Recognition of Written Cues System for Users of General Paper Media

Daiki Yamaji(✉) and Jiro Tanaka

University of Tsukuba, Tsukuba, Japan
{yamaji,jiro}@iplab.cs.tsukuba.ac.jp

Abstract. This paper proposes a system for users of "general paper media (newspaper, books, publications, etc.)" using recognition of written cues (made by handwritten entries) and performing digital processing. Users are able to use this system by a smartphone and on paper-media to save a favorite paragraph or image on the paper, illustrate data associations, and search for English translations, all the while being able to use the paper-media in a natural way. Moreover, users are able to browse the interaction logs from both the paper-media and smartphone. Experiments to evaluate the performance of this system shows the high recognition accuracy, and high discrimination accuracy depending on written cues.

Keywords: Document recognition · Handwriting · Image processing · Image recognition · Data management · Smartphone

1 Introduction

Due to recent development in digital technology, there has been much research going on to improve usability by connecting digital technology with the real world. For instance, in the field of NUI and TUI, it is possible to capture real-world actions and perform feedback based on those actions. However, considering our daily activities, there have not been many cases of smoothly transitioning our real-world actions into digital processing. In such cases, we focus on using the general paper medium (newspaper, books and publications), and consider the following scenarios. (1) when discovering a content, figure or photo that you like, putting a sign around it and capturing by a camera or scanning it, (2) classifying paper media including various writing and related articles and storing them in files together, (3) when trying to check the meaning of a word, putting a sign around it and searching the word by using a computer or smartphone, etc. Based on these cases, in this paper we propose a system of applying digital processing by using the natural human action of "writing something on a paper medium" as a trigger to smoothly realize transitioning our real-world actions into digital processing. Basically, a pen is used on a paper medium to (1) draw "⌈⌋" at diagonal ends of the desired region to save it, (2) draw the same characters on the upper-left side of the "⌈⌋" to associate that regions, (3) enclose "□" the

S. Yamamoto (Ed.): HIMI 2015, Part I, LNCS 9172, pp. 466–476, 2015.
DOI: 10.1007/978-3-319-20612-7_45

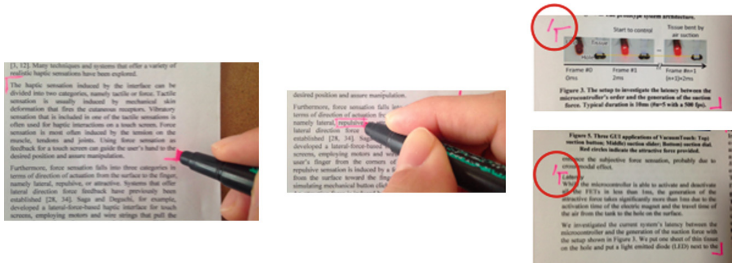

Fig. 1. (Left) Favorite figures, pictures or paragraphs enclosed by "⌈⌋", (middle) English word enclosed by "□" to search for its translated meaning, (right) similar characters ("1" is written in this case) written on the upper-left side of the regions enclosed by "⌈⌋" to associate the regions.

English word user does not understand the Japanese meaning. Using analog actions on the paper medium as input allows users to enjoy the benefits of the digital platform during casual use of the paper medium (Fig. 1).

2 Related Work

2.1 Link Between Paper Medium and Digital Data

Among the field of seamless integration of digital and analog media, there have been a number of research focussing on the link between paper medium and digital data. Koike et al. [1] and Do-Lenh et al. [2] have developed a system that links the real world with digital data, where, upon placing something like a book with a marker pasted on it on a table, the system can project digital information related to the book in the vicinity. Although these systems are similar to our research in terms of connecting paper medium with digital data, the aforementioned systems require markers to be pasted and digital data prepared beforehand for the systems to work. These are not required for our system. Sangsubhan and Tanaka [8] have developed an idea generation support system by automatically digitizing data written using a digital pen on a paper medium. This research also focused on the uses of the digital data after digitizing analog data, and depending on written cues, digitized English words can be looked up for meanings and kept for later study, or multiple data can be grouped together for combined browsing.

