

Multimedia representation and compression

1. (a) For the top edge of the sphere, there are DCT values along the left column of each 8×8 DCT block and near zero elsewhere. This is because the cosine functions with $v = 0$ are higher and higher frequencies in the y -direction, and no change in the x -direction; so there is a contribution to these cosines whereas the cosines with $u = 0$ are sinusoids changing in the x -direction and not in the y -direction, but there is no corresponding change in the x -direction at the top of the sphere, so there result DCTs with values near zero. For the left (or right) edge of the sphere, there are DCT values along the top row of each 8×8 and near zero elsewhere.
- (b) Still lots of low-frequency; but also more higher frequencies because of the sharp edges of the beach ball colors.
- (c) For the block that straddle the top edge, as in (a) there are DCT values along the left column of each 8×8 block and near zero elsewhere, although some coefficients in the column, e.g., $F(1;0)$, will now have much larger (absolute) value in response to the change from black to grey.
- (d) For the block that straddle the left edge, this change happens in the top row of each 8×8 block.
2. You can full marks as long as you turn in all the plots.
3. Based on the relation between the good images \mathbf{X}_G and the bad images \mathbf{X}_B , we can break \mathbf{X} into two components:

$$\mathbf{X} = \begin{pmatrix} \mathbf{I}_p \\ \mathbf{A} \end{pmatrix} \mathbf{X}_G + \begin{pmatrix} \mathbf{0} \\ \mathbf{N} \end{pmatrix} \quad (1)$$

In the limiting case when $k = p$, the entire X_G is received at the central server and the best prediction of X given X_G is just the first component of the right hand side of (1):

$$E(\mathbf{X}|\mathbf{X}_G) = \begin{pmatrix} \mathbf{I}_p \\ \mathbf{A} \end{pmatrix} \mathbf{X}_G \quad (2)$$

When $k < p$, the best one can hope for is to use the best k -dimensional approximation of $E(\mathbf{X}|\mathbf{X}_G)$ (not X_G as we will see in the following toy example). The KLT \mathbf{V} can be computed in the following way:

$$\begin{aligned} E\left[\begin{pmatrix} \mathbf{I}_p \\ \mathbf{A} \end{pmatrix} \mathbf{X}_G \mathbf{X}_G^T (\mathbf{I}_p \mathbf{A}^T)\right] &= \begin{pmatrix} \mathbf{I}_p \\ \mathbf{A} \end{pmatrix} \Sigma_{X_G} (\mathbf{I}_p \mathbf{A}^T) \\ &= \mathbf{V}^T \mathbf{\Lambda}' \mathbf{V} \end{aligned} \quad (3)$$

and the result follows.

A toy example illustrating the basic issue is the following: Suppose that a Gaussian random vector \mathbf{X} has mean zero and the following covariance matrix:

$$\Sigma_X = \begin{pmatrix} \sigma_1^2 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix} \quad (4)$$

Suppose that the first two components constitute the good camera \mathbf{X}_G . Since \mathbf{X} is jointly Gaussian, the matrix \mathbf{A} is found to be $\mathbf{A} = E(\mathbf{X}_B \mathbf{X}_G^T) \Sigma_{\mathbf{X}_G}^{-1}$. The encoder is asked to provide a 1-dimensional approximation. For $\sigma_1^2 = 0.11$, applying the usual KLT to the first two components is simple in this example: the first two components are uncorrelated, hence the KLT is the identity. Picking the eigenvector corresponding to the larger eigenvalue of $\Sigma_{\mathbf{X}_G}$ incurs a distortion of 1.9182. Using KLT as described in (4), and hence making the optimal choice, results in a distortion of 1.3795, and the transform is

$$\mathbf{V} = \begin{pmatrix} 1.1119 & 2.6353 \\ 1.1902 & -0.5524 \end{pmatrix} \quad (5)$$

It is clear that this matrix is substantially different from the usual KLT (applied to the first two components).