

Mining in Forestry Images

Mufida, Yue, Inkyo, Li

12/12/00

Cmput 695 project

1

Background

- 10¹¹ trees in Canada
- What we need for each tree:
 - Tree coordination, Tree species, Tree height, Crown shape / diameter, Tree volume
- Related Research Survey
 - Image Based Approach (Gougeon, Larsen)
 - Graphic Model Based Approach (Pollock)

12/12/00

Cmput 695 project

2

What We want to do?

- Pre-process: Use the image Based Approach method, to find tree apex candidates.
- Use Clustering Methods, to get some meaningful results.
 - AutoClass, SOM, K-Means
- Compare these clustering results, to see what happens...

12/12/00

Cmput 695 project

3

Forest Image Example



12/12/00

Cmput 695 project

4

About Pre-processing:

- To find the local maxima of the image data
 - Smooth the image data by kernel matrix
 - Transform to grey image by only taking the intensity component from its HIS space
 - Establish the histogram
 - Get tree apex candidates from the brightest part of the histogram
 - Use greedy algorithm to accept the tree apices

12/12/00

Cmput 695 project

5

About Pre-processing, Cont.

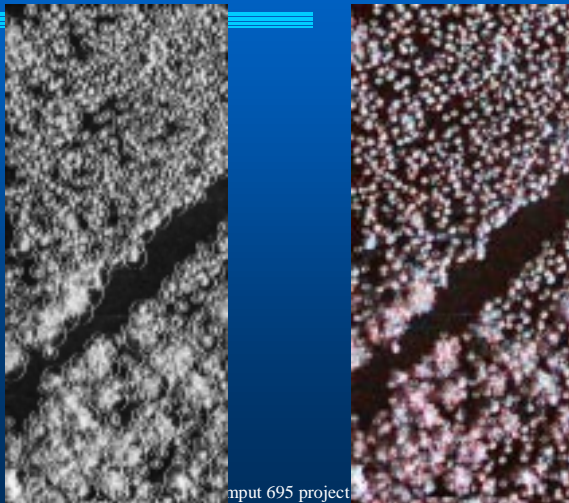
- To get distinguishable attributes for tree apex (two ways we used)
 - The neighbourhood window
 - (3*3 window) * (3 for RGB)
 - The log-polar method

12/12/00

Cmput 695 project

6

An Example of Pre-processing result



12/12/00

Cmput 695 project

7

About Pre-processing, Cont.

- What we learned:
 - Simply Image Based technique doesn't work well → need to combine with the Graphic based model



Figure 1: Model tree and in the right part, template with optimal bounding ellipse.

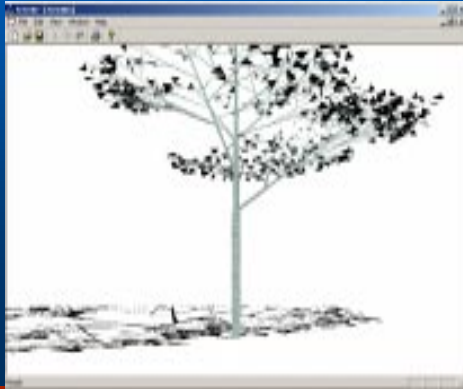
12/12/00

Reproduced from [Larsen97]. Cmput 695 project

8

About Pre-processing, Cont.

- We plan to use L-System to generate more realistic tree model



12/12/00

9

Autoclass

- **AutoClass** is an unsupervised Bayesian classification system that seeks a maximum posterior probability classification.
 - **AutoClass III** -- An unsupervised Bayesian classification program in Lisp
 - **AutoClass X** -- An experimental extension of AutoClass III in Lisp
 - **AutoClass C** -- A public domain version of AutoClass III, with some improvements from AutoClass X, implemented in the C language.

12/12/00

Cmput 695 project

10

Autoclass (Cont.)

- **Key features:**
 - determines the number of classes automatically;
 - can use mixed discrete and real valued data;
 - can handle missing values;
 - processing time is linear in the amount of the data;
 - data cases have probabilistic class membership;
 - allows correlation between attributes within a class; and
 - generates reports describing the classes found.