2.2 Extracting Written Cues

Nakai et al. [4] have proposed a method of extracting written cues on a paper medium by comparing the image of the paper with an original digital version and detecting the position of the cues made. Iwata et al. [5], by attaching a miniature camera on the tip of a pen, have managed to detect the position of the written cues made without scanning, and by using subtraction technique were able to

extract written cues. However, the aforementioned systems require a digitized version of the data present on the paper medium, which is not required for our system. Moreover, the recognition process is huge and time consuming. Stevens et al. [12]. have developed a high performance system for extraction of written cues by restricting the color used for the written cue. Guo and Ma [10] and Zheng et al. [11] have developed a system that can extract written cues only from paper medium where written cues have been made. However, the system is only capable of extracting handwritten letters but unable to do so for handwritten lines or figures. In our system, we have used color information of written cues for detection, in order to be able to use the system on smartphones which have significantly lower specs compared to PCs. Here, high precision extraction of written cues has been achieved. In addition, by allowing users to choose the color used for extraction, the system provides flexibility.

3 Data and Commands Arrangement Design

In this section, we will introduce the method of use and application processing. There are two modes in this system.

- Recognition mode
- View mode

In use of recognition mode, system applies digital processing by recognizing cues written by the user. In use of view mode, the user are able to browse digital datas stored by recognition mode.

3.1 Recognition Mode

To use this system, users hold a pen and use a smartphone. The user holds the phone up over the paper while writing or after he finished writing. There are three types of written cues (made by handwritten entries) recognized by this system.

(1). Enclosure by "⌈⌋". The user likes an image or paragraph on the paper, and would like to save it. By drawing "⌈⌋" at the diagonal ends of the desired region with a pen, "⌈" is enclosed by red square and "⌋" is enclosed by green square on smartphone, the user is able to confirm that this system recognizes "⌈⌋" (Fig. 2(a)). By touching the region, the rectangular area enclosed by "⌈⌋" will be saved on the smartphone as digital data (Fig. 2(b)).

(2). Drawing the same character. By writing a character with a pen on the upper-left side of the "⌈⌋" used in (1)., associations can be made in between regions designated by the same character. The character is enclosed by a yellow square on the smartphone, and the user is able to confirm that this system recognizes it. In Fig. 3(a) the character "1" is written; the corresponding region and the other region where the same character (i.e. "1") is written are automatically associated by touching the region (Fig. 3(b), (c)). This function can be useful in

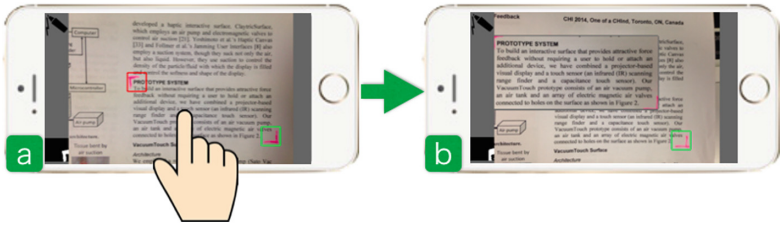

Fig. 2. (a) By drawing "⌈⌋" and touching the region the user would like to save, (b) the region will be saved as digital data.

scenarios where the user wants to save an figure and the description associated with it as a set. The user can read the description while looking at the figure on the smartphone.

(3). Enclosure by "□". The user does not understand the meaning of an English word on the paper. By enclosing the word with "□" with a pen, the word is enclosed by a blue rectangle on the smartphone; the user is able to confirm that this system recognizes "□" (Fig. 4(a)). Touching the region prompts the system to show the translated (Japanese) meaning on the upper left side of the display (Fig. 4(b)). In addition, the word enclosed by "□" and the corresponding translation will be saved on the smartphone.

