J8 Construction of a Handwriting Database of Japanese Writers

Yoko Seki, MA*, 6-3-1, Kashiwanoha, Kashiwa-shi, Chiba-ken, 277-0882, JAPAN

After attending this presentation, attendees will be given an overview of the handwriting database of Japanese writers and its application to a fundamental experiment on writer identification.

This presentation will impact the forensic science community by demonstrating the validity of the database through the introduction of the findings obtained using the database.

Two-hundred and sixty-eight subjects (139 males and 129 females) participated in the data collection. Ages varied from 20 to 59 (mean age: 36.2). The subjects were requested to write 548 Kanji characters, 46 Katakana characters, 46 Hiragana characters, 52 Latin alphabet letters, and 10 Arabic numerals. Kanji characters selected for the database were characters listed in the national list of Chinese characters in common use and characters used for Japanese prefecture and prefectural capital names. Subjects were instructed to repeat the task five times and they wrote 3,510 characters in total. The subjects were also instructed to write a character in a 1cm square box printed on a data sheet. There were 117 boxes printed on a data sheet and six sheets were assigned to one subject per task. A digital pen was used for the data acquisition and x-y coordinates of the pen tip were acquired online. Average time necessary to one task was 40 minutes. The database had x-y coordinate data and handwriting image data. Image data was acquired by scanning the data sheet at a resolution of 800 dpi. Image data can be viewed on a monitor. An operator can load and observe one writer's image data or load several writers' handwriting and compare them. The database will be of use to both the research on handwriting and the forensic document examination.

A writer identification experiment was conducted using the database. In this experiment, writer identification was accomplished using characters that were different but having the same component characters for questioned and known handwriting. All the characters used for the experiment were constructed with two components. Four construction types (divided into right and left components, divided into upper and bottom components, inside component surrounded by another component at left and bottom sides, and inside component surrounded by another component at upper and left sides) were examined. Two characters (assuming character number one and number two) per construction type were selected for the examination. The experiment was conducted using 10 writers' handwriting samples (10 subjects times 5 repetition = 50 samples per character). The x-y coordinates on the measuring points, which had been defined beforehand on the common component, were obtained and then standardized. Euclidean distance of corresponding measuring points between questioned and known samples was calculated and compared. Identification was determined as follows: Any one sample of character number one was selected as the questioned handwriting. All samples of character number two (50 samples) were defined as the known handwriting. Euclidean distance between the questioned handwriting and any known handwriting samples were calculated and compared. This procedure was conducted on all character number one samples (50 samples), where this procedure was repeated 2,500 times on one character. Then, any one sample of character number two was selected as the questioned handwriting and all character number one samples were used as the known handwriting and the same procedure as the previous case was taken. The correct identification rate was calculated and defined as the case where the smallest distance was observed between the questioned handwriting and the known handwriting of the genuine writer of the questioned handwriting. The average correct identification rate was 58%. Right and left component construction showed the highest correct identification rate. Top and bottom component construction and inside component surrounded at top and left side construction showed low correct identification rate. These results showed the comparison of the questioned and known handwriting using only a part of a character was not valid. These results also suggested that a writer paid attention to write a character in a well-balanced shape as a whole and modifying the shape of the components according to the character he or she was going to write.

Handwriting Identification, Database, Japanese Character