12/12/00

Cmput 695 project

11

Autoclass (Cont.)

- **Limitations**
 - AutoClass C is limited by memory requirements
 - For very large data sets, you may well find that even if you can handle the data, the processing time is excessive. If that is the case, it may be worthwhile to try class generation on random subsets of the data set. **This should pick out the major classes, although it will miss small ones**

12/12/00

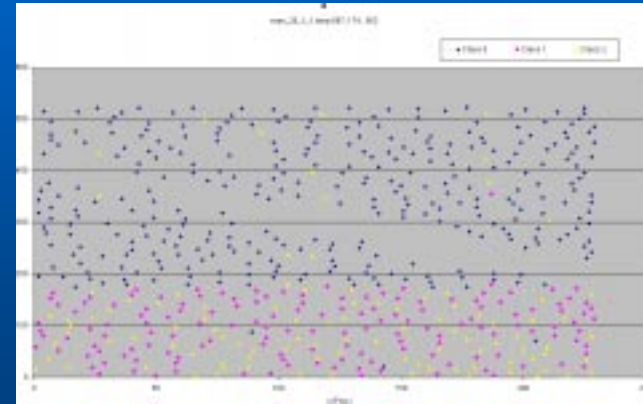
Cmput 695 project

12

Autoclass (Cont.)

- AutoClass has been used and tested on many data sets, both within NASA and by industry, academia and other agencies. These applications typically find surprising classifications that show patterns in the data unknown to the user.

Example of Autoclass Result



Introduction to SOM

- A kind of neural network clustering algorithm
- Mapping from a high dimensional data space to a two dimensional data space
- Each node on SOM map has a reference vector in the input data space
- BMU

Procedure of mapping

- Initialize the SOM map
 - random value for each node
 - bubble or hexagonal for lattice type
- Map training
 - ordering phase
 - fine-tuning phase
 - radius & learning rate

Visualizing the SOM map

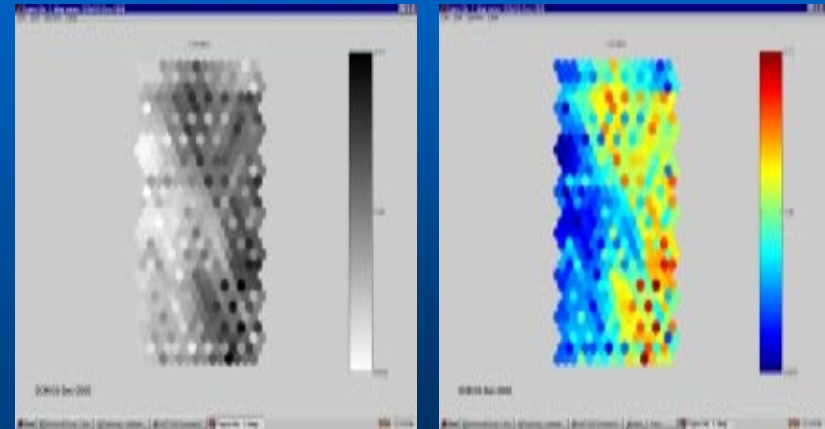
- U-matrix map
- Black & White Vs Color visualization
- We can see the clusters from the picture

12/12/00

Cmput 695 project

17

The U-matrix graph



12/12/00

Cmput 695 project

18

Clustering algorithm.

- K-means clustering algorithm using on the SOM map find the clusters.
- Finding the BMUs for each data sample of the original dataset.
- Summary the cluster that each data sample belongs to.

