) = tr(\mathbf{BA})$,

$$tr\Sigma_y = tr(\Phi^T \Sigma_x \Phi) = tr(\Phi^T \Phi \Sigma_x) = tr\Sigma_x$$

The total energy of the signal is thus conserved after the KLT transform.

However, the energy distribution among all N components can be very different before and after the KLT transform.

Define the energy contained in the first $m < N$ components after the transform $\mathbf{y} = \mathbf{A}^T\mathbf{x}$ as

$$P_m(\mathbf{A}) \triangleq \sum_{i=0}^{m-1} E[(y_i - \mu_{y_1})^2] = \sum_{i=0}^{m-1} \sigma_{y_i}^2 = \sum_{i=0}^{m-1} E(e_{y_i})$$

$P_m(\mathbf{A})$ will be maximized if and only if (iff) the transform matrix is the same as that of the KLT:

$$P_m(\Phi) \geq P_m(\mathbf{A}).$$

The KLT is optimal in compacting the energy into a few components of the transformed signal.

Consider

$$
\begin{aligned}
P_m(\mathbf{A}) &\triangleq \sum_{i=0}^{m-1} E(y_i - \mu_{y_1})^2 = \sum_{i=0}^{m-1} E[\mathbf{a}_i^T(\mathbf{x}-\mathbf{m}_{x_i})\,\mathbf{a}_i^T(\mathbf{x}-\mathbf{m}_{x_i})] \\
&= \sum_{i=0}^{m-1} E[\mathbf{a}_i^T(\mathbf{x}-\mathbf{m}_{x_i})\,(\mathbf{x}-\mathbf{m}_{x_i})^T\mathbf{a}_i] \\
&= \sum_{i=0}^{m-1} \mathbf{a}_i^T E[(\mathbf{x}-\mathbf{m}_{x_i})(\mathbf{x}-\mathbf{m}_{x_i})^T]\mathbf{a}_i \\
&= \sum_{i=0}^{m-1} \mathbf{a}_i^T \Sigma_x \mathbf{a}_i
\end{aligned}
$$

The objective is to find a transform matrix **A** such that

$$\arg\max_{\mathbf{A}} P_m(\mathbf{A}) \quad \text{subject to:} \quad \mathbf{a}_j^T\mathbf{a}_j = 1 \quad (j = 0, \cdots, m-1)$$

The constraint $\mathbf{a}_j^T\mathbf{a}_j = 1$ is to guarantee that the column vectors in **A** are normalized. This constrained optimization problem can be solved using Lagrange multipliers by letting the following partial derivative be zero:

$$\frac{\partial}{\partial \mathbf{a}_i}[P_m(\mathbf{A}) - \sum_{j=0}^{m-1} \lambda_j(\mathbf{a}_j^T\mathbf{a}_j - 1)] = \frac{\partial}{\partial \mathbf{a}_i}[\sum_{j=0}^{m-1}(\mathbf{a}_j^T\Sigma_x\mathbf{a}_j - \lambda_j\mathbf{a}_j^T\mathbf{a}_j + \lambda_j)]$$

$$= \quad \frac{\partial}{\partial \mathbf{a}_i}[\mathbf{a}_i^T\Sigma_x\mathbf{a}_i - \lambda_i\mathbf{a}_i^T\mathbf{a}_i] \stackrel{*}{=} 2\Sigma_x\mathbf{a}_i - 2\lambda_i\mathbf{a}_i = 0$$

The column vectors of **A** must then be the eigenvectors of Σ_x:

$$\Sigma_x\mathbf{a}_i = \lambda_i\mathbf{a}_i \quad (i = 0, \cdots, m-1)$$

The transform matrix must be

$$\mathbf{A} = [\mathbf{a}_0, \cdots, \mathbf{a}_{N-1}] = \Phi = [\phi_0, \cdots, \phi_{N-1}]$$

where ϕ_i's are the orthogonal eigenvectors of Σ_x corresponding to eigenvalues λ_i ($i = 0, \cdots, N-1$):

$$\Sigma_x \phi_i = \lambda_i \phi_i, \quad \text{i.e.} \quad \phi^T \Sigma_x \phi_i = \lambda_i \phi_i^T \phi_i = \lambda_i$$

The optimal transform is thus the KLT with

$$P_m(\Phi) = \sum_{i=0}^{m-1} \phi_i^T \Sigma_x \phi_i = \sum_{i=0}^{m-1} \lambda_i$$

where the eigenvalue λ_i of Σ_x is also the energy contained in the ith component.

