The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Robotic system for reading Japanese characters on a written document in real time

Narumi Habu[1], Yukiharu Yamauchi[1], Sei-ichiro Kamata[2]
Youhei Shinfuku[2], and Takahiro Kawabuchi[1]

[1] Kitakyushu National College of Technology, Kitakyushu, Fukuoka 802-0985, Japan
(Tel: 81-93-964-7297, Fax: 81-93-964-7298)
(yamauchi@kct.ac.jp)
[2] Waseda University Graduate School, Kitakyushu, Fukuoka 808-0135, Japan

Abstract: This paper presents a robotic system in which a robot reads a printed document in Japanese. Currently, Apple's iPad, Amazon's Kindle, Sony Table, and other electronic book readers, are stimulating the electronic book market. These are, essentially smaller versions of personal computers and can be used to send emails, watch movies, or play games. In cooperation with Waseda University Graduate School and Shanghai Jiao Tong University, we have developed a robotic system called Ninomiya-kun that has the ability to read a book. This robotic system is the size of a child and has an appearance resembling a human being. The robot is equipped with two cameras as part of a visual information processing mechanism and can communicate with humans via voice and gestures. It automatically extracts character areas that contain Japanese characters, which are recognized using character recognition procedures. The system configuration is described, as well as the Japanese character recognition procedures performed by the robot. The experimental results are also discussed.

Keywords: Robot vision, entertainment robots, music score recognition, singing robotic system

1 INTRODUCTION

In fictional stories such as movies and cartoons, domestic robots are commonly used to assist humans; for example, Doraemon, a cat-like robot from the future, helps children fulfill their dreams using various secret devices. In a typical episode of this cartoon, whenever the central character becomes a target of bullying or fails in something, Doraemon uses unique technological devices from his belly pouch to help him. In the same way, robots will be working in offices, factories, and homes in the near future.

In 2004, a group of doctors suggested that parents limit their children's contact with TV and computer games. They claimed that TV and computer games reduced children's time for playing outside with their friends and having conversations with their families. They also recommended that parents read picture books to their children. Working in conjunction with Waseda University Graduate School and Shanghai Jiao Tong University, we have jointly developed a robotic system called Ninomiya-kun that can read a book (Fig. 1) [1, 2]. This robotic system is equipped with two cameras on the head and is able to communicate with humans via voice and gestures. It has an appearance that resembles a human being. The image of a printed document such as a picture book, textbook, or magazine, each of which has a large number of characters, is captured with the camera placed on the head, and the characters are extracted from the captured image.

Fig. 1. Robotic system Ninomiya-kun

Character recognition software in the robot translates these characters into spoken words, which are produced using a voice synthesizer. We have already proposed a book reading robotic system, Ninomiya-kun, by training its TV camera. This system could be applied to reads a book out loud for children or a visually impaired person. In the parents' stead, the proposed system can be used to read picture books to their children. This robotic system automatically extracts areas with Japanese characters, which are recognized by the character recognition application software. It has a character dictionary with a layered structure that has the ability to recognize over 2,000 kanji, hiragana, and katakana characters and can currently read elementary texts. This paper describes the method for segmenting a sentence

into Japanese kanji, hiragana, and katakana characters from the image data and retrieving them from the dictionary [3]. The proposed technique can be applied to read printed documents out loud using tablet computers such as Apple's iPad, Amazon's Kindle, Sony Table, or other electronic book readers. In the following, the system configuration is introduced, followed by the procedures for the Japanese character recognition. The experimental results are also discussed.

