Leaf Classification from Boundary Analysis

Anne Jorstad
AMSC 663
Midterm Progress Report
Fall 2007

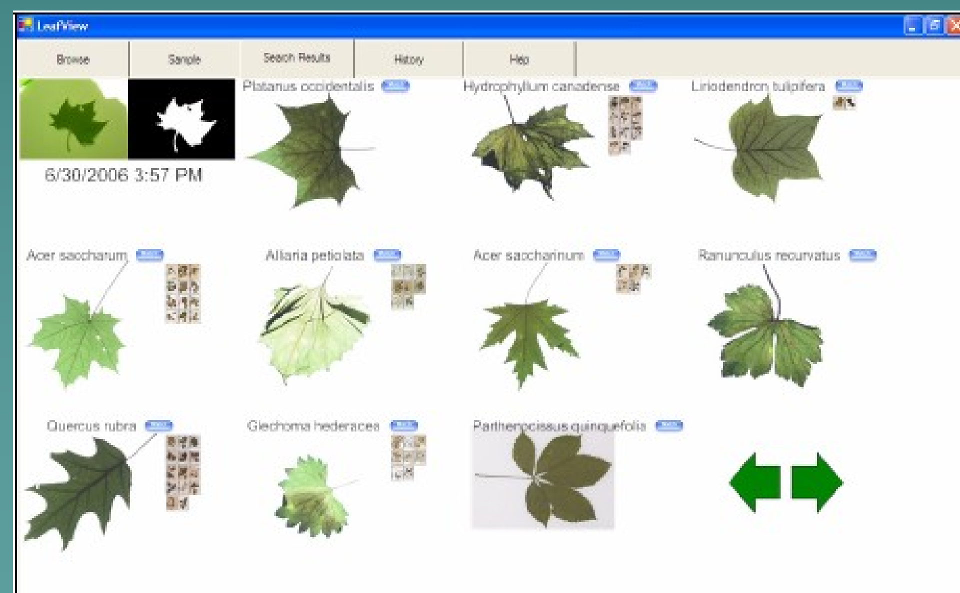
Advisor: Dr. David Jacobs, Computer Science

Outline

- Background, Problem Statement
- Algorithm
- Validation
- Test Results
- Schedule, Future Work

Background

- Electronic Field Guide for Plants

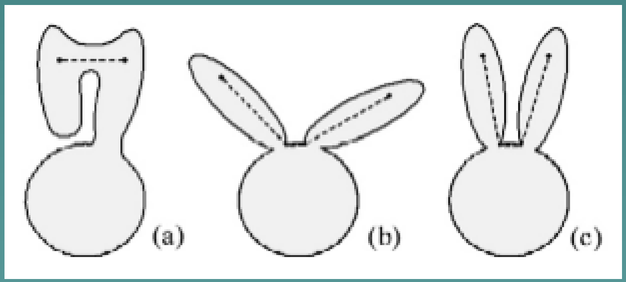


Background

- Current System:
  - Inputs photo of leaf on plain background
  - Segments leaf from background
  - Compares leaf to all leaves in database, using global shape information
  - Returns images of closest matches to the user

Background

- Current System:
  - Inner-Distance Measure
    - Measures the shortest distance between two points on a path contained entirely within a figure



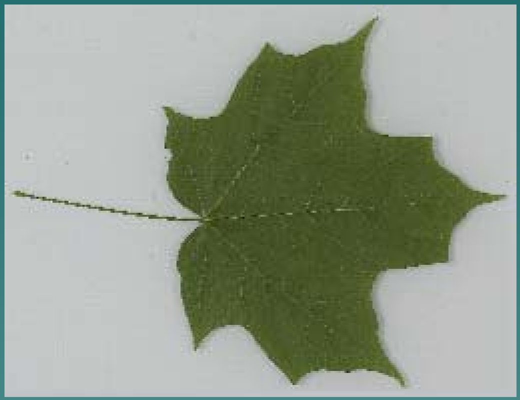
    - Good for detecting similarities between deformable structures

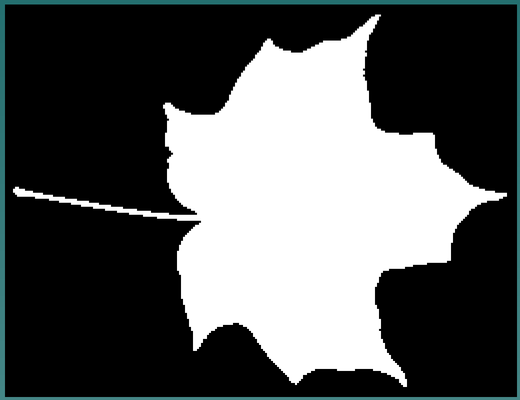
Problem Statement

- Current System:
  - All shape information is compared at a global level, no specific consideration of edge types
- My Project:
  - Incorporate local boundary information to complement existing system