If we choose those $\phi_i's$ that correspond to the m largest eigenvalues of Σ_x: $\lambda_0 \geq \lambda_1 \geq \cdots \lambda_m \cdots \geq \lambda_{N-1}$, then $P_m(\Phi)$ will be maximized.

The KLT can be used for data compression by reducing the dimensionality of the data as follows:

- Find the mean vector $\mathbf{m}_x$ and covariance matrix Σ_x.
- Find the eigenvalues λ_i of Σ, sorted in descending order, and their corresponding eigenvectors ϕ_i $(i = 0, \cdots, N-1)$.
- Choose a lowered dimensionality $m < N$ so that the percentage of energy contained $\sum_{i=0}^{m-1} \lambda_i / \sum_{i=0}^{N-1} \lambda_i$ is no less than a given threshold (e.g., 95%).

- Construct an N by m transform matrix composed of the m largest eigenvectors of Σ_x:

$$\Phi_m = [\phi_0, \cdots, \phi_{m-1}]$$

and compute the KLT based on Φ_m:

$$\mathbf{y} = \Phi_m^T \mathbf{x}$$

or

$$\begin{bmatrix} y_0 \\ \cdots \\ y_{m-1} \end{bmatrix}_{m\times 1} = \begin{bmatrix} \phi_0^T \\ \cdots \\ \phi_{m-1}^T \end{bmatrix}_{m\times N} \begin{bmatrix} x_0 \\ \cdots \\ x_{N-1} \end{bmatrix}_{N\times 1}$$

Since the dimensionality m of $\mathbf{y}$ is less than the dimensionality N of $\mathbf{x}$, data compression is attained.

This is a lossy compression with the error representing the percentage of information lost: $\sum_{i=m}^{N-1} \lambda_i / \sum_{i=0}^{N-1} \lambda_i$.
But as these λ_i's are the smallest eigenvalues, the error is small (e.g., 5%).
Carry out inverse KLT for reconstruction:

$$\mathbf{x} = \Phi_m \mathbf{y}$$

$$\begin{bmatrix} x_0 \\ \cdots \\ x_{N-1} \end{bmatrix}_{N\times 1} = \begin{bmatrix} \phi_0 \cdots \phi_{m-1} \end{bmatrix}_{N\times m} \begin{bmatrix} y_0 \\ \cdots \\ y_{m-1} \end{bmatrix}_{m\times 1}$$

Geometric Interpretation of the KLT

Assume $\mathbf{x} = [x_0, \cdots, x_{N-1}]^T$ is a multivariate Gaussian vector with a joint probability density function:

$$p(x_0, \cdots, x_{N-1}) = N(\mathbf{x}, \mathbf{m}_x, \Sigma_x) = \frac{1}{(2\pi)^{N/2} |\Sigma_x|^{1/2}} exp[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_x)^T \Sigma_x^{-1} (\mathbf{x} - \mathbf{m}_x)]$$

where $\mathbf{m}_x$ and Σ_x are the mean vector and covariance matrix of $\mathbf{x}$

When $N = 1$, Σ_x and $\mathbf{m}_x$ become σ_x and μ_x, and $p(\mathbf{x})$ becomes the single variable normal distribution:

$$p(x) = N(x, \mu_x, \sigma_x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} exp[-\frac{(x - \mu_x)^2}{2\sigma_x^2}].$$

