

Title	深層学習を用いた単眼カメラによる空書の自動認識
Author(s)	藤本, 一文
Citation	
Issue Date	2020-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/16417
Rights	
Description	Supervisor: 長谷川 忍, 先端科学技術研究科, 修士 (情報科学)

Automatic recognition of Kusyo by a monocular camera using deep learning

1810165 Fujimoto Takafumi

Sign language is one of the methods used for Deaf people to communicate with hearing people. However, only a handful of hearing people can use or understand sign language. The number of Deaf people in Japan is 340,000, but there are only less than 50,000 people can support them, such as sign language interpreters and summary scribes. The shortage of sign language interpreters is one of the critical social issues. Moreover, the words we use are constantly being updated, so sign language must adapt to the new words. Research on automatic recognition of sign language is progressing to solve the problem. In recent years, because deep learning in the field of AI research is rapidly evolving, research on time-series recognition from images and moving images are also in progress. The recognition accuracy has dramatically improved. In order to deal with new words, however, we focus on the next most common method of communication, characters such as air writing (Kusyo). Kusyo refers to writing characters in the air without using a brush or paper. This can be written mainly with one finger. Kusyo is often used when it cannot be represented by sign language words or finger letters, or when it is necessary to convey that “shape” has meaning. Also, this is faster than sign language for simple interaction with hearing people. Therefore, an environment that can automatically recognize Kusyo is required as well as sign language. However, there is not much research supported up to Kusyo.

The purpose of this research is to extend the range of automatic sign language recognition by developing a method for recognizing Hiragana’s Kusyo written in front of a monocular camera using deep learning. A monocular camera can be installed at a lower cost than a dual camera, so it is easier to set up the environment. We also aim to develop a non-contact environment that does not use special sensors other than cameras.

The Hiragana used in this research consists of 50 characters (excluding ゐ and ゑ), 71 voiced and semi-voiced characters. There are three research questions in detecting and recognizing Kusyo. 1)It is not possible to identify where to start writing and where to end writing since it is an aerial character, 2)It is not know the line cut (between the first and second strokes) because it is a one-stroke, 3)It needs to prepare because there is no data of Kusyo by Hiragana for learning.

In this research, we developed an image of Kusyo and identified Hiragana by CNN. First, a region of a hand using colored gloves is detected, and a feature amount is extracted. These features are used for the following purposes. One is to set the start and end points for writing based on the feature amount on the screen. The other is to take the center of gravity of the hand area. By displaying the trajectory of the center of gravity, Kusyo was reproduced. At this time, the line toward the upper left direction was deleted, and the line cut was reproduced. In the experiment, these uncorrected data and corrected data are compared. Next, the obtained trajectory of the center of gravity is output as images data and cut out. The clipped image data is sent to the discriminator. In the discriminator, we use CNN that one of the deep learning. In addition, we use the fine tuning which allows learning with less training data. It uses ResNet for the network. Finally, Hiragana was identified from the result of the output from the discriminator.

The data to be learned was a hiragana's Kusyo by the participants using a video camera. In this research, we set up a camera in front of the participants and acquire the image of the hiragana's Kusyo because we often see a face-to-face when writing it. Kusyo is a part of handwritten characters. Handwritten characters include cursive writing. For this reason, we also conducted experiments that added publicly available handwritten character data to the learning data as well as Kusyo data collected from the video. Since the handwritten character data is different from the Kusyo data, the recognition accuracy may decrease depending on the difference in data amount. It was added in three patterns. In the comparative experiments, we compared a total of five models: a leaning model without correction, a leaning model with correction, and a leaning model with handwritten character data divided into three patterns and added with correction.

We have obtained generalized results by using 5-fold cross validation as an evaluation method. In the experimental results, the discrimination rate was 96[%] without correction. There were an identification rate of 97[%] with correction and 98[%] when handwritten character data was added there. We got good results by the effects of correction and handwritten character data.