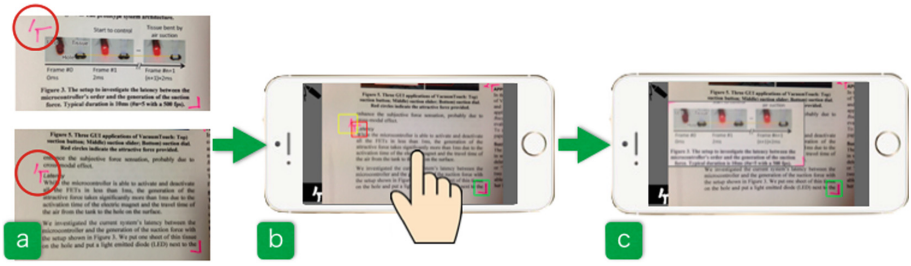

Fig. 3. (a) By writing "1" on the upper-left side of the "⌈⌋" and (b) touching one of the region (the lower figure of (a) in this case), (c) the other region associated with the region will be displayed.

Selection of region by touching on screen. The regions saved using the above methods can be selected by touching them on the smartphone. With the selection of a region, if created with (1)., the digital image of that region will be displayed on the screen. If created with (2)., the digital images associated with the region will be displayed, then one of the digital images will be displayed larger. If created with (3)., the translated meaning of the word within the region will be displayed on the top-left side of the screen.

Registration of the pen. Due to different pens being used by users, the system allows for any pen to be used for making cues. The process for registering a pen for making cues is as follows. While the system is running, the user scribbles something on a scrap of paper with the desired marker pen (Fig. 5(a)).

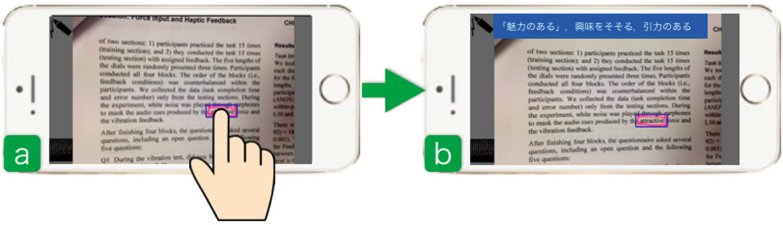

Fig. 4. (a) By drawing "□" and touching the word, (b) display the translated (Japanese) meaning on the upper left side of the display

By touching the pen icon on the upper left corner of the screen, the pen icon turns yellow and the system goes to pen registration state. Then, by touching the scribbled region on the smartphone screen (Fig. 5(b)), the system extracts the color of the ink (Fig. 5(c)), and the color is registered as the default color for future recognition and extraction.

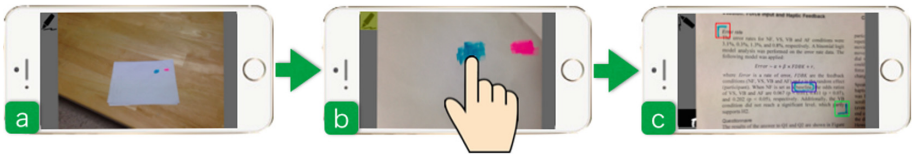

Fig. 5. (a) Scribbling something on a scrap of paper with the desired marker pen, (b) by touching the pen icon and the scribbled region, (c) the system extracts the color of the ink

3.2 View Mode

Since the data extracted using the written cue are saved on the smartphone, the data can be accessed and browsed anytime. The data may consist of figures, pictures, images of paragraphs, English words and their translations, and are divided into two categories: "IMAGE" and "WORD". The "IMAGE" category allows users to collectively browse extracted figures, pictures or paragraphs (Fig. 6(a)). In addition, image data that have been grouped together using the same letter show a yellow triangle mark on the upper right corner and the letter used for grouping on the upper left corner, and can be browsed as a set of grouped images (Fig. 6(b)). By selecting the "WORD" category, English words previously extracted and looked up for meanings can be rechecked (Fig. 6(c)) and their meanings rechecked as well.

4 Implementation

This system has been implemented as an application running on iOS. In this chapter, we will explain the implementation method used for the "recognition mode" described in the previous chapter.