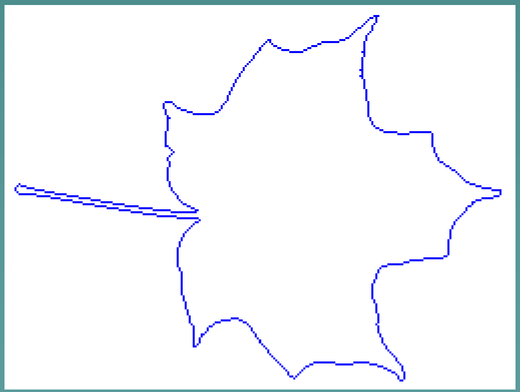
The Algorithm

- Input:





- Capture boundary curve:



The Algorithm

- Each leaf:
  - Vector of ~2000 2-D points
  - Treat as 1-D complex points

The Algorithm: Wavelets

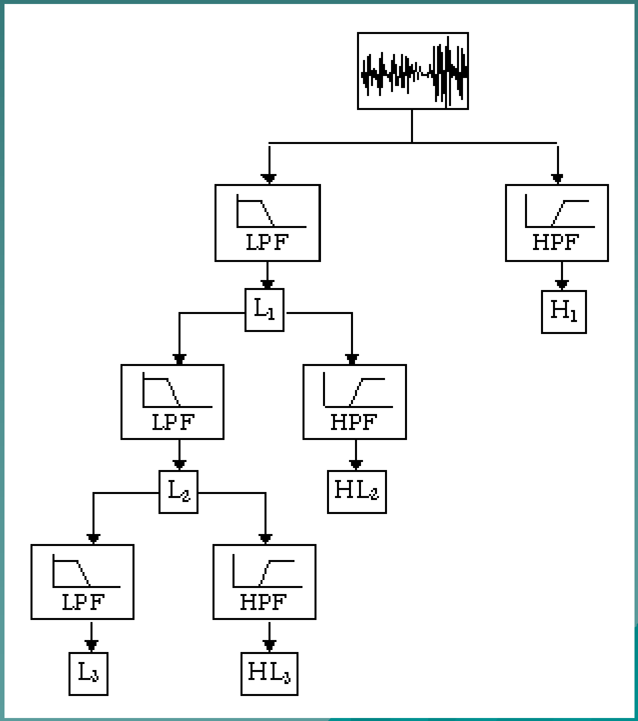
- Discrete wavelet transform
  - In: vector of points
  - Out: two vectors, each half original length

"approximation coefficients"

= low pass

"detail coefficients"

