05 Shape and Texture features for Classification

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

This assignment has four parts. For image pre-processing, please use **DIP-Image (PyDIP)**.

In this assignment you are analysing images of 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 aged 7 days of development, 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 1

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

1. Develop and apply a segmentation technique to prepare the embryos for further processing. Explain the algorithm and any pre-processing steps you have used to compute the segmentation. (3)
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. (2)
3. Subsequently, provide a similar panel for the resulting binary images that you will use for the further analysis. These image panels constitute the reference set for this assignment. (2)

Part 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 size and its shape. As mentioned, the blue colour is the result of an *in-situ* hybridization representing the expression of a particular gene; in some cases, it is abundantly present while in others the gene expression seems to be near absent.

1. Compute size features (area and perimeter) and shape features (such as roundness and solidity) for all embryos in the images. Also, include both the relative (relative to the entire mask) and absolute amounts of gene expression (blue regions) within each embryo. (4)

2. According to the measurements, try to establish three groups in plant embryos that have a similar response after 7 days of treatment. Explain your approach to group the embryos and what thresholds for the values for size and shape you have chosen. (2)

Part 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.

First you will calculate the texture features such as uniformity and standard deviation. In addition, one hand-crafted feature will be evaluated: i.e. the Histogram of Oriented Gradient (HOG). For each of your images, a HOG feature vector with two cell sizes is created. ***Refer to the additional documentation on Brightspace for more information on Histogram of Gradients and code snippets required for computing these (placed under the Assignment).***

1. Per shape, compute the texture (uniformity and standard deviation) in the blue part as well as in the remaining part. Report your values in a table and a histogram. (4)
2. Per shape compute the HOG of each the embryos for just the gene-expression part as well as for the whole embryo. As mentioned, this is done for two different cell sizes; i.e. 16 and 64. Provide the resultant images in the Appendix of the report. (4)

Part 4

You have developed an initial idea of the number of different groups that are present in this dataset. You 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 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.

1. Apply the SVM using the HOG feature set. Report and explain important metrics such as precision, recall and F1 score on the outcome. (1)
2. Apply the SVM using the Texture feature set. Report and explain important metrics such as precision, recall and F1 score on the outcome. (1)
3. Apply the SVM on the Shape feature set. Report and explain important metrics such as precision, recall and F1 score on the outcome. (1)
4. Apply the SVM using the combined feature sets. Report important metrics such as precision, recall and F1 score on the outcome. (1)
5. Show the quality of separation via a confusion matrix of the classification results. (1)

6. Does the texture classification follow the grouping that you have established earlier? Motivate your answer. (2)
7. Provide an assessment of how shape, texture and HOG features can separate the patterns in this dataset. (2)

Please submit your code as a separate .py or .ipynb file. Kindly adhere to the numbering in the document in your assignment. Explain your choices wherever you can. For the report itself make sure that the images in the pdf are scaled down for the size used in the document. It is a control for the size of your submission and should prevent your report from becoming unnecessarily large.