COURSE PROJECT – Handwriting Image Processing

Preliminary Submission Deadline: Sunday, December 10 2023, no later than 22:00 **SHARP**
Final Submission Deadline: Wednesday, December 20 2023, no later than 22:00 **SHARP**
Submission format: Electronic / repository
Submit to: skhachat@aua.am
Subject line: CS260/371 Project

Project Objective

    The course project is a part of a larger project that aims at predicting of student progress in CS and IT disciplines based of their handwriting samples [1]. It is expected to apply in its framework several Image Processing methods to identify and extract graphical features that could be used in the subsequent studies. Several Project stages are outlined below. After Stage 0, the stages can be addressed and implemented in an arbitrary order. Depending on the progress, additional or modified stages may also be posted. You are welcome to import and use OCR methods after proper acknowledgement and brief description.

    Stage 0 – store the images: Create a new folder **\CSHandwriting** in the same repository your HW01 is uploaded. Hard copies of handwriting samples were given for scan / digitalization with an individual code on each page. The code format is **CRS.TST.ddmmyy.Cxxx** (course name, test, date, lookup index – a letter H, M or L followed by 3 digits). Create in the folder **\CSHandwriting** subfolder for each **Cxxx** and save in them the images of all pages with such Cxxx. Name the files after the code specifying the page number as a suffix after the underscore, like **CRS.TST.ddmmyy.Cxxx_p1.png** or **CRS.TST.ddmmyy.Cxxx_p1.jpg**. If there are several images of the same page specify the copy in parentheses, like **CRS.TST.ddmmyy.Cxxx_p1(2).jpg**.

    Stage 1 – eliminate the printed text (Chapters 3, 4): Study the color distribution in each image using RGB and grayscale histograms. Suggest and test a method for automated removal of the printed test. If needed, suggest and test a method for automated removal of instructor's marks (normally appearing in red shades). Having the handwriting extracted, suggest a test a method to crop it. Convert the cropped image to binary format and save it under a filename **CRS.TST.ddmmyy.Cxxx_p1_bin.png**. Adjust the brightness / contrast before the conversion as needed (for example, auto-contrast, histogram match, etc.).

    Stage 2 – evaluate page features (Chapters 5, 6): Evaluate the size and orientation of the page in the image using the regions of the removed printed text. Among others, consider smoothening, sharpening, borders / edge detection, etc. Evaluate also the font size or average letter width in the printed text. The evaluated quantities will be used for normalization of to be detected handwriting features.

    Stage 3 – straight lines in handwriting samples (Chapter 8): Apply Hough Transform and detect straight lines / segments in the cropped binary handwriting samples. Investigate the possibilities to interpret them in terms of the baseline, slant, spacing, character size and / or other handwriting features.

    Stage 4 – binary regions in handwriting samples (Chapter 9): Apply morphological filters to construct convenient binary regions in the cropped handwriting samples. Using the constructed regions, suggest and test methods to separate characters, keywords (such as "if", "for", "while", "return", etc.), statements, etc. Try skeletonization to detect the baselines, label readability (clear vs. messy writing style), etc.

Stage 5 – labeled features of specific characters (Chapter 10): Using the geometric and statistical properties of binary regions, suggest and test methods to detect specific characters ("i", "f", "r", "t", "w", etc.) and label them. The labeling may refer but not limit to the style, circularity, position of the horizontal or vertical components, connectivity, etc.

Submission Conditions:
1. This is an individual assignment. Identical or similar submissions / files / results / reports / diagrams etc. will be disqualified – both the source(s) and receiver(s) will collect 0 point.
2. Group work will be accepted only if all group members are explicitly indicated in the submission. The individual contribution of each group member must also be explicitly stated, including all reasons of forming the group.
3. The submission deadline is rigidly strict. Submit even an unfinished work to get points and feedback. Late submissions will be disqualified and collect 0 point.
4. Not only precise solutions, but also free-format descriptions of ideas, difficulties, algorithms, simplifications, assumptions, etc. may be submitted.
5. All used external sources must be explicitly acknowledged.

Reference
1. Naira Khachatrya. Using Handwriting Features to Predict Student Performance in IT and Engineering Domain, BS in DS Capstone Thesis, College of Science and Engineering, AUA, 2022.