

Homework 4
Classification of images using wavelet features
Due: May 1, 2009
(accepted without penalty until 8:30 AM May 11, 2009)

50 points

Your task is to write and test a program in C, C++, Java, Perl, Python or Matlab to train and test a classifier to distinguish two classes of images using Daubechies 4 (db4) wavelets. The program must run on the Andrew Linux servers and should read required input values input from the command line and write required outputs to a text file (it should not use any interactive or graphical user interface). **Output should be limited to the items requested (and should not include prompts of any type).**

The program should parse from the command line the following inputs

- a. two paths (including wildcards) specifying the sets of files in two classes
- b. the level of decomposition desired (integer)
- c. the number of folds of cross-validation desired (integer)
- d. the name of an output file to be created for the features
- e. the name of an output file to be created for the final confusion matrix

The program should calculate the db4 decomposition up to the level specified and then calculate features by taking the average energy at each level. The feature values should be saved into a comma-separated text file with the name specified on the command line.

The images should be divided into training and test sets based on the number of cross-validation trials specified. For each trial, the features from the training images should be used to train a classifier of your choice (e.g., LDA, neural network, SVM, etc.) and then its performance should be evaluated using the test data. The results from all trials should then be averaged to create a confusion matrix that should be saved into a comma-separated text file with the name specified on the command line.

You should use your program to analyze the two sets of images in the zip file on the homework page (16 bit images of endosomes and lysosomes in PNG format. Use values of 1 to 10 for the decomposition level. Write a description of your results (include the confusion matrices) for various levels of decomposition and the conclusions that you draw from them.

Submit (1) the source code for your programs, (2) a description of the test procedures you used to verify that the program operates properly, and (3) the description of your results on the test images.

Extra Credit

You may choose to also calculate average amplitude for each basic frequency of a 2D FFT as additional features (the average amplitude in “rings” around the center of the 2D FFT). If you chose this option, your grade will be multiplied by 1.25 (e.g., a score of 44/50 will become 55/50).