The shape of this N-dimensional density is depicted by the iso-value hyper-surface

$$(\mathbf{x}-\mathbf{m}_x)^T\Sigma_x^{-1}(\mathbf{x}-\mathbf{m}_x)=c_1$$

where c_1 is a constant. For $N=2$, we have

$$\begin{aligned}(\mathbf{x}-\mathbf{m}_x)^T\Sigma_x^{-1}(\mathbf{x}-\mathbf{m}_x) &= [x_0-\mu_{x_0}, x_1-\mu_{x_1}]\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}\begin{bmatrix} x_0-\mu_{x_0} \\ x_1-\mu_{x_1} \end{bmatrix} \\ &= a(x_0-\mu_{x_0})^2+b(x_0-\mu_{x_0})(x_1-\mu_{x_1})+c(x_1-\mu_{x_1})^2=\end{aligned}$$

Here we have assumed

$$\Sigma_x^{-1} = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$$

The above quadratic equation represents an ellipse (instead of other quadratic curves such as hyperbola and parabola) centered at $\mathbf{m}_x = [\mu_0, \mu_1]^T$, because Σ_x^{-1} and Σ_x are both positive definite, i.e.,

$$\left|\Sigma_x^{-1}\right| = ac - b^2/4 > 0$$

For $N > 2$, $N(\mathbf{x}, \mathbf{m}_x, \Sigma_x) = c_0$ represents a hyper ellipsoid in N-dimensions.

The center and spatial distribution of this ellipsoid are determined by $\mathbf{m}_x$ and Σ_x.

The KLT $\mathbf{y} = \Phi^T \mathbf{x}$ completely decorrelates $\mathbf{x}$ such that

$$\Sigma_y = \Lambda = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_{N-1} \end{bmatrix} = \begin{bmatrix} \sigma_{y_0}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{y_1}^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_{y_{N-1}}^2 \end{bmatrix}$$

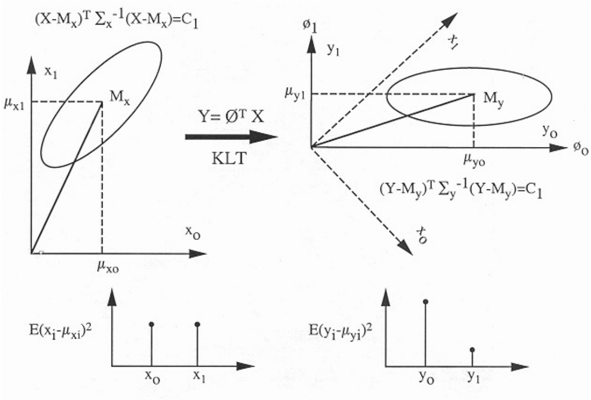
Equation $N(\mathbf{x}, \mathbf{m}_x, \Sigma_x) = c_0$ becomes $N(\mathbf{y}, \mathbf{m}_y, \Sigma_y) = c_0$, or

$$(\mathbf{y} - \mathbf{m}_y)^T \Sigma_y^{-1} (\mathbf{y} - \mathbf{m}_y) = \sum_{i=0}^{N-1} \frac{(y_i - \mu_{y_i})^2}{\lambda_i} = \sum_{i=0}^{N-1} \frac{(y_i - \mu_{y_i})^2}{\sigma_{y_i}^2} = c_1$$

This equation represents a standard ellipsoid in the N-dimensional space.

The KLT rotates the coordinate system so that the semi-principal axes of the ellipsoid associated with the normal distribution are in parallel with ϕ_i, the axes of the new coordinate system.

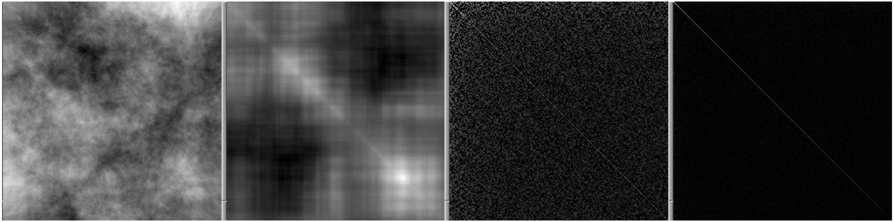
Moreover, the length of the semi-principal axis parallel to ϕ_i is equal to $\sqrt{\lambda_i} = \sigma_{y_i}$.



