Hiragana Recognition using Optical Character Recognition

Author: Joonas Lõmps - b12027@ut.ee - Computer Science Masters Year I
Faculty of Science and Technology - Institute of Computer Science
Supervisor: Amnir Hadachi
https://github.com/joonaslomps/hiragana-ocr

Objective

Implement a program that takes an image of handwritten hiragana characters as an input and as an output adds pronunciations of the detected charactes below them.

What is hiragana?

Hiragana is a Japanese syllabary, one basic component of Japanese writing system, along with katakana and kanji [1]. Hiragana consists of 46 distinct syllables as well as some functional marks to add double consonants or soften them.

Dataset

Using optical character recognition (OCR) requires machine learning and the latter requires a dataset to train on. The dataset that was used for this project was made by the author, because he did not manage to find any usable dataset online. The generated dataset consisted of 460 images, 10 for every syllable.

Classifiers and data features

Classifiers:

- *k*-Nearest Neighbor - *k* = 4.
- Support Vector Machine with Linear kernel

Data features:

- Raw features - pixel intensity in grayscale
- Histogram of oriented gradients (HOG) - histogram of gradient directions of small connected regions of the image.

Image preprocessing

To enchance the results images were preprocessed before using them. Every image was converted to grayscale, thresholded and then dialated. Figure 1 demonstrates the differences.

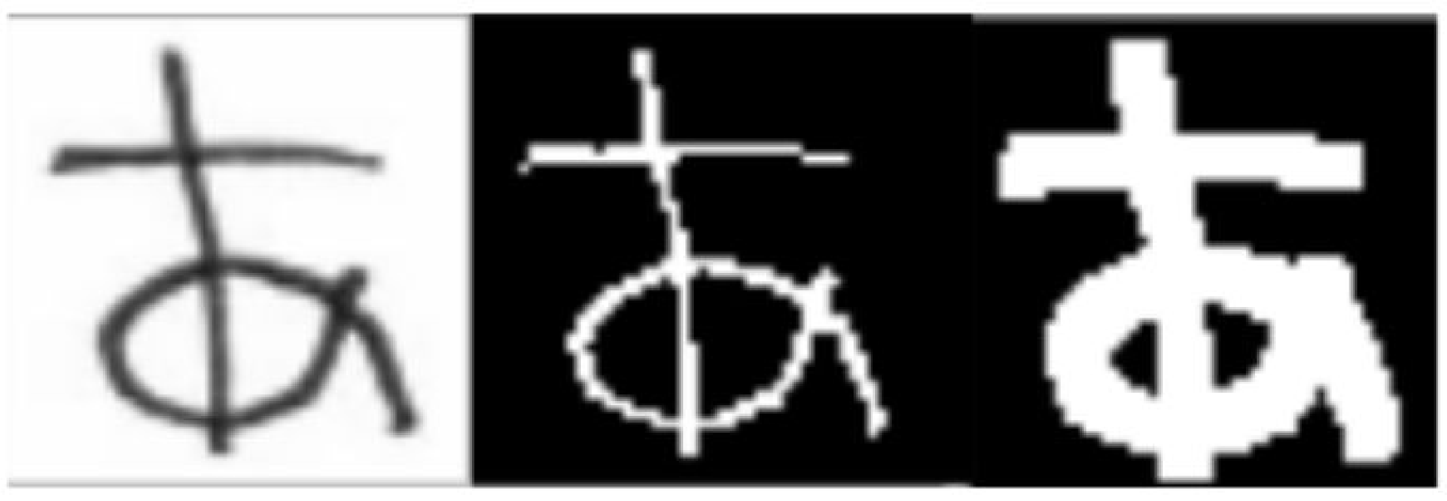

Figure 1. Left: intial image. Center: threshold image. Right: dialated image.

Recognition and accuracy

Instead of implementing the classifiers a computer vision library called OpenCV was used [2]. After testing several image sizes the best results yielded with 50x50px images. Out of every 10 image of a syllable, 8 were used in the training set and 2 in the testing set. Per syllable accuracy rate was rather high with any combination of classifiers and data features as it can be seen on Table 1.

Classifier	Feature	Accuracy
k-NN	Raw	92.39%
k-NN	HOG	91.30%
SVM	Raw	94.57%
SVM	HOG	96.74%

Table 1. Accuracy rates for recognising single syllables

Application

As we take an image of some syllables as an input we have to detect every syllable on that picture. For that, the author decided to use background substraction. It raises an issue: not all syllables are connected and are not detected as a whole. After thinking about it, the author saw that there are three different cases for it:

- On top of each other
- Side by side
- Both together

For the case where the unconnected parts are on top of each other, their starting points from the left are really close to each other. The two parts are then connected if |x1 - x2| ≤ 10. Similar logic is used for the case when they are side by side, but the allowed difference is 50. Result of first and second combining can be see on figure 2.