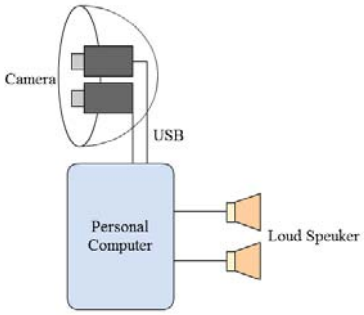

Fig. 2. System configuration of Ninomiya-kun

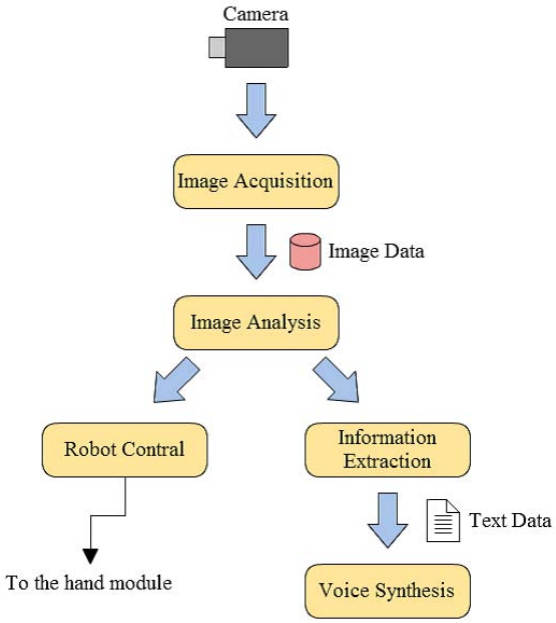

Fig. 3. Structure of software system

2 SYSTEM CONFIGURATION

The system configuration of Ninomiya-kun is shown in Fig. 2. This robotic system has three units in the robot's backpack: a personal computer unit, an image capture unit, and a loudspeaker unit, which is an electro-acoustic transducer that produces sound in response to the texts that are translated into spoken words. It has two hands that are driven by servomotors that are used to open a book and turn a page; the rotational speed and angles of the servomotors are given by the controlling personal computer unit [4]. Two cameras on the robots's head are used as visual sensors. The video signals from the cameras are transferred to the personal computer unit through an image acquisition interface unit, which provides digitized images at high resolution [6].

Processing a document to extract its content in an automated fashion is the essential task of this system. The following software modules are used to process a document, as depicted in Fig. 3.

1. The image acquisition module captures a document's image using the camera and the image acquisition unit attached to the personal computer and performs a binarization process on the image.

2. The image analysis module then extracts distinctive characteristics from the document being processed.

3. The information extraction module is a classification stage that identifies the characters and groups them according to certain classes.

4. The voice synthesis module reproduces human voices using a voice synthesizer.

5. The robot control module controls the posture of two cameras and drives the servomotors for the two hands.

3 CHARACTER SEGMENTATION FROM A PRINTED DOCUMENT

This section provides an outline of the image processing and analysis for the character segmentation from the image captured by the camera [7, 8]. Real-time recognition is realized by recognizing the document when the robot opens it and turns a page. However, the recognition depends on an algorithm and the speed of the computer. No delay is caused in the task of reading the document out loud because the character segmentation is carried out in the background. The system needs be fast enough to keep up with this reading task.

3.1 Distortion correction and image processing

Off-line character recognition uses a scanned image. The proposed system uses an image captured by the robot equipped with the camera. When the captured image is used as an input image, there are two differences from the image scanned with a scanner.

1. Image distortion: The image scanned with a scanner has no distorted characters, whereas the image captured by the camera has numerous distorted characters in many cases.

2. Image resolution: The number of pixels that is assigned to a character in the image captured from the camera is considerably smaller than that assigned to a character in the considerably as image from the scanner.

The character recognition algorithm needs to recognize an input image with low resolution. The character distortion must be detected and corrected. Before the correction of a distorted character, an input image is digitized. The digitized image is divided into eight domains, and the histogram for these domains are calculated (Fig. 4).

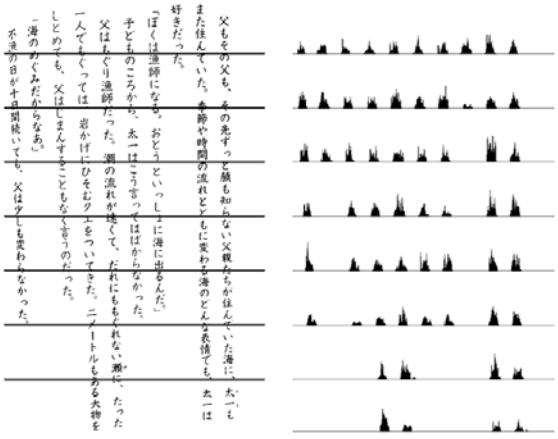

Fig. 4. Divided image and histogram

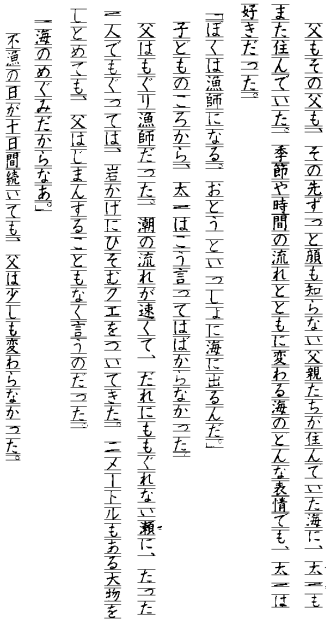

Fig. 5. Results for correction of distorted image

The inclination is calculated by using the histogram of a neighboring domain. In an image without inclination, we focus on the matching in the peak and trough positions in the histogram of the domain. With this matching, one side of the histogram of two neighboring domains is moved from side to side. The position that best matches the peak and trough in the histogram is calculated. The results of this correction are shown in Fig. 5.

