Leaf Shape Project
(automated species clusterization)

CMPT400, University of Saskatchewan
Yuqing Tan (Beatrice)

Supervised by: Prof. Ian Stavness

Guidelines

- I. Motivation
- II. Problem statement
- III. Approaches
- IV. Results
- V. Conclusions

I. Motivation

- "Leafsnap: An Electronic Field Guide“ (http://leafsnap.com/)

Home | Species | Dataset | About

Northeast US | New York | Washington D.C. | Canada

Leaf	Flower	Fruit	Common Name	Scientific Name
			Allegheny Serviceberry	*Amelanchier laevis*
			American Beech	*Fagus grandifolia*
			American Bladdernut	*Staphylea trifolia*
			American Chestnut	*Castanea dentata*
				

(a) Thumbnails of all 184 tree species from the Northeastern United States

II. Problem statement

- My CMPT400 project simulates the shape recognition process of the Leafsnap project, starting with segmented images.
- Work flow:
  - Choose the species
  - Extract features (using Matlab)
  - Run clustering algorithm (using Weka)
- Measurement criteria:
  - Grouping
  - Distinguishing
  - Time/space complexity

III. Approaches

- Choose the species
  - Download the dataset
  - Select species (fine segmented, distinguishable)
- Extract features (using Matlab)
  - 3 models
- Run clustering algorithm (using Weka)
  - Study how clustering works
  - Choose an algorithm → K-Means

III. Approaches

- Extract features (using Matlab)
- 1st model: 3-meta-properties model
- 3 features derived from axis lengths, area and perimeter.
  - Oblong: (short axis) / (long axis)
    range: (0,1)
    value: smaller → bigger
    shape: more oblong → more round
  - Fluffiness: area / ((short axis) * (long axis))
    range: (0,1)
    value: smaller → bigger
    shape: more fluffy → more fulfilled
  - Jaggedness: area * pi * 4 / (perimeter^2)
    range: (0,1)
    value: smaller → bigger
    shape: more jagged → more smooth

III. Approaches

- Extract features (using Matlab)
- 2nd model: round-scanning model
- detecting stem/symmetry and scanning boundary point positions around (one feature with multiple values).

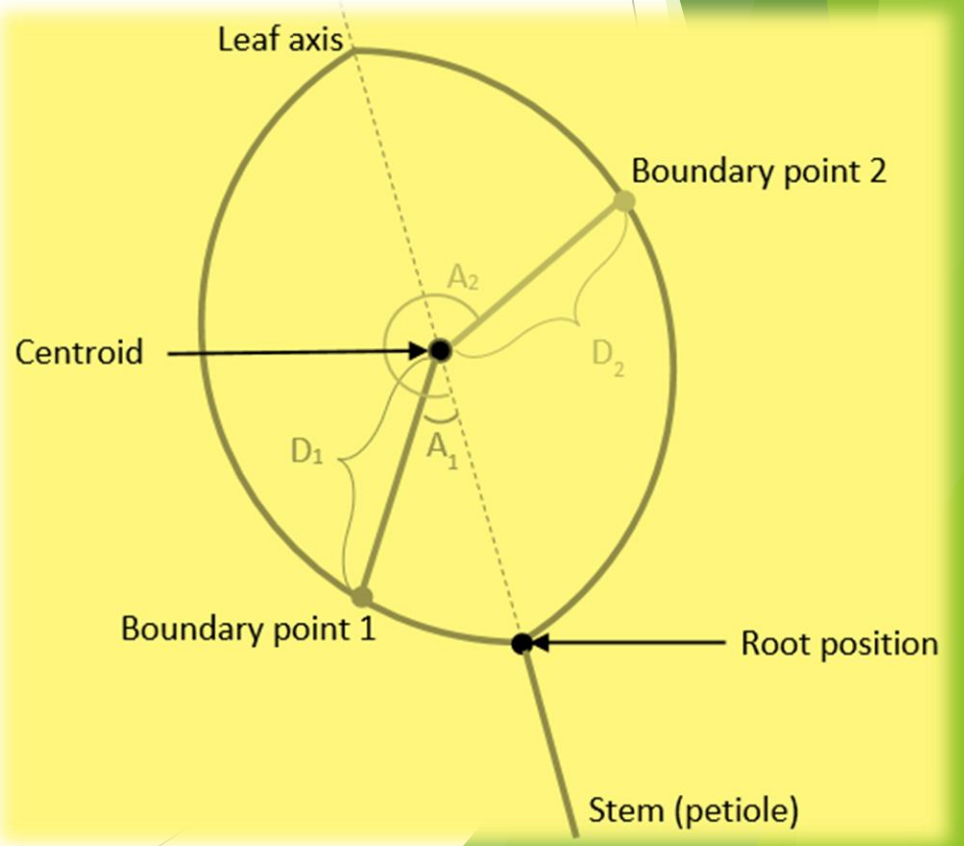
(a) The algorithm to detect stem (as sharp clip shape) if it appears in the picture