Comparison with Other Orthogonal Transforms

Example 1 Consider a 256×256 image *clouds* where each row is treated as an observation sample of a $1-$D random vector $\mathbf{x}$ (with 256 components). The various transforms $\mathbf{y} = \mathbf{A}^T\mathbf{x}$ are applied to $\mathbf{x}$, and the corresponding covariance matrices Σ_y are compared.

Results for the DFT and WHT are very similar to that of DCT. A conversion $y = x^{0.3}$ has applied to the intensity of the images for covariance matrices for the low values to be still visible.



The energy compaction is illustrated in the following table showing the number of components needed to keep a percentage of the energy.

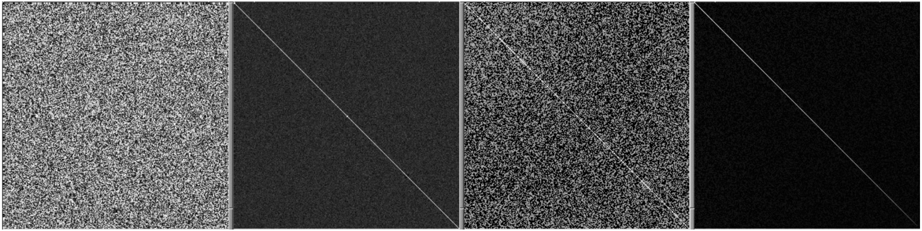
To keep 99% of the total energy, 250 out of 256 components are needed for no transform, 97 out of 256 are needed after the DCT, and 55 after the KLT.

Percentage	90	95	99	100
Identity	209	230	250	256 (all)
DCT	10	22	97	256 (all)
KLT	7	13	55	256 (all)

- All orthogonal transforms tend to decorrelate signals, and the KLT does it optimally.
- Orthogonal transforms tend to compact the energy into a small number of components. For example, after the DFT most of the energy will be concentrated in the DC component and a relatively small number of low frequency components. The same is also true for DCT and WHT.

Orthogonal transforms tend to reduce correlation when signals are continuous and smooth. In other cases, orthogonal transforms may not perform well, and the energy will not necessarily be compacted.

Example The left figure shows texture of sand, where the pixels are not correlated as in the previous example, since the color of a grain of sand is not related to those of the neighboring grains. The second image shows the covariance matrix of the row vectors of the image. The 3*rd* and 4*th* figures shows the covariances after DCT and KLT.



Although the KLT is optimal, other transforms are widely used for two reasons.

- First, KLT transform matrix is composed of the eigenvectors of the covariance matrix Σ_x, which can be estimated only if sufficient amount of data is available.
- Second, the computational cost for KLT is much higher than other transforms,
    - We need to estimate the covariance matrix Σ_x and solve its eigenvalue problem to obtain the transform matrix Φ.
    - Fast KLT transforms do not exist. The complexity of the transform is $O(N^2)$, instead of $Nlog_2N$ for most of other transforms.

Assume a set of N images of size $K = rows \times columns$ are to be stored or transmitted.

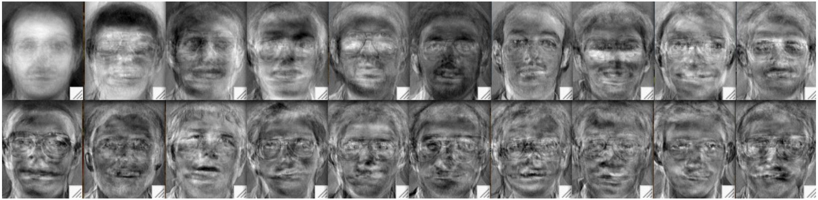
The pixels of the same position in all these images are used to form a N-dimensional vector and there are in total K such vectors.

Treating these vectors as random vectors, we can find their mean vector $\mathbf{m}$ and covariance matrix Σ, and the KLT can be carried out to transform these vectors into a lower dimensional space of $m << N$ dimensions.