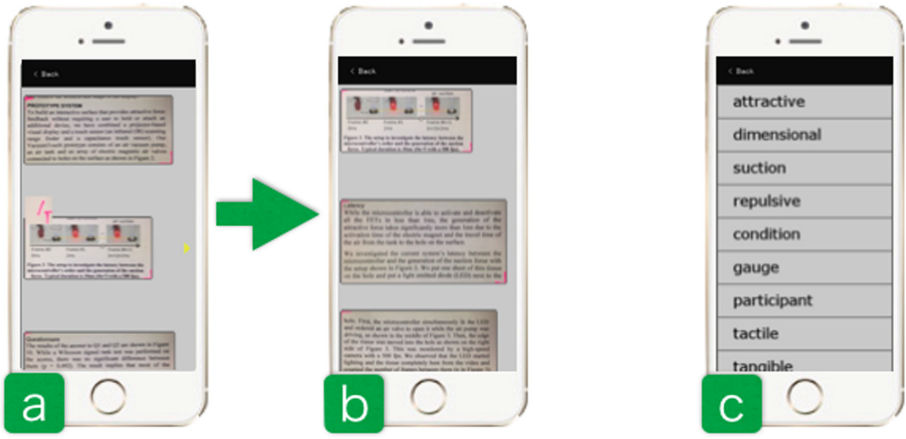

Fig. 6. (a) "IMAGE" category shows images list saved by recognition mode, (b) images list grouped by "1", (c) "WORD" category shows words list saved by recognition mode

4.1 Recognition of Written Cues

In this mode, simple shapes such as "⌈⌋" or "□", as well as letters used for grouping can be extracted. In addition, by taking into account the processing power of smartphones, we have used extraction of specific colors to detect written cues; thus reducing the processing load and improving accuracy.

4.2 Color Extraction

The robust HSV color space has been used for color extraction. The written cues extracted with regard to the specific color are distinguished as a separate region from the main text, and then binarized for the shape recognition process explained in the following.

4.3 Shape Recognition

After detection of written cues, recognition is performed. To recognize the written shape in the image obtained from the smartphone camera, "template matching" has been used as the recognition algorithm. In this mode, the digital processing performed differs depending on the detected shape.

4.4 Processing for the Shape "⌈⌋"

Using template matching, the coordinates of the upper left positions of "⌈" and "⌋" are obtained. A rectangular region is determined using these coordinates, and then cut out from the RGB image obtained from the camera.

4.5 Processing for Written Letters

In order to recognize letters, the region containing the written letter is cut out as image data. The method used for this is described as follows. First, a rectangular region has to be obtained using "⌈" and "⌋". Then, another rectangular region of 30 px containing the letter is cut out, with upper left corner of the rectangle starting at 15 px above and 20 px to the left of the upper left corner of "⌈". Using this image as a template and by performing template matching on images containing letters obtained later, the letters are compared, and upon finding similar letters, would group the regions enclosed by "⌈" and "⌋" that are associated with the letters.

4.6 Processing for the Shape "□"

Using template matching, the coordinates for the upper left corner of "□" can be obtained, and by applying these coordinates on the RGB image obtained from the camera, the English word enclosed within the "□" can be cut out. Then, from the image data obtained, the text can be extracted from the image file using OCR.

4.7 Text Extraction Using OCR

In this mode, OCR (optical character recognition) has been used for text extraction from image data (using "tesseract-ios" from OCR libraries). However, in case of noise in the image backdrop or blur in the image, the accuracy tends to drop significantly. Therefore, the images used in the system are magnified, converted to grayscale, sharpened and the contrast increased in order to improve the accuracy.

4.8 Translation of English Words

To look up the word extracted using OCR for its meaning, an English-Japanese dictionary web service called "Dejizo" has been used. By including the English word in the request URL while accessing Dejizo, the corresponding translation can be obtained included in an XML file.

4.9 Data Processing

The figures, pictures, paragraphs, English words and the corresponding meanings obtained from the recognition mode are then stored on the iPhone storage. Figures, pictures and paragraphs are stored as image files; the metadata of the image files and English words extracted for translation, and the corresponding translations are stored as XML files. While using the recognition mode, the data stored in the storage is imported into the application memory. This is to reduce the processing time for the application. When saving data, the data present on the application memory is compared to the one in the storage in order to determine if the data is the same or not. In addition, the data is also used when viewing the data in the "view mode".

5 Evaluation

A preliminary experiment has been performed for performance evaluation of the recognition accuracy of the recognition mode and the discrimination accuracy for the generated data.