(b) The algorithm to find symmetry axis (leaf axis)

(c) Collection of fixed amount of data (D, A) from each image starting from leaf stem

D = the distance from centroid to boundary point (in pixels), divided by average radius (in pixels)

A = the direction relative to leaf axis (in degrees)



III. Approaches

- Extract features (using Matlab)
- 3rd model: concave-triangle model
- 10 features from doing arithmetic on concave shapes.

Get the following values between any two adjacent points in convex set.

(a) **number of maxima**: how many trivial jags it contains

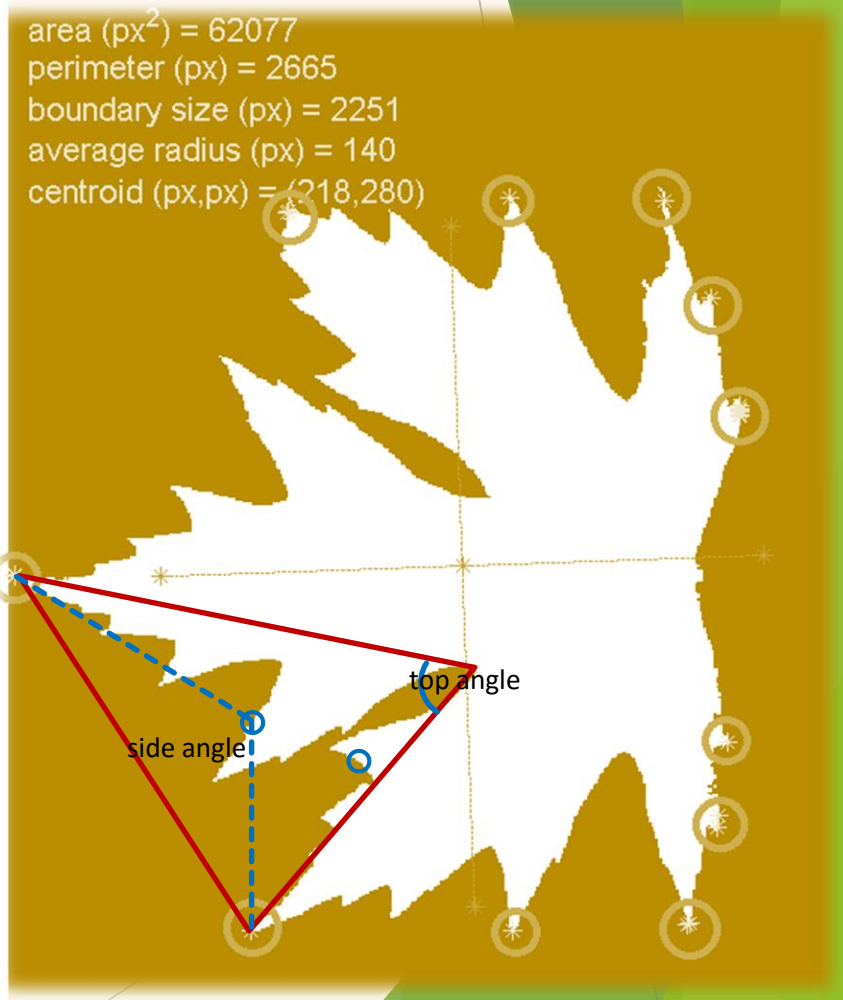
(b) (**triangle height**) / (**average radius of the shape**): how deep the jag is

(c) (**triangle height**) / (**triangle bottom**): shape of the jag

(d) (**curve length**) / (**triangle bottom**): how much the boundary is folded

(e) **top angle**

(f) **average of two side angles**



IV. Results

Acer palmatum
(Japanese maple)

Acer saccharinum
(Silver maple)

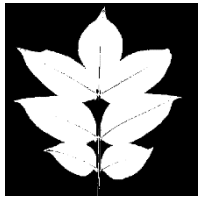
Acer saccharum
(Sugar maple)

Aesculus flava
(Buckeye)

Ailanthus altissima
(Ailanthus)

Betula nigra
(River birch)

Catalpa speciose
(Catalpa)



Evodia daniellii
(Evodia)

Liquidambar
styraciflua
(Sweetgum)

Metasequoia
glyptostroboides
(Dawn redwood)

Pinus peucea
(Balkan pine)