Example:

Twenty images of faces:

The eigen-images after KLT:



Percentage of energy contained in the

Percentage of energy contained in the

components	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
percentage energy	48.5	11.6	6.1	4.6	3.8	3.7	2.6	2.5	1.9	1.9	1.8	1.6	1.5	1.4	1.3	1.2	1.1	1.1	0.9	0.8
accumulative energy	48.5	60.1	66.2	70.8	74.6	78.3	81.0	83.5	85.4	87.3	89.	90.7	92.2	93.6	94.9	96.1	97.2	98.2	99.2	100.0

Reconstructed faces using 95% of the total information (15 out of 20 components):

Singular Value Decomposition

Let a 2D image be represented by a matrix $\mathbf{A} = [a_{ij}]_{M\times N}$ with rank equal to R . Here we assume $R \leq M \leq N$. Consider next the eigenvalue decompositions for $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_i = \lambda i\mathbf{u}_i, \qquad \mathbf{A}^T\mathbf{A}\mathbf{v}_i = \lambda i\mathbf{v}_i, \tag{5}$$

where λ_i, are the eigenvalues of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$, for $(i = 1, 2, \cdots, R)$. Since these matrices are symmetric, their eigenvectors are orthogonal:

$$\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij} = \mathbf{v}_i^T\mathbf{v}_j = \delta_{ij} \tag{6}$$

forming two orthogonal matrices $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_N]_{M\times N}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_N]_{N\times N}$ such that

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}, \qquad \mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}. \tag{7}$$

Since there are only R non-zero eigenvalues, These matrices satisfy

$$\mathbf{U}^T(\mathbf{A}\mathbf{A}^T)\mathbf{U} = \Lambda_{M\times M} = \mathrm{diag}[\lambda_1, \cdots, \lambda_R] \tag{8}$$

and

$$\mathbf{V}^T(\mathbf{A}^T\mathbf{A})\mathbf{V} = \Lambda_{N\times N} = \mathrm{diag}[\lambda_1, \cdots, \lambda_R] \tag{9}$$

The singular value decomposition (SVD) is of a matrix **A** is defined as

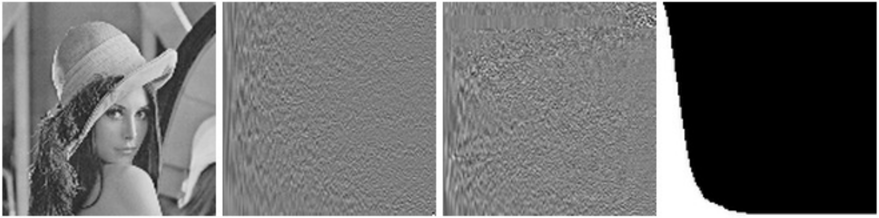
$$\mathbf{U}^T\mathbf{A}\mathbf{V} = \Lambda^{1/2}_{M\times M} = \mathrm{diag}[\sqrt{\lambda_1}, \cdots, \sqrt{\lambda}_R] \tag{10}$$

From the properties of **U** and **V**, the inverse of the SVD decomposition is

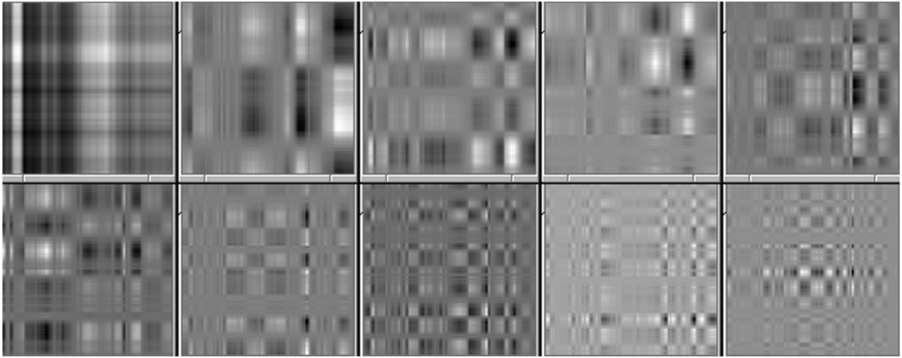
$$
\begin{aligned}
\mathbf{A} &= \mathbf{U}\Lambda^{1/2}\mathbf{V}^T = [\mathbf{u_1},...,\mathbf{u_N}]\begin{bmatrix} \sqrt{\lambda_1} & ... & 0 \\ ... & ... & ... \\ ... & \sqrt{\lambda_R} & ... \\ ... & ... & ... \\ 0 & ... & 0 \end{bmatrix}[\mathbf{v_1},...,\mathbf{v_N}]^T \qquad (11)\\
&= [\sqrt{\lambda_1}\mathbf{u}_1 \sqrt{\lambda_2}\mathbf{u}_2 ... \sqrt{\lambda_R}\mathbf{u}_R ... 0]\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ ... \\ \mathbf{v}_N^T \end{bmatrix} \qquad (12)\\
&= \begin{bmatrix} \sqrt{\lambda_1}u_{11}v_{11} + ... + \sqrt{\lambda_R}u_{R1}v_{R1} & \sqrt{\lambda_1}u_{11}v_{12} + ... + \sqrt{\lambda_R}u_{R1}v_{R2} & ... \\ ... & ... & ... \\ \sqrt{\lambda_1}u_{1M}v_{11} + ... + \sqrt{\lambda_R}u_{RM}v_{R1} & \sqrt{\lambda_2}u_{1M}v_{12} + ... + \sqrt{\lambda_R}u_{RM}v_{R2} & ... \end{bmatrix} \qquad (13)\\
&= \sum_{i=1}^{R}\sqrt{\lambda_i}[\mathbf{u}_i\mathbf{v}_i^T]. \qquad (14)
\end{aligned}
$$