5.1 Experiment Outline

We evaluated the following: the recognition and discrimination accuracy for "⌈" and "⌋", the discrimination accuracy for the letters written to the upper left corner of "⌈", the recognition accuracy for "□" and the discrimination accuracy for the enclosed region (English word). In addition, three types of marker pens have been used to evaluate difference of recognition by color of pen. A thesis paper written in English has been used as the paper medium for the experiments. Three university students of age 21–22 have been chosen as participants for the experiments.

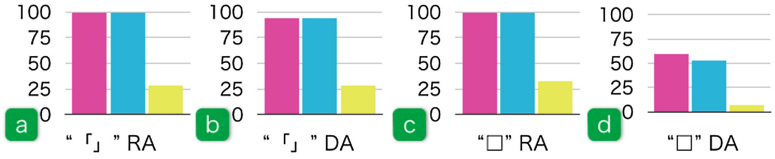

Fig. 7. (a) "IMAGE" category shows images list saved by recognition mode, (b) images list grouped by "1", (c) "WORD" category shows words list saved by recognition mode

The color of the graph represents the color of the ink of the marker pen used. "⌈⌋ recognition accuracy" represents the recognition accuracy for "⌈" and "⌋", and "⌈⌋ discrimination accuracy" represents the discrimination accuracy for the data obtained (such as paragraphs, figures, etc.). Similarly, "□ recognition accuracy" represents the recognition accuracy for "□", and "□ discrimination accuracy" represents the discrimination accuracy for the data obtained (English word).

The color of the graph represents the color of the ink of the marker pen used. Each graph represents the discrimination accuracy for the letter written on the upper left corner of "⌈".

5.2 Considerations

From the graphs it can be seen that, while using the red and blue pen, the results for the recognition and discrimination accuracies were similar, while the recognition and discrimination accuracy for the yellow pen was lower. The reason for this can be assumed to be the low contrast created by yellow ink on a white paper. From this it can be inferred that, for the system to work properly, the color of the ink used for writing cues needs to have high contrast with regards to the color of the paper used as the medium. However, when using red and blue pens, the recognition for "⌈⌋" and "□" tend to be accurate (Fig. 7(a), (c)). In addition,

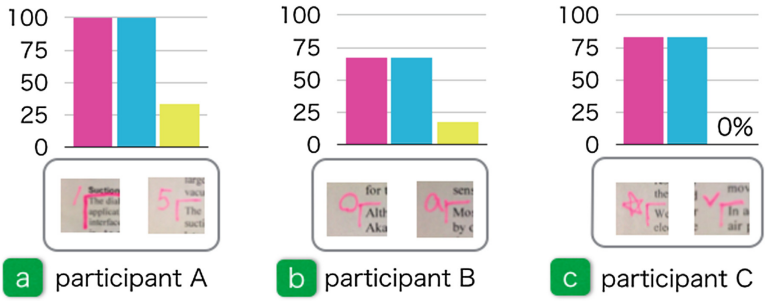

Fig. 8. The discrimination accuracy for the letter written by (a) participant A, (b) participant B, (c) participant C

data such as paragraphs and figures obtained using "⌈⌋" also have high discrimination accuracy (Fig. 7(b)). On the other hand, the English word obtained from "□" appears to have a discrimination accuracy of about 60 % (Fig. 7(d)). This result is seemingly affected by the accuracy of the text extraction of OCR. In this system, to increase the discrimination accuracy, several image processing have been applied on the image files of the data. For implementation of this system in the real world, the discrimination accuracy for "□" needs to be further increased by using improved processes. However, the ability to extract an English word enclosed within "□", with an accuracy of 60 %, might become the foundation of a real-world implemented version of a system based on paper medium. On the matter of discrimination accuracy for letters, for participant A, 100 % accuracy was obtained. However, for participants B and C, the accuracy turned out to be lower (Fig. 8). The decrease in accuracy for participant C might have been caused due to the inability of the participant to recreate the subtle intricacies of a star symbol (Fig. 8(c)), and for participant B, the similarity between the two symbols (Fig. 8(b)) might have caused confusion in the recognition process thus reducing accuracy. Although using simple letters and symbols gave accurate results, the system needs to be further improved to accommodate complex symbols, and the ability to differentiate between similar-looking but different symbols.

