A12 – Feature Extraction

Introduction

If a pattern recognition task is fairly generative, the features necessary for classification is more or less known. Consider the task of automatically identifying fruits, say bananas, oranges and ripe mangoes. From visual observation, one can use two features, shape and color, to differentiate these fruits. The challenge is to represent these features as numbers which we can manipulate. One way is to express shape as the eccentricity of the fruit outline, (e=0 for a circle, e <1 for an ellipse). Color can be expressed as a number from 1 to 7 to indicate hue (e.g. ROYGBIV into to 7). We can then plot measurements of several fruit samples in feature space and we expect a clustering of feature points per fruit class.

Experiencing firsthand feature extraction will make you appreciate why data is the currency of today's knowledge economy.

Procedure

1. Download images of bananas, oranges and mangoes and tabulate their features. Alternatively you may use other classes of objects but you must decide on at most three visual features by which each of the objects can be differentiated by. These features must be represented as numbers. Measure these features using image processing techniques.
2. Collect as much data as possible, for example, around >20 samples per class.
3. In most cases it is helpful to normalize the data to have values in the range [0,1]. Per feature and across all classes, normalize the data to lie within this range.
4. Plot the feature points in 2D or 3D space and discuss if class clustering is observed.