

HOMWORK ASSIGNMENT #4

Optical Character Recognition (OCR)

Due Date: 11:59am on 05/30/2018

Please read the submission guideline (posted on the class website) carefully before getting started.

All images in this homework can be downloaded from our class website:

<https://ceiba.ntu.edu.tw/1062DIP>. Images are in the raw file format. The size of each image is listed in the appendix.

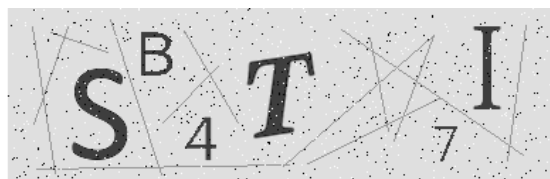
For MATLAB users, you are **NOT** allowed to use the MATLAB Image Processing toolbox except the `imshow()` and `image()` functions.

PROBLEM 1: Optical Character Recognition (OCR)

Please design an algorithm to recognize the CAPTCHA images as shown in Fig. 1. The training set is given in Fig. 2. Please describe the proposed algorithm in detail along with a flowchart and discuss the recognition results.



(a) sample1.raw



(b) sample2.raw

Fig. 1: Two CAPTCHA images



Fig. 2: TrainingSet.raw

Appendix:

Image files

Problem1: Optical Character Recognition

| | | | |
|-----------------|-----------|-----------------|------------|
| sample1.raw | Fig. 1(a) | 390 x 125 image | gray-scale |
| sample2.raw | Fig. 1(b) | 390 x 125 image | gray-scale |
| TrainingSet.raw | Fig. 2 | 450 x 248 image | gray-scale |