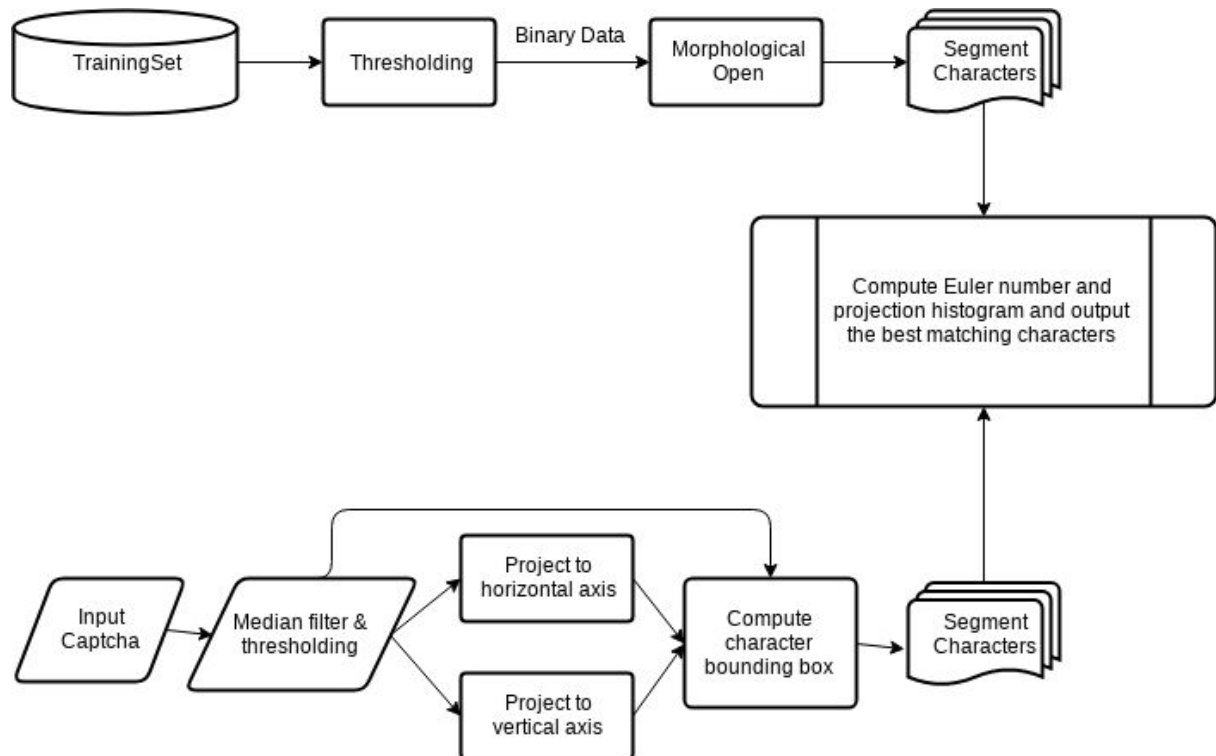# DIP HW4 Report

B03902082 資工四 江懿友 big2632@gmail.com

Problem 1: OCR

1. Threshold the TrainingSet image to turn it to binary data.
2. Apply morphological open to TrainingSet to smooth character contour.
3. Segment & label each characters. Segmentation result:

4. Perform data augmentation and compute Euler number and projection histogram for each characters in TrainingSet.
5. Apply median filter to input CAPTCHA to remove salt & pepper noise.

6. Threshold the input CAPTCHA to turn it to binary data.
7. Project the image to horizontal axis, and use the projection to segment characters.
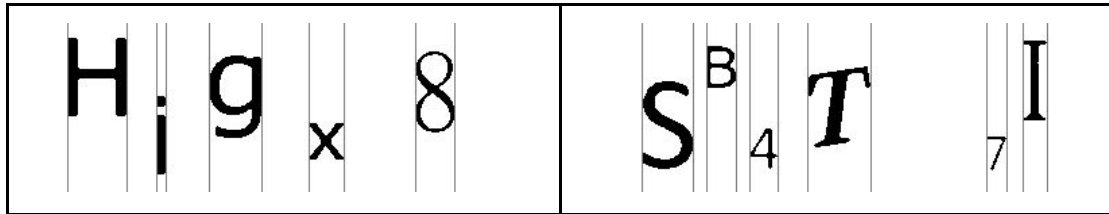
8. Compute Euler number and projection histogram of CAPTCHA character.
9. Find the best matching in TrainingSet.

At step (4) I augment the training data by flipping "!" upside down to be "i", so there are two versions of "i" in TrainingSet with different fonts.
Projection histogram is computed by projecting the image onto x and y axis:

$$h_x(x) = \frac{1}{A}\sum_y I(x,\ y),\ h_y(y) = \frac{1}{A}\sum_x I(x,\ y),\ A = \sum_{x,y} I(x,\ y)$$

For example $h_x(x)$ is the projection histogram w.r.t x axis, the histogram can be interpreted as the distribution of pixel coordinates. So the histogram can be seen as a feature vector of a image.
At step (9) I first use Euler number to narrow down the search space, so that a querying character is only compared with characters that have the same Euler number. For example, the first character in sample1, 'H', will never be predicted to be 'g' because they have different Euler numbers. Then I compute the MSE of projection histogram between querying character and candidate character (from TrainingSet).

The predicted results are:

Higx8	SBAt7I

sample1's predict accuracy is 100%.
sample2's predict accuracy is 66%.

In sample2, '4' and 'T' are mispredicted to be 'A' and 't'.
Character '4' is mispredicted because '4' looks like a skewed 'A' and the '4' in Training Set has a different font from the '4' in sample2, Training Set's '4' Euler

number = 1 and sample2's '4' Euler number = 0. This is a kind of Out-Of-Vocabulary error.
Character 'T' is mispredicted because the fonts of Trainging Set and sample2 are different. The font of sample2 is a kind of serif font and the small line at the bottom of 'T' makes it looks like a 't'.

I also tried using image orientation, image eigenvalue ratio, bounding box width-height ratio and Hu's invariant image moment as image features, but all of them yielded poor result and only worsens the prediction accuracy.