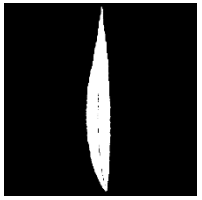
Populous deltoides
(Cottonwood)

Prunus sargentii
(Sargent's cherry)

Ptelea trifoliate
(Wafer ash)

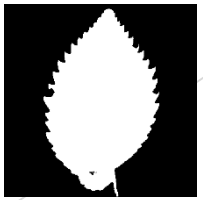
Quercus montana
(Chestnut oak)

Quercus velutina
(Black oak)



Salix babylonica
(Babylon willow)

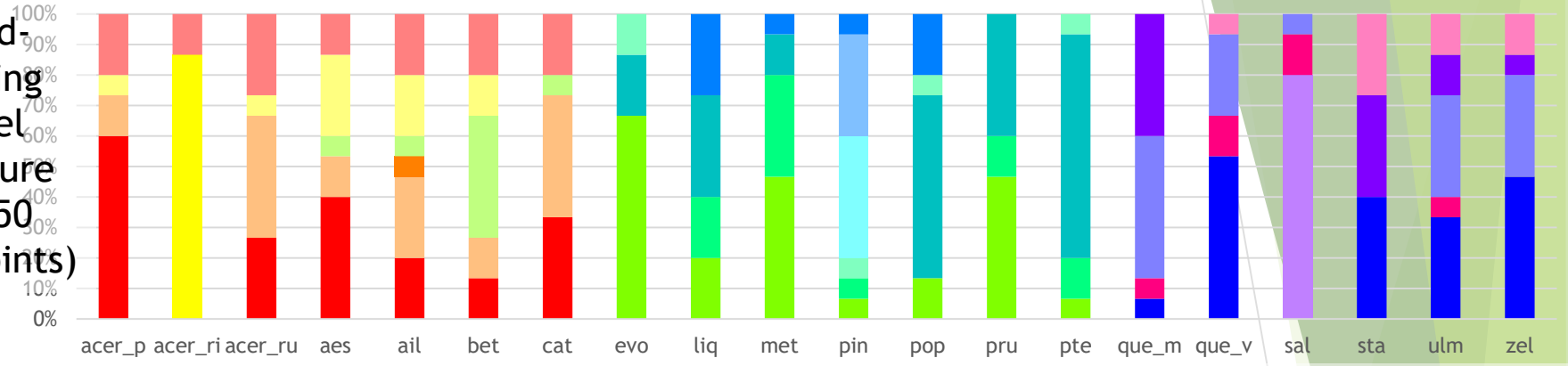
Staphylea trifolia
(American bladdernut)



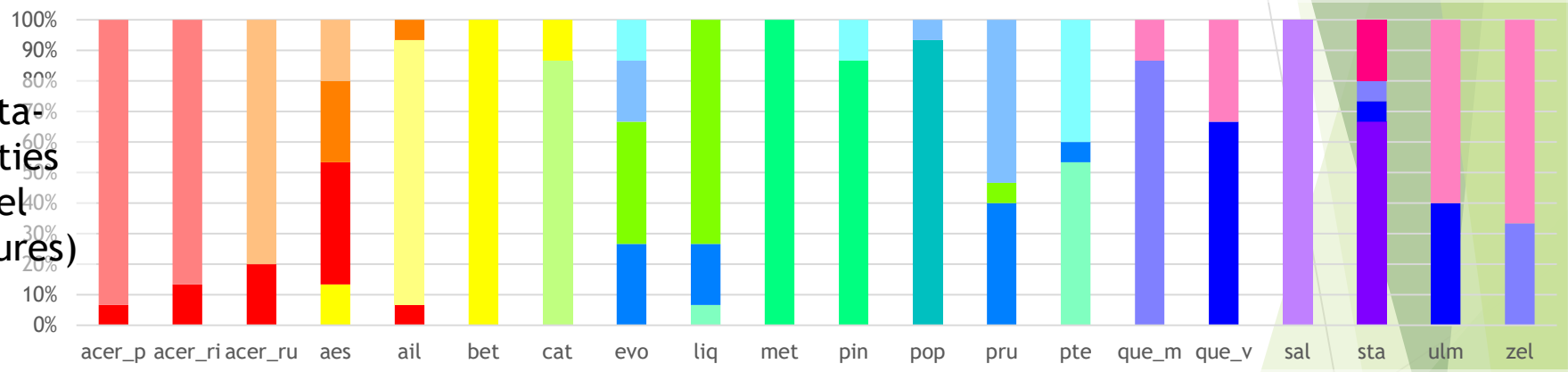
Ulmus Americana
(American elm)

Zelkova serrata
(Japanese elm)

