Fuzzy Sets and Fuzzy Techniques
VT 2010

Exercise 1: Fuzzy Thresholding and Clustering

Formality

- We would prefer if you work in groups of two persons. That way, you can discuss with each other and solve difficulties more easily.
- You should prepare a written report, one for each group. Remember when you write the report that it should show that you have learned about the methods you have worked with. We are not that interested in how well you solved the particular problem, as long as you describe precisely what you have understood.
- Hand in the report to us manually, or send it by e-mail to milan@cb.uu.se and/or joakim@cb.uu.se.
- Deadline: Tuesday, 2 March 2010.

Before you begin, study the lecture notes and articles suggested for reading:

[1] S.K. Pal, and A.Rosenfeld. Image enhancement and thresholding by optimization of fuzzy compactness. *Pattern Recognition Letters*, 7(2):77-86, 1988.
[2] L.K. Huang, and M.J.J. Wang. Image thresholding by minimizing the measures of fuzziness. *Pattern Recognition*, 28(1):41-51, 1995.
[3] R.L. Cannon, J.V. Dave, and J.C. Bezdek. Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-8(2):248–255, 1986.

Fuzzy Thresholding

- Implement fuzzy thresholding by minimization of fuzziness, i.e., Algorithm 1 of Pal and Rosenfeld [1].
- Test the different measures of fuzziness, using different bandwidths, on at least 2 images of your choice.
- Implement fuzzy thresholding according to Huang and Wang [2].
- Compare and comment the methods and results, give pros and cons.
- Compare with Otsu crisp thresholding (Matlab function *graythresh*).

Unsupervised Fuzzy Clustering

The fuzzy c-means clustering (FCM) algorithm tries to partition numerical data into clusters. The belongingness of a data point to a specific cluster is given by the membership value of the data point to that cluster. The membership value is calculated by minimization of a FCM function which searches for the belongingness that gives the least error. The function needs approximate cluster centres, as well as a metric for membership evaluation as input, e.g., the Euclidean distance. The minimization is an iterative process where new cluster centres are computed as weighted averages of all data points, where the membership values are the

weights. The stop criterion is when the membership values have changed very little in comparison with the previous iteration.

Study the code fcm.c and compare with the information given in the manuscript [3]. No programming of your own is required, but of course you are free to change parts of the code that you are not satisfied with. However, you need to compile and run the program several times (hint: to compile fcm.c execute the following command **cc fcm.c -o fcm -lm**). As input to the program you give an initial guess of each cluster.

The task concerns classification of three different tree species. Ground truth is that we have 333 of each kind. But is it possible to classify them accordingly using FCM? The trees were identified in an image with two layers, giving one (R,G) value for each tree. The program reads data from the file clusters.dat, which contains 999 rows of (R,G)-data for each tree. The first 333 lines are for cluster 1, lines 334–666 are for cluster 2, and the rest is for cluster 3, i.e., this is their true classes. The same data are also available as three separate files: species1.dat, species2.dat, species3.dat. Since this is simulated data, we have the information that the three clusters are normal distributions:

	mean	std.dev.	N
cluster 1			
R-band	100.0	10.0	333
G-band	100.0	10.0	333
cluster 2			
R-band	120.0	10.0	333
G-band	120.0	10.0	333
cluster 3			
R-band	120.0	15.0	333
G-band	90.0	15.0	333

- Choose initial guesses according to the normal distributions.
- Choose initial guesses, somewhat different from the normal distribution. How robust is the method?
- Choose initial poor guesses, way off the true cluster centres. Will convergence be reached?
- What happens if the same initial guess is given to all cluster centres?

For each choice of initial guesses, plot the cluster (R,G)-data and check the classification result. Report the final stable positions of the cluster centres for each guess, as well as how many iterations are needed before convergence in each case.