3.2 Character segmentation

Next, the procedure for segmenting a character from an input image is described. The segmentation of a character is performed for every other line or column. One line is segmented out, and the segmenting candidate position of the character is calculated from the histogram of the line (Fig. 6 (a)). In the next procedure, the segmented characters merge sequentially from the top character. As an example, we take a case where two segmenting candidate characters are merged. If the length to width ratio is 1:1.2 or less, two candidate characters are merged. The merged case is shown in Fig. 6 (b) and (c). Another case is shown in Fig. 6 (d) and (e).

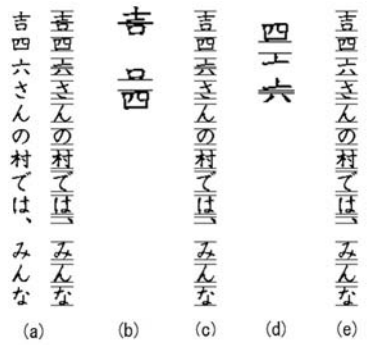

Fig. 6. Merging of two candidate characters

3.3 Feature vector detection

The features for every segmented character are calculated and compared with the system dictionary. The segmented characters differ in size. The outline box of a character is calculated and the size is normalized to 64 × 64 pixels. The directional element feature, which determines the amount of directional change in the outline, is used as a recognition feature. The outline pixels that match four directional changes are detected (Fig. 7).

Fig. 7. Four directional changes

An outline from the normalized image is calculated, and 12 kinds of 3 × 3 pixel masked images are used to extract the directional change in the outline (Fig. 8). The mask and outline pixels are measured, and the directional change is given to the outline pixel that matches the mask. The directional

change is then compared with the feature vectors stored in the dictionary.

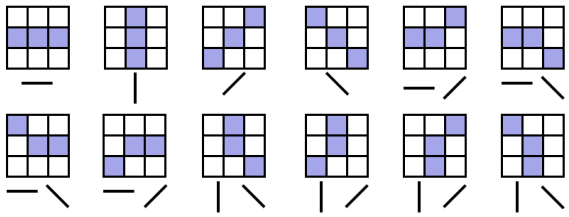

Fig. 8. 12 kinds of masked images

4 EXPERIMENTAL RESULT

An experiment was conducted using the robotic system Ninomiya-kun. A captured document and the character segmentation results are shown in Fig. 9. A portion of the recognition results is shown in Fig. 10. In the experimental results, for a document with 262 characters, 244 characters were correctly recognized.

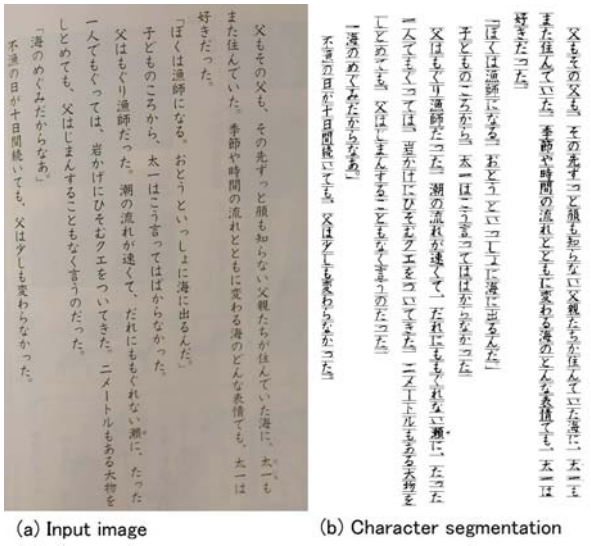

Fig. 9. Input image and character segmentation

5 CONCLUSION

We described a novel robotic entertainment system that can recognize Japanese characters from the a printed document. This system can recognize a printed document with nearly ninety percent accuracy. However, the recognition of a document printed with multiple letter types is difficult. The recognition accuracy is also greatly influenced by the lighting. The system needs to be further developed to improve the recognition accuracy for multiple letter types and the influence of lighting, which are now under investigation [9].