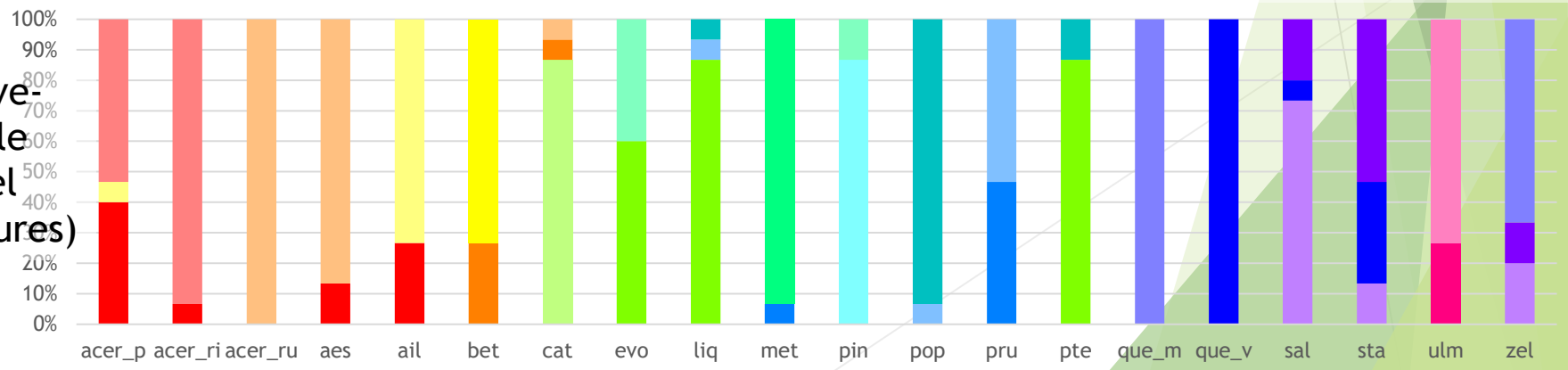
round-scanning Model (1 feature with 50 data points)



3-meta-properties Model (3 features)

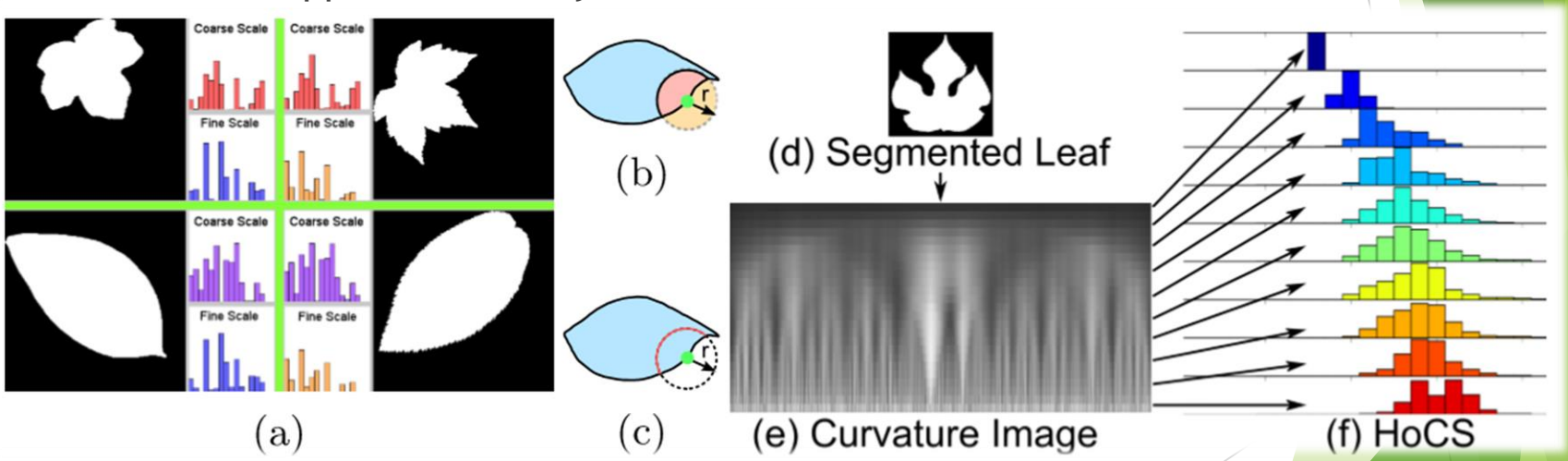


concave-triangle Model (10 features)



V. Conclusions

- Choose the species
- Feature extraction
  - The number is not important
  - What catches the essence
  - Possible improvements
  - Other approaches to try



- Clustering
  - Distance calculation
  - Design my own algorithm

Leave this page blank for questions