Spring 2023

Course Assignments for Advanced Visual Data Analysis (TNM098) Spring 2023

Lab Assignment 3

The deadline for this assignment is May 5, 2023 (at 23:59).

Content-based analysis of data

This lab comprises two tasks involving different content analysis and comparison methods.

Task 1 *Image comparison*

Take the zipped data file and extract the 12 images it contains. Create a visual data analysis solution that loads each image (suitable image loader libraries can easily be found online) and analyses it to produce a feature vector. Suitable features might include:

1. Colour content
2. Colour distribution around the central point
3. Colour distribution around several points
4. Luminance distributions around one or several points
5. Edge positions and orientations
6. Anything else that you can think of

Pick one image from the set and use the comparison of the feature vector to rank the other 11 in order of similarity to your chosen image.

Questions of interest: How did you make the feature vector comparison? Did you apply any weighting of features over one another? Why?

Task 2 *Text comparison*

Take the zipped file and extract the 10 text files it contains. Two of the ten files contain plagiarised content from one of the other files. In one it is a single sentence, in the other an entire paragraph.
Create a visual data analysis solution to analyze the 11 texts for textual content and then use comparison to identify the plagiarised sections.
Tips:

1. Remove the punctuation using a linear replacement and convert the sentences into single lines of text. Also, convert everything to upper or lower case.
2. Analyse the files to create a word list (a hashed dictionary) replacing the words with numerical values of the word positions in the list.

3. Linked lists are the ideal mechanism for creating this dictionary
4. Compare the numerical sentence sequences to identify the copied text and the source and destination files for each copied element.

The final step may be quite a lengthy process.

Prepare a short report describing your analysis approach and your proposed solution. Include a description of your solution, a discussion of your results and their quality, the questions of interest, and any additional insights your solution makes possible.

Present your solution to a teaching assistant in the scheduled lab session for Lab 3: **Monday 24 April, 13:15-17:00 in TP4003**. Upload your short report (~2 pages, PDF format) under Submissions Lab 3 in Lisam.