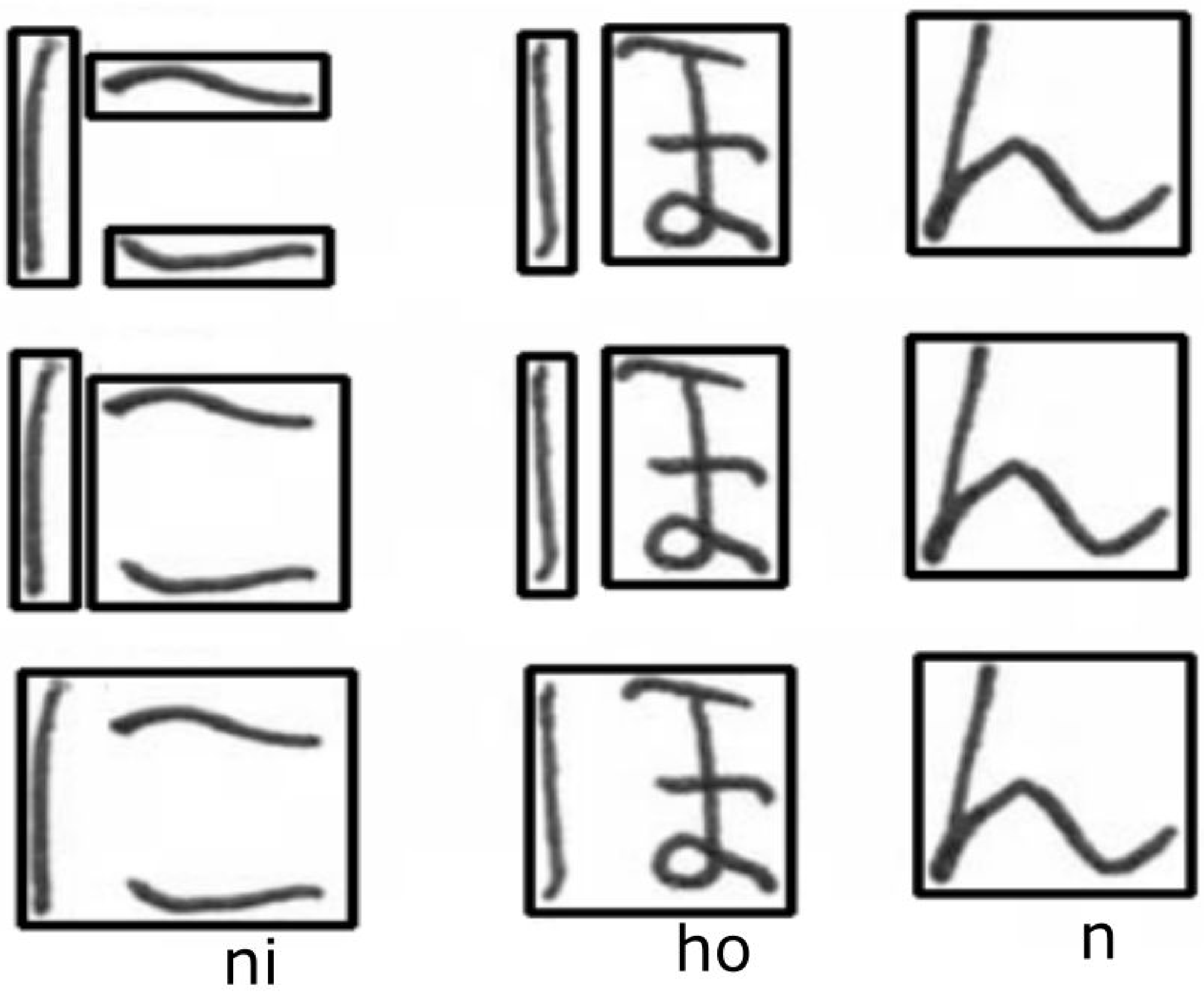

Figure 2. Top: Background subtraction.
Middle: After combining ones on top of each other
Bottom: Final result

Future work

The author has collected a much larger dataset of handwriting from 20+ people with 160+ writings per syllable. There are plans to use clustering to combine unconnected parts of the same syllable, as well as try other classifiers and data featurs to get increased accuracy.

[1] http://www.oxforddictionaries.com/definition/english/hiragana
[2] http://opencv.org/