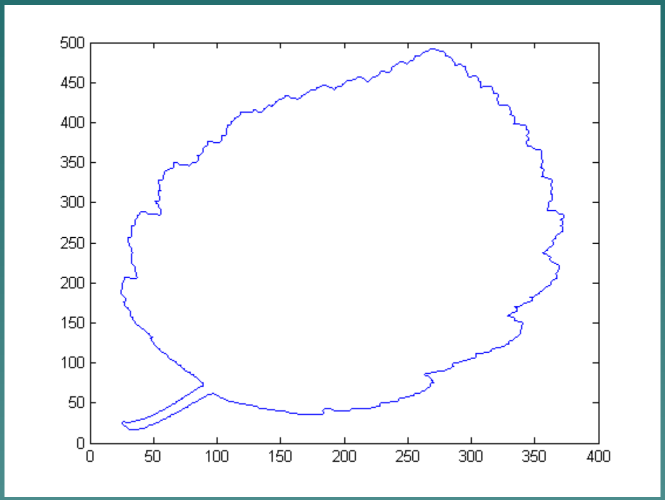
= high pass



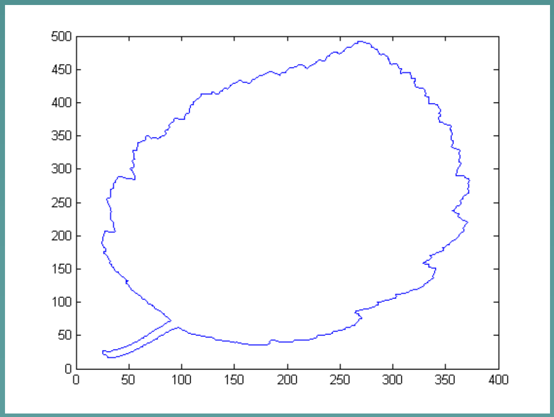
The Algorithm: Wavelets

- Apply repeatedly
  - → hierarchy of scales

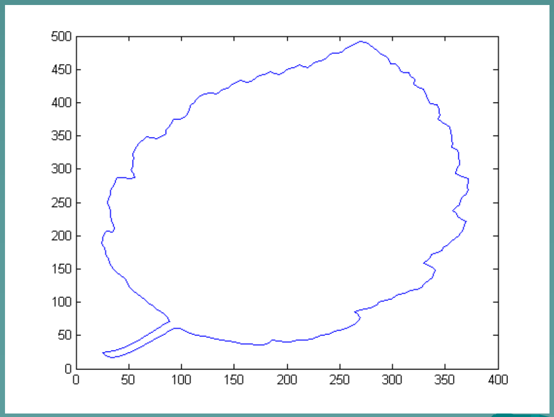
Original



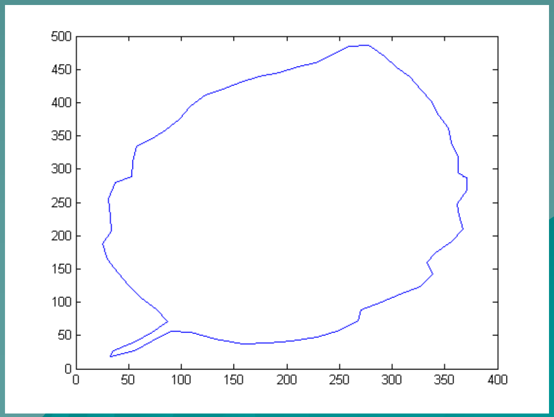
Scale 1



Scale 3



Scale 5



The Algorithm: Wavelets

- Goal: classify by local information

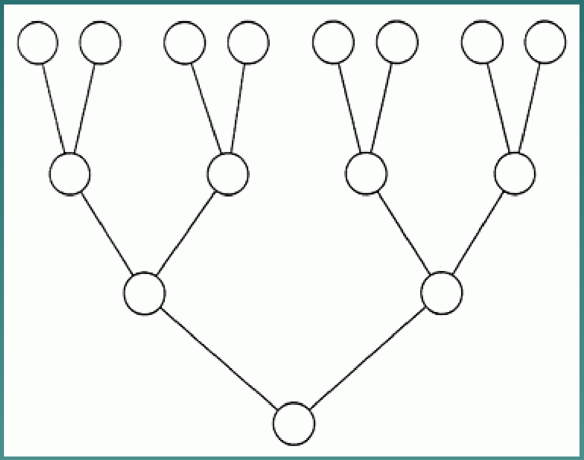
  → Use only detail coefficients

- Want: n-D vector for each boundary point (where n = # scales)

  → More work to get all points at all scales

The Algorithm: Wavelets

- Each coarser scale has ½ as many detail coefficients as the previous
  - → Scale n has $(½)^n$ as many data points as original boundary



- Must calculate (sub-sampling) wavelet decomposition with boundary vector starting at each of the first 2^n points for complete information

The Algorithm

- Now forget about leaves:
  - Data is ~2000 n-D points

- Combine data for all leaves:
  - #leaves x ~2000 n-D points

- Classify!

The Algorithm: Clustering

- Goal: get unique distribution of coefficients for each leaf species

- Direct method:
  Histograms of each scale separately

  – Sort coeffs into ascending, equally spaced intervals
  – Count number of coeffs in each interval, divide by # coeffs
    → distribution

The Algorithm: Clustering

- Considering each scale separately did NOT provide enough information to distinguish between leaves

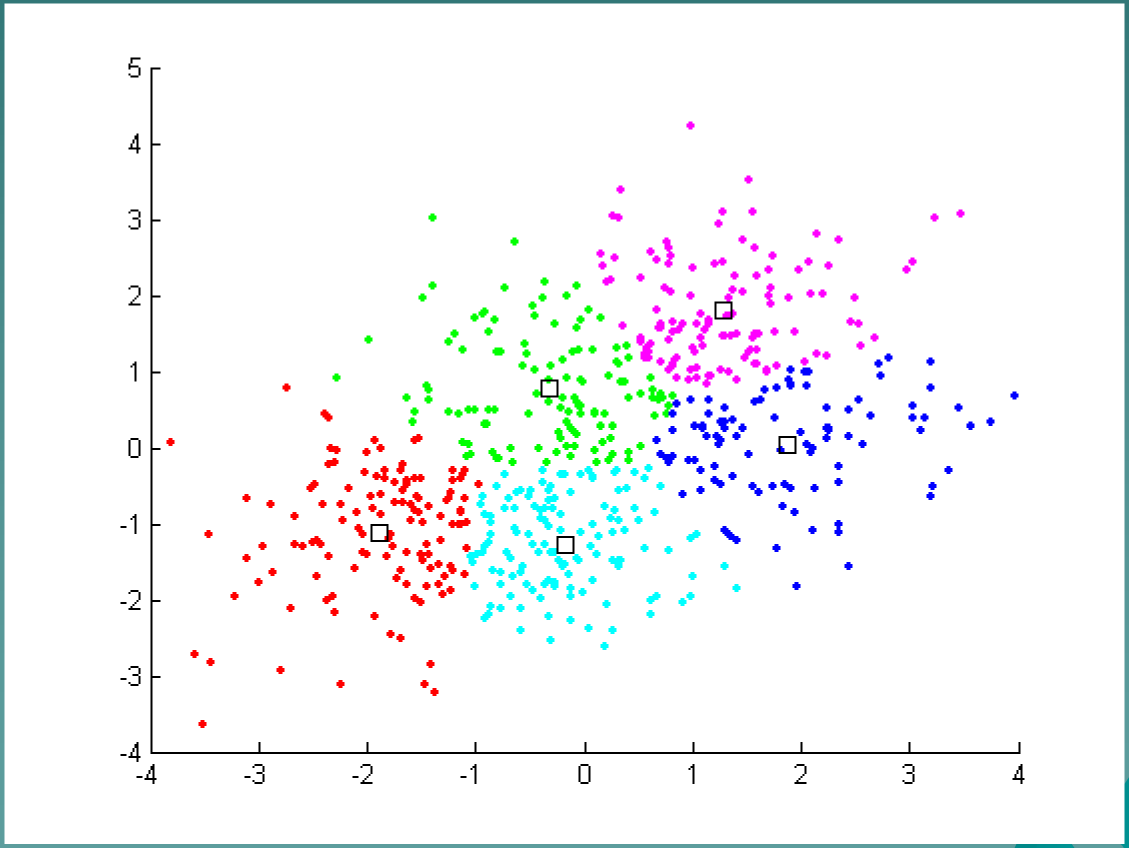
- Use full n-D vector!