6 Discussion and Conclusion

In this research, we have developed a system for performing digital processing on data obtained through written cues on the "general paper medium" such as books, newspaper, publications, etc. Users are able to use the system with a smartphone, and by writing cues on a paper medium, the system can be used to save desired figures, pictures or paragraphs from the physical paper, associate aforementioned figures, pictures and paragraphs with each other, or show meanings of English words chosen by the user. In addition, the saved data and associated data can be browsed anytime on the smartphone, and since the

written cues persist on the paper medium, they can also be browsed on the physical paper. Moreover, since the system only requires a smartphone and pen in addition to the paper medium, and the digital processing requiring the natural process of writing something on paper, the system is very accessible and easy to use for the general people. From the preliminary experiments conducted, we have obtained high accuracy for recognition and discrimination depending on the pen used and written cues. However, on the matter of extraction of English words, we have deduced the need of further study to improve processing method to increase accuracy. In addition, to give accurate results, we have inferred the need of thick marker pens with the color of the ink having high contrast with regards to the color of the paper, which gives rise to some inflexibility for the system. Furthermore, while this system requires the source text material to be in English, we want to further improve the system to accommodate for other languages such as Japanese.

References

1. Koike, H., Sato, Y., Kobayashi, Y.: Integrating paper and digital information on enhanceddesk: a method for realtime finger tracking on an augmented desk system. ACM Trans. Comput. Hum. Interact. **8**(4), 307–322 (2001)
2. Do-Lenh, S., Kaplan, F., Sharma, A., Dillenbourg, P.: MultiFinger interactions with papers on augmented tabletops. In: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 267–274 (2009)
3. Brandl, P., Richter, C., Haller, M.: NiCEBook: supporting natural note taking. In: CHI 2010: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 599–608 (2010)
4. Nakai, T., Kise, K., Iwamura, M.: A method of annotation extraction from paper documents using alignment based on local arrangements of feature points. In: Ninth International Conference on Document Analysis and Recognition, ICDAR 2007, vol. 1, pp. 23–27 (2007)
5. Iwata, K., Kise, K., Iwamura, M., Uchida, S., Omachi, S.: Tracking and retrieval of pen tip positions for an intelligent camera pen. In: Proceedings of ICFR 2010, pp. 277–282 (2010)
6. Yoon, D., Chen, N., Guimbretire, F.: TextTearing: expanding whitespace for digital ink annotation. In: Proceedings of UIST 2013, pp. 107–112 (2013)
7. Harrison, C., Xiao, R., Iwamura, M., Schwarz, J., Hudson, S.E.: TouchTools: leveraging familiarity and skill with physical tools to augment touch interaction. In: Proceedings of CHI 2014, pp. 2913–2916 (2014)
8. Sangsubhan, P., Tanaka, J.: Idea generation support system utilizing digital pen and paper. Master Thesis, University of Tsukuba (2013)
9. Mazzei, A.: Extraction and classification of handwritten annotations for pedagogical use. In: Proceedings of EDIC 2009 (2009)
10. Guo, J.K., Ma, J.K.: Separating handwritten material from machine printed text using hidden markov models. In: Proceedings of 6th international Conference on Document Analysis and Recognition, pp. 436–443 (2001)
11. Zheng, Y., Li, H., Doermann, D.: The segmentation and identification of handwriting in noisy document images. In: Lopresti, D.P., Hu, J., Kashi, R.S. (eds.) DAS 2002. LNCS, vol. 2423, p. 95. Springer, Heidelberg (2002)

12. Stevens, J., Gee, A., Dance, C.: Automatic proceessing of document annotations. In: Proceedings of 1998 British Machine Vision Conference, vol. 2, pp. 438–448 (1998)
13. Yi, C., Tian, Y.: Text extraction from scene images by character appearance and structure modeling. In: Proceedings of CVIU 2013, pp. 182–194 (2013)
14. Huang, R., Shivakumara, P., Uchida, S.: Scene character detection by an edge-ray filter. In: Proceedings of ICDAR 2013, pp. 462–466 (2013)
15. Jain, A., Sharma, J.: Classification and interpretation of characters in multi-application OCR system. In: Proceedings of ICDMIC 2013, pp. 1–6 (2013)