12/12/00

Cmput 695 project

19

What's CViz

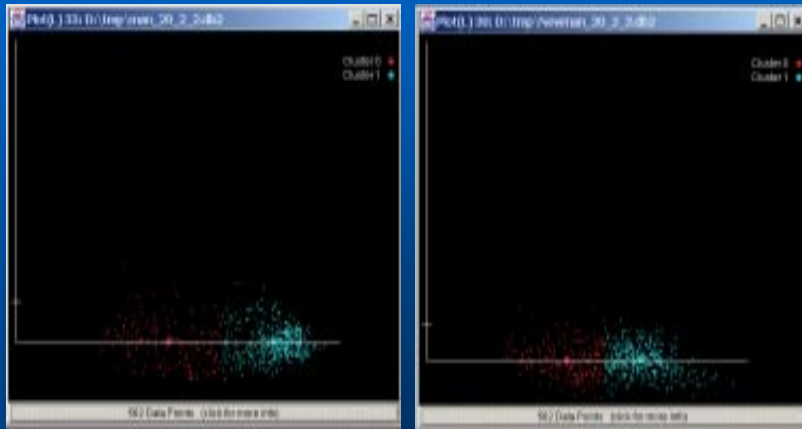
- CViz is a visualization tool for clustering and analysis of multi-dimensional data sets.
- CViz is most valuable when applied to situations where little or no information is known about the relationship between attributes and class.
- Cviz utilizes a K-means clustering algorithm to find interesting concepts in the data.
- Cviz takes in flat file data in Comma Separated Value format.