The Algorithm: Clustering

- K-Means Clustering:
  - Given lots of data points, choose k of them at random
  - Assign every point to its nearest chosen point → k clusters
  - Find the mean value of each cluster, these are the k new points
  - Iterate
  - End with k "cluster centers"

The Algorithm: Clustering

- K-Means Clustering



Example of final clustering in 2-D

The Algorithm: Clustering

- Cluster 8000*2000 n-D data points into 36 clusters

- For individual leaf, can find distribution of its 2000 points over the 36 clusters
  - → Distribution represents leaf

The Algorithm

- Classification:
  - Get distribution of new leaf over the 36 cluster centers
  - Compare this distribution to that of all leaves in the system (chi-squared test)
  - Smallest difference in distribution is closest match = the best guess at leaf species!
  - In practice, return 10 closest matches

Some Details

- How to choose the wavelet basis?
  - For our purposes, does not make much difference

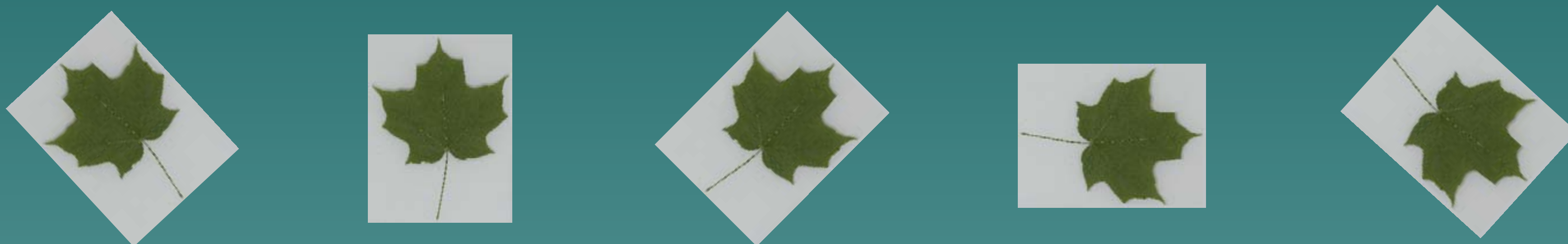
    → Daubechies 2

Some Details

- How to chose n, the number of scales?
  - Trial cases: n = 5
  - Update empirically

Some Details

- Requirement: Rotation Invariance

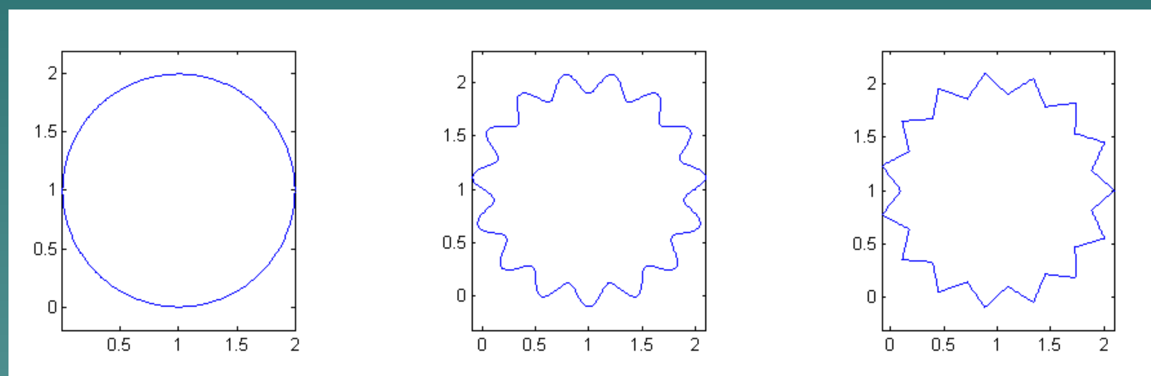


→ Before clustering, for each point:

