EXERCISES FOR IMAGE PROCESSING I
PROJECT

Due date: 28.01.16 before 12:00h

Topics: Characterization of historical (Asian) manuscripts

Submission: Please send your solutions via email to seppke@informatik.uni-hamburg.de.

Note: Have your system ready for demonstration on January 28!

MANUSCRIPT WRITER IDENTIFICATION

One goal of analyzing historical manuscripts is to decide whether a manuscript was written by the same or by different persons. Since each (scanned) manuscript is unique and of high-resolution, such an analysis is mainly performed in feature space.

For your project, we provide you a total of 12 manuscripts to analyze (6 for each of the two authors). Based on at least one image per author, you shall program two different systems to find out, if two manuscripts have been written by one person or not. To begin with the project, please select one image of the dataset O.8210-S.2717 from the folder "front" and one from the folder "back". If necessary, you may add the other images, too. If you would like to use/experiment with more images, you may use the (publicly available) database of the International Dunhuang Project, e.g.: http://idp.bl.uk/database/oo_scroll_h.a4d?uid=4600447812;bst=1;recnum=10528;index=1;img=1

I LOCAL/FEATURE-BASED ANALYSIS   35 P.

1. **(Manual) Character separation**
    Look carefully at both manuscripts and determine at least four different characters, which are imaged in each of the two manuscripts. Mark all occurrences of the characters in each manuscript, e.g. by a bounding box and the corresponding filename:
    charA = [["fileA.jpg", [0, 0, 100, 80], … [2500, 240, 2650, 330]]
        …
            ["fileB.jpg", [3000, 1000, 3100, 1080], … [200, 2040, 290, 2110]]   ]

2. **Feature selection**
    For local approaches, features are computed separately for each character, and it is up to you to design your system with suitable features.
    Here are some examples (which may or may not be suitable):
    *   Histograms of greyvalues
    *   Histograms of gradient orientations
    *   Mean and variance of greyvalues
    *   Shape parameters
    *   Correlation with fixed set of templates
    *   Principle components etc.

Select some features for further application and discuss, which kind of features may be applicable for your task. Note, that (same) characters may be imaged at slightly different sizes.

3. **Implementation and feature derivation**
Design and implement a system, which computes your selected features for all of the characters you have defined in (I.1).

4. **Comparison of feature applicability and decision function**
Compare the results of your features with respect to the given task. Use the (multi-dimensional) result of all of your (combined) features to develop a decision, discriminant, or probability function, which either:
   - Returns *True*, if two manuscripts are assumed to be written by the same author, else: *False*.
   - Gives a confidence measure in the range [0,…,1] where 100% means *True* and 0% means *False* w.r.t. the upper definition.

II GLOBAL ANALYSIS   25 P.

1. **Feature selection**
For global approaches, features are **not** computed separately for each character, but once per document. Thus, no features need to be extracted at prior. Again, it is up to you to design your system with suitable features. Here are some examples (which may or may not be suitable):
   - Histograms of greyvalues
   - Histograms of gradient orientations
   - (Accumulated) Projections of the image
   - Principle components
   - Fourier-based features etc.

Select some features for further application and discuss, which kind of features may be applicable for your task. Remember, that even documents written by the same writer may have different image dimensions.
Note: For this task you may use the total amount of 12 documents! Please remove/crop the (black) surroundings from the images before feature derivation!

2. **Implementation and feature derivation**
Design and implement a system, which computes your selected features for the different manuscripts, which we provided for this project.

3. **Comparison of feature applicability and decision function**
Compare the results of your features with respect to the given task. Use the (multi-dimensional) result of all of your (combined) features to develop a decision, discriminant, or probability function, which either:
   - Returns *True*, if two manuscripts are assumed to be written by the same author, else: *False*.
   - Gives a confidence measure in the range [0,…,1] where 100% means *True* and 0% means *False* w.r.t. the upper definition.