12/12/00

Cmput 695 project

20

2-NUM-CLUSTER

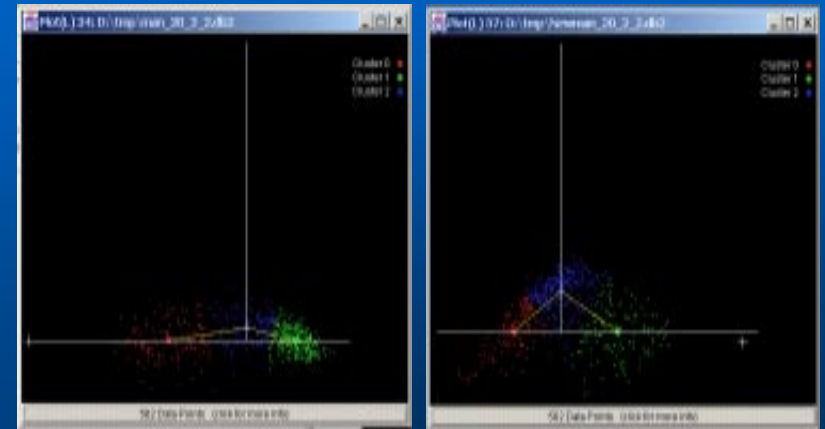


12/12/00

Cmput 695 project

21

3-NUM-CLUSTER

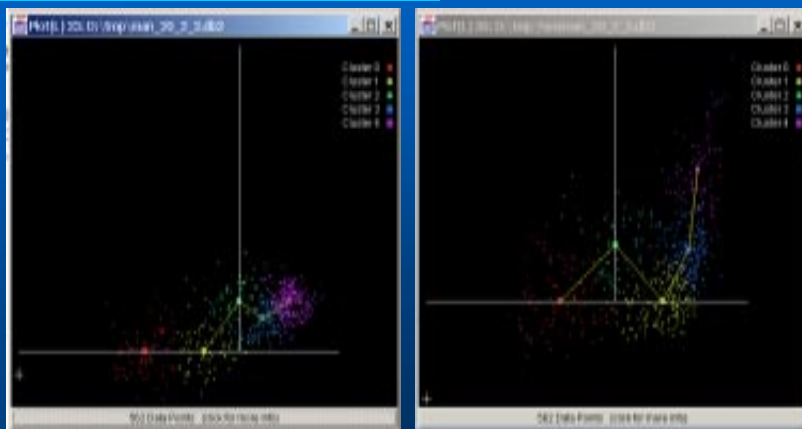


12/12/00

Cmput 695 project

22

5-NUM-CLUSTER



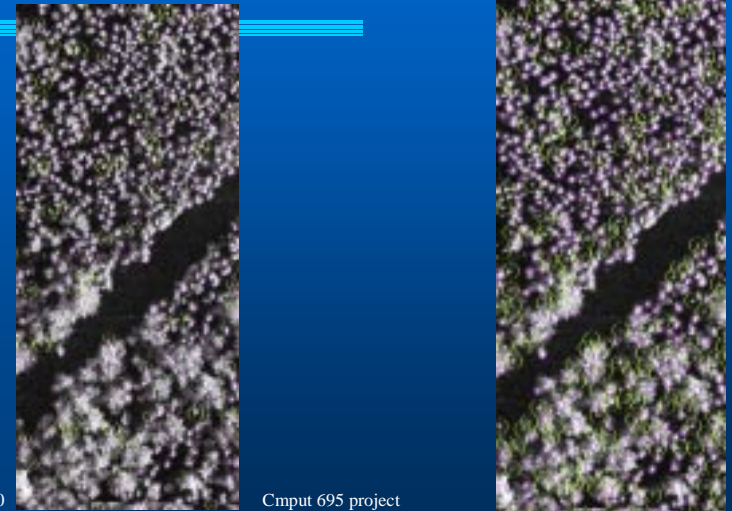
12/12/00

Cmput 695 project

23

Viz the results of Autoclass

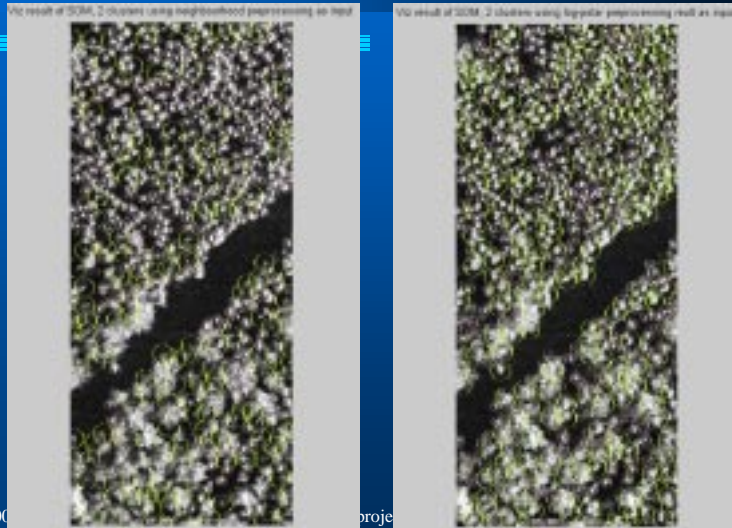
Neighborhood method, 4 clusters Log-Polar method, 2 clusters



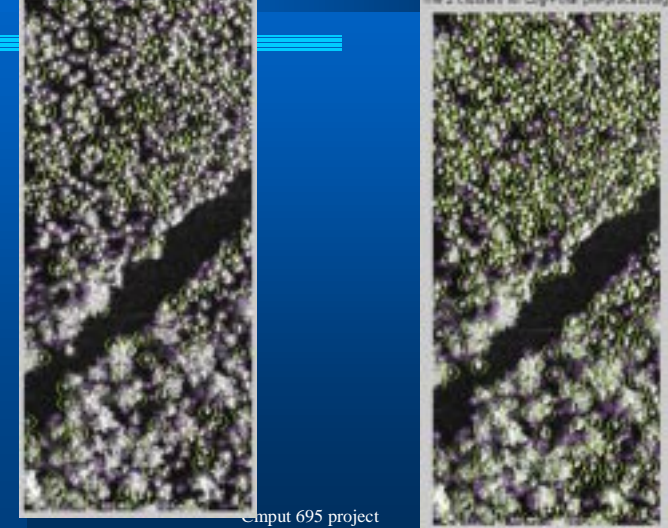
12/12/00

Cmput 695 project

Viz the results of SOM



Viz the results of K-Means



What we learned from this project

- Pre-processing is hard but crucial
- Fairly comparing among different algorithms is hard
 - Different parameter set
 - Different characteristic profile
 - Autoclass need less care, most CPU time
 - SOM has its own interpretation mechanism, still hard to interpret the result
 - K-Means still hold its stand for its simple mechanism, easy to use & acceptable results