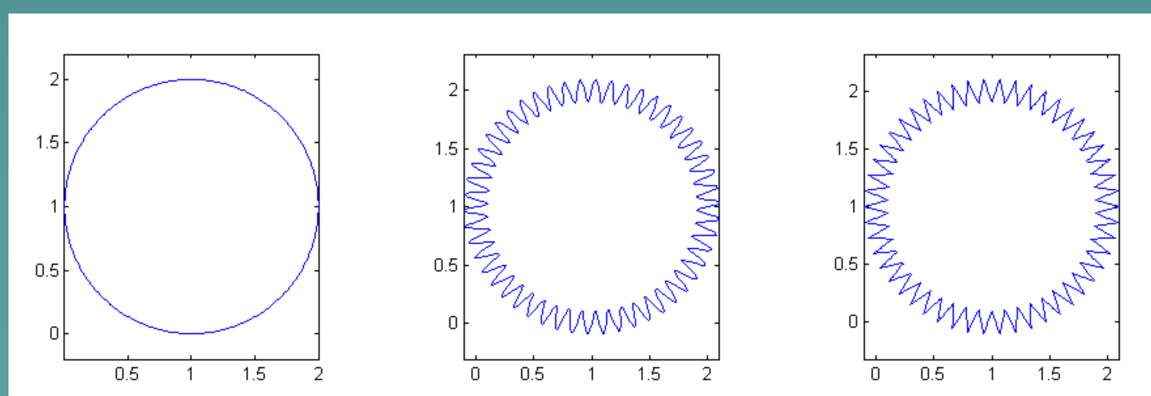
- Rotate coarsest scale to lie strictly on the x-axis
- Rotate each entry of point vector by this same angle
- Effectively reduces the degrees of freedom by 1

Validation

- Simple test cases:

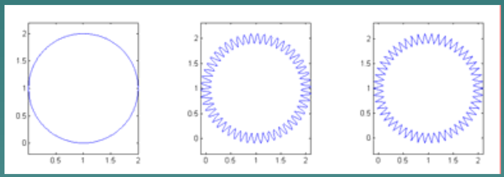


circle | circle + sine curve | circle with sharp points

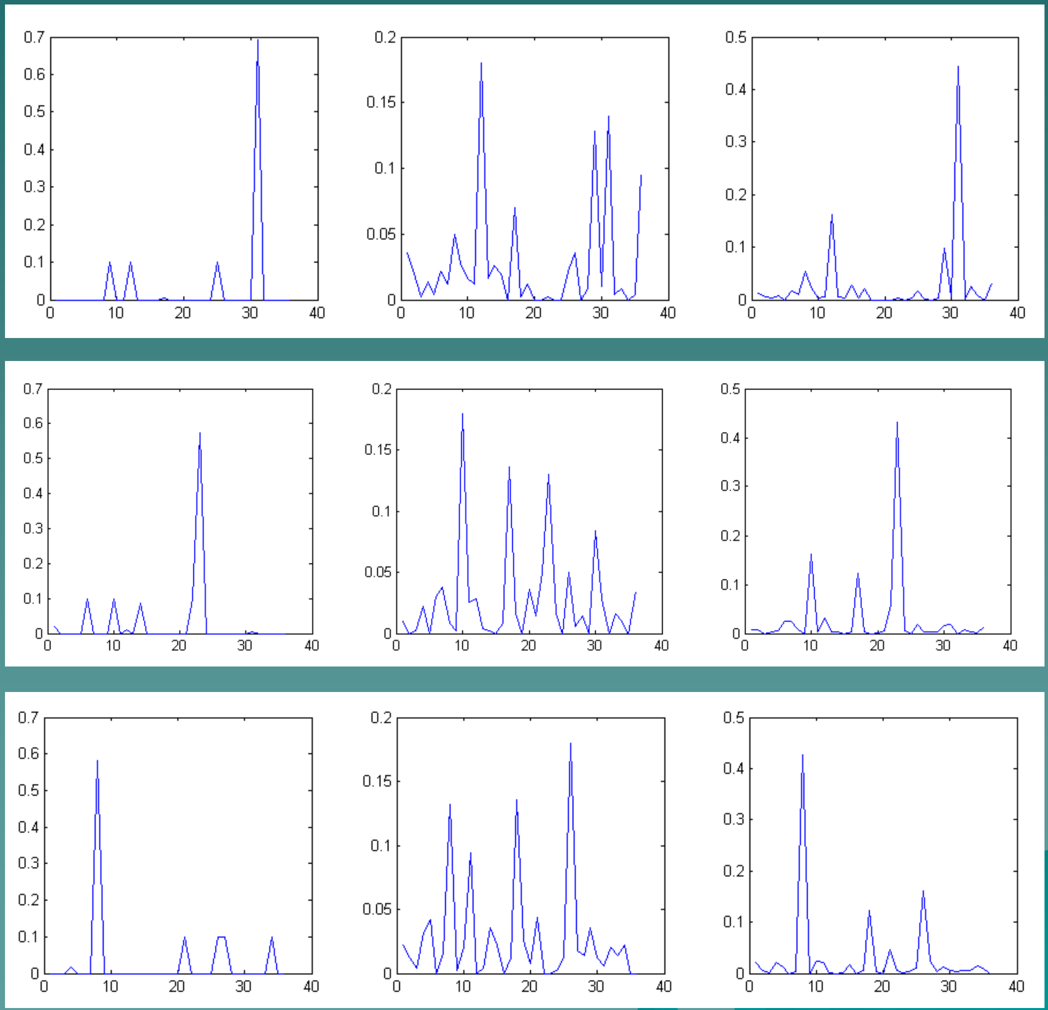


Validation

- Histograms:



