05 Shape and Texture features for Classification

Assignment 5, 30 points in total; 20% of the total score for the practical part.

In this assignment you are analysing embryos of a mutant of the plant *Arabidopsis thaliana*. In the set of images provided, there are about three classes that express the mutant form differently. These classes can be expressed by their size and shape. The embryos are 7 days old and the blue colour represents the expression of one particular gene.
The general idea of this assignment is to evaluate shape and texture features and assess the combination of features that performs well in a classification.

Part 5.1

The first thing accomplish is to segment the embryos so that the form can be analysed. Each image contains one embryo which is, in general, centralized in the image. As we deal with a 3D structure, the objects are not always completely in focus. Nevertheless, also out-of-focus parts should be included in the segmentation – in some cases an extra image of the same embryo is provided (suffix "b"). You will need to derive a mask for each of the images to be able to extract the "plant embryo" for further analysis.

1. (3) Develop and apply a segmentation technique to prepare the embryos for further processing. Report on the results by briefly explaining the algorithm and pre-processing you have used to realize segmentation.
2. (2) Provide a panel of all original images that are used in the analysis. Make sure that you are not using the original resolution but rescale to a size that is workable for the presentation of these original images. This panel contains 58 images, crop the images and use a 4*15 template for the panel to display all the images. For referencing, keep the original numbering. Try and automate this process from the functions that are available form DIP-image.
3. (2) Subsequently, provide a similar panel for the resulting binary images that you will use for the further analysis. Adapt your small program to accomplish this page of results. These are the masks. The two pages are the reference set for this assignment.

Part 5.2

Once the shapes are available, a measurement strategy is designed. You will need to define the shape and size and the amount of "blue" per shape and per size. The blue represents the expression of one particular gene; in some cases, it is abundantly present while in other the gene expression seems to be absent.

4. (4) Compute the size and shape for the embryos in all images; including the amount of gene-expression (blue), relative and absolute.

5. (2) According to the measurements, try to establish three groups in the plant-embryo's that have a similar response after 7 days of treatment. Argue what thresholds for the values for size and shape you have chosen.

Part 5.3

In the area of the blue colour different textures can be seen. We need to know if texture can play a role in distinguishing the different clusters. Therefore, you are going to analyse that part specifically. For texture compute the standard deviation and the Uniformity in the blue region, i.e. the gene-expression and in the whole embryo.
In addition, one "handcrafted" feature is evaluated: i.e., the Histogram of Oriented Gradient (HOG). You will be pointed to python code in order to compute this feature. For each of your images a HOG feature vector, with two "cell" sizes is created. In order to assess the colour of the berries, the (average) hue is computed. This requires to separate the object from the background and thus apply a segmentation. The mask for the object/embryo can be used for the computation. The same mask can be used to compute a number of shape features. All these features will be used to see how well they can characterize the embryos. For thresholding you might need to do some pre-processing on the images.

6. (4) Per "shape" compute the texture in the blue part as well as in the remaining part. Use the 2 different texture measures as indicated and report these in a table. Plot a histogram of the different values you find.
7. (4) Per "shape" compute the HOG of each the embryos for just the gene-expression part as well as for the whole for the whole embryo. As mentioned, this is done for two different "cell" sizes; i.e. 16 and 64.

Part 5.4

You have developed an initial idea of the number of different groups that are present in this dataset. We should check if the intuitive grouping can be confirmed by a classification over the features that are computed per embryo.

Now we obtain the complementary dataset of all HOG, Texture and Shape features in vectors. We will use a Support Vector Machine (SVM) to classify the images so as to see how well the classes can be separated on the basis of these features. This will be used for training set; separate a part that will be used as the test set. A division of 80%-20% is used for the division of training and test set. Indicate the number of classes that you hope to identify. A python implementation of the SVM will be made available to you by the course administration.

8. (1) Apply the SVM using the HOG feature set – report on the outcome.
9. (1) Apply the SVM using the Texture feature set – report on the outcome
10. (1) Apply the SVM on the Shape feature set – report on the outcome.

11. (1) Apply the SVM using the combined feature sets – report on the outcome.
12. (1) Show the quality of separation through graphs of the classification results.
13. (2) Does the texture classification follow the grouping that you have established earlier? Motivate the answer.
14. (2) Provide an assessment how shape, texture and HOG features can separate the patterns in this dataset.

PS Instructions are provided as a separate file. Use the python libraries that are made available via the course administration.