

Document and Content Analysis (SS 2009)

Exercise Sheet 4

to be submitted by E-Mail to faisal.shafait@dfki.de by: 29.06.2009

Use Python for the following exercises.

Exercise 4.1 (Binarization)

Implement Otsu binarization algorithm and apply it on the image: sample.png.

Exercise 4.2 (Page Segmentation)

Implement the run-length smearing algorithm with all four steps:

- Horizontally smear the binary image obtained in the previous task.
- Vertically smear the binary image obtained in the previous task.
- Obtain an image by performing pixel-wise AND operation on the two smeared images.
- Horizontally smear the obtained image to obtain the final bitmap.

Exercise 4.3 (Connected Component Analysis)

Perform connected component analysis on the final image to extract image segments. Draw rectangular bounding boxes around the segments in the original input image.