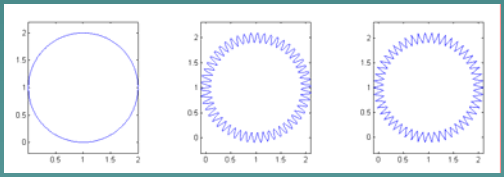
3 runs →
3 sets of cluster centers



Validation

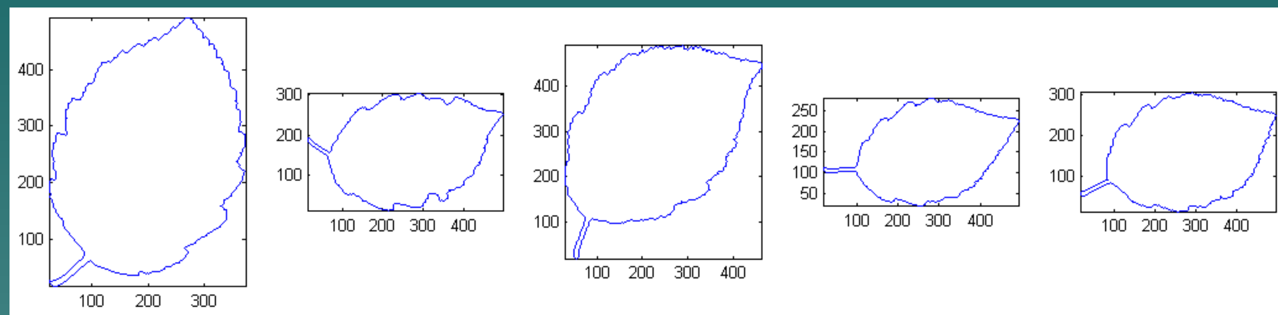
- Chi-Squared Distances between distributions over several runs

D(1,2)	D(1,3)	D(2,3)
0.2435	0.0546	0.1038
0.2048	0.0514	0.1012
0.2219	0.0650	0.1059
0.2068	0.0504	0.1015
0.2745	0.0680	0.0905
0.2575	0.0552	0.0976
0.2126	0.0651	0.1028
0.2629	0.0591	0.0997
0.2276	0.0686	0.1049
0.2679	0.0668	0.0942
0.2604	0.0599	0.0944