The inverse SVD transform can be interpreted as the representation of the image matrix **A** decomposed into a set of eigenimages $\sqrt{\lambda_i}[\mathbf{u}_i\mathbf{v}_i^T]$ of size M by N.

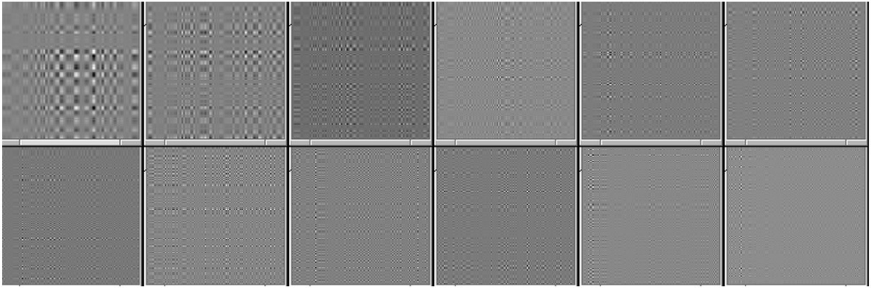
Lenna Image, U, V and Singular Values



First 10 eigen-images of the Lenna image



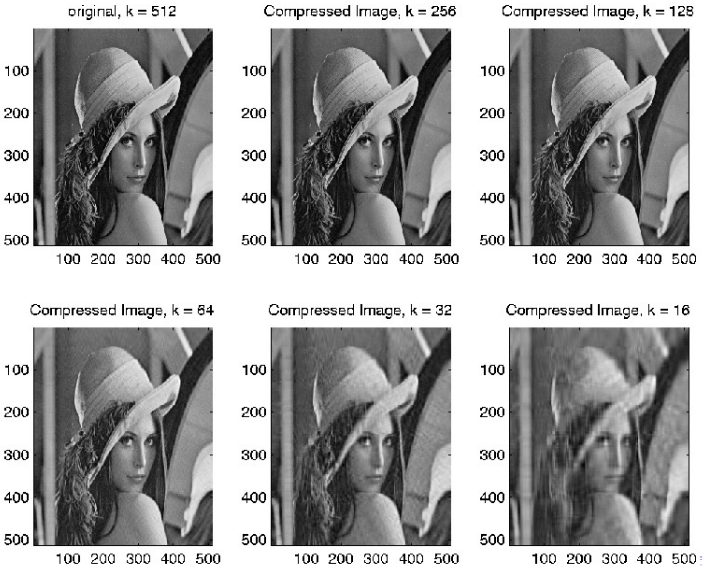
Eigen-images (from 10 to 120 with increment of 10)



First 10 Partial Sum Images

Partial Sum Images (from 10 to 120 eigen-images with increment 10)

Reconstructed image for different ranks



Absolute difference between the original and the reconstructed images