

# Principal Component Analysis

Principal component analysis (PCA) is a statistical procedure that provides a sequence of best linear approximations to a given high-dimensional observation. This PCA application takes a classic image (lena) in image processing as an example to illustrate the dimension reduction.

To the left is the original jpeg image (512 by 512). As a result of the PCA calculation, the total 512 of principal components are ordered in such a way that the first principal component explains the highest variance in the data and the succeeding component accounts for less variance than the preceding one but highest variance in the rest of the data.

Please use the slider on the left to select the number of principal components ( $n$ ) to keep. Accordingly, the results on all the 4 panels on the right will be updated:

1 **Image:** The left image is produced using the first  $n$  principal components, and you will see that the higher  $n$  is, the clearer the image is. The right figure shows the percentage of variance explained by the first  $n$  principal components. Note that the first 31 principal components account for 90% of the variance and the resulted right image acquires most characters of the original image.

2 **Variance:** The variance figure shows that the principal components are ordered in such a way that the variance each of the components explains is decreasing. Note that the variance decreases sharply for the first 10 principal components, and each of the rest principal components explains less than 1% of the variance.

3 **PCA:** The PCA figure illustrates that each principal component is a linear combination of the original variables (the columns). Note that principal component that accounts for  $<1\%$  of the variance seems a random combination of the original variables.

4 **Data:** The Data panel displays the original data. there is a range slider for selecting the columns of the data to display.