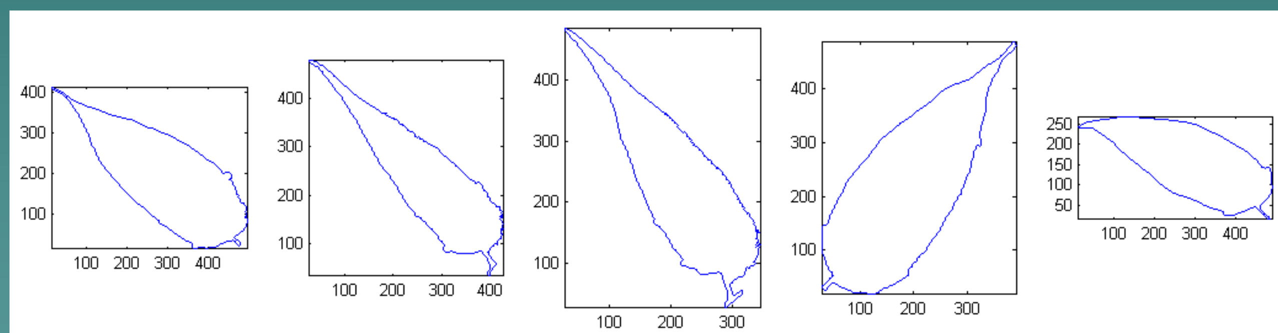


Test Results

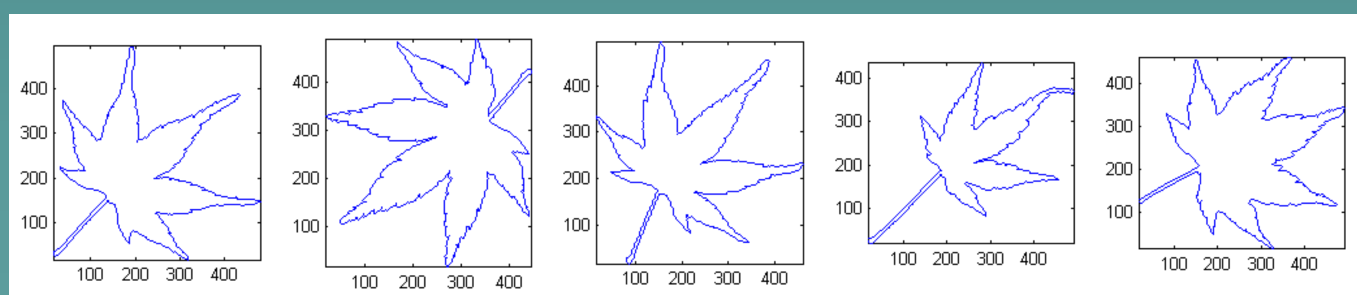
Leaf 1



Leaf 2



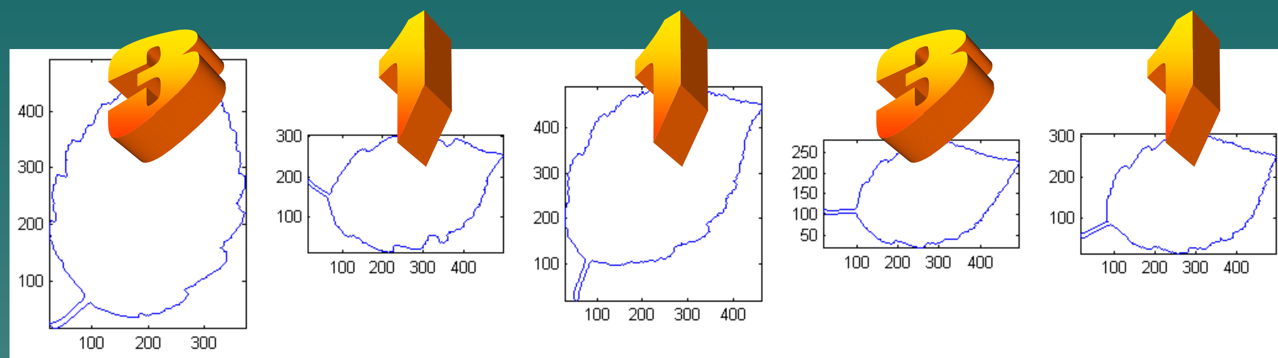
Leaf 3



Test Results

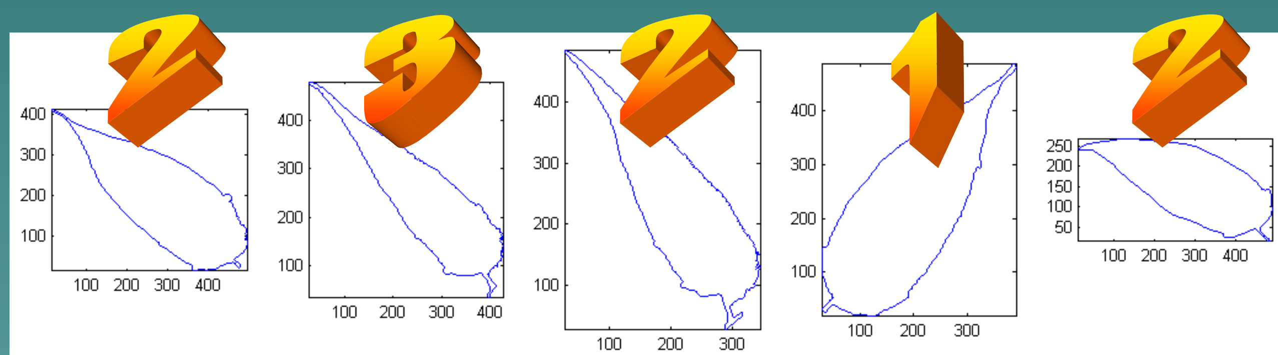
Leaf 1

60%



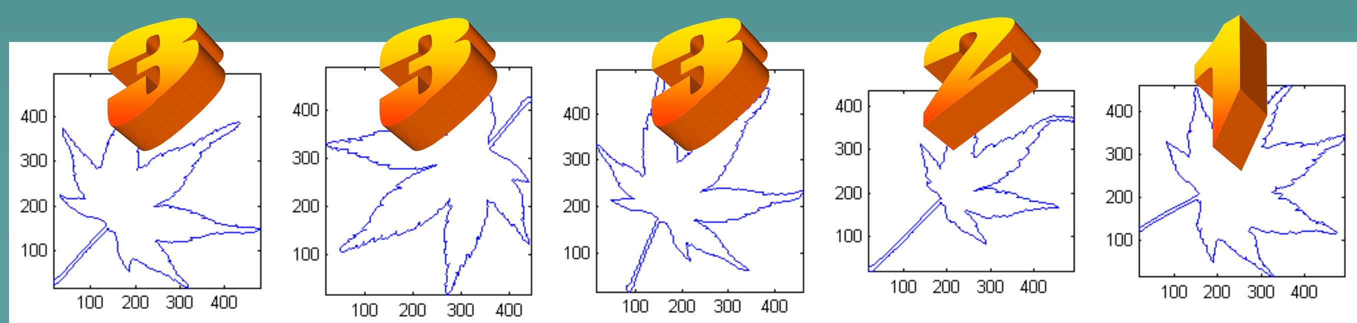
Leaf 2

60%



Leaf 3

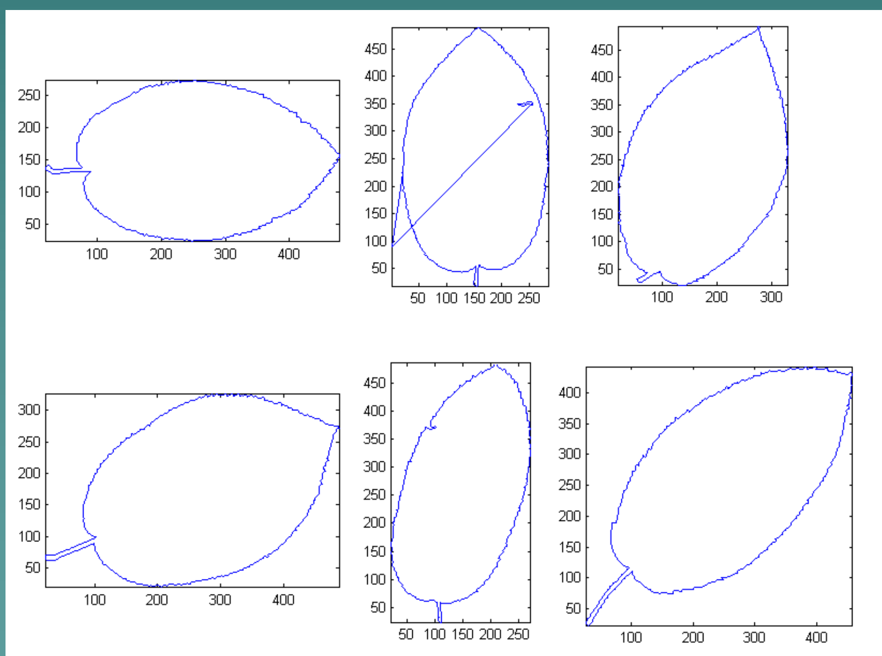
60%



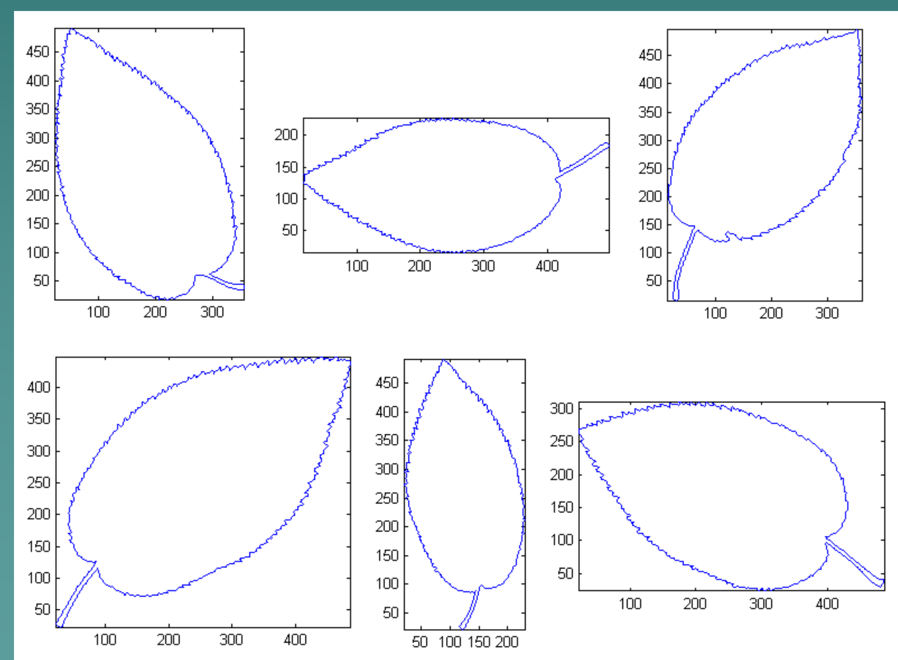
Test Results

- Cases the current system classifies poorly

Leaf 1



Leaf 2



Test Results

	leaf 1	*leaf 2*
◆ #scales = 1:	33%	50%
◆ #scales = 2:	83%	67%
◆ #scales = 3:	100%	100%
◆ #scales = 4:	67%	87%
◆ #scales = 5:	50%	67%

Future Work

- Wavelets: (January)
  - Test on larger data set
  - Experiment with variables
  - Combine with current system
    - Combine wavelet distance with Inner-Distance
    - Test new classification

Future Work

- Beyond wavelets: (February-March)
  - Extend Inner-Distance algorithm over a hierarchy of scales
  - Compare/combine methods
- Final report (April-May)

References

- Gaurav Agarwal, Haibin Ling, David Jacobs, Sameer Shirdhonkar, W. John Kress, Rusty Russell, Peter Belhumeur, Nandan Dixit, Steve Feiner, Dhruv Mahajan, Kalyan Sunkavalli, Ravi Ramamoorthi, Sean White. "First Steps Toward an Electronic Field Guide for Plants". Taxon, vol. 55, no. 3, Aug. 2006.
- Cene C.-H. Chuang, C.-C. Jay Kuo. "Wavelet Descriptor of Planar Curves: Theory and Applications". IEEE Transactions of Image Processing, Vol. 5, No. 1, January 1996.
- Ingrid Daubeshies. "Ten Lectures on Wavelets". Society for Industrial and Applied Mathematicians, Philadelphia, PA, 1992.
- Pedro F. Felzenszwalb, Jushua D. Schwartz. "Hierarchical Matching of Deformable Shapes". IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- Jitendra Malik, Serge Belongie, Thomas Leung, Jainbo Shi. "Contour and Texture Analysis for Image Segmentation". International Journal of Computer Vision, vol. 34, no. 1, July 2001.
- Stephane Mallat. "A Wavelet Tour of Signal Processing". Academic Press, Chestnut Hill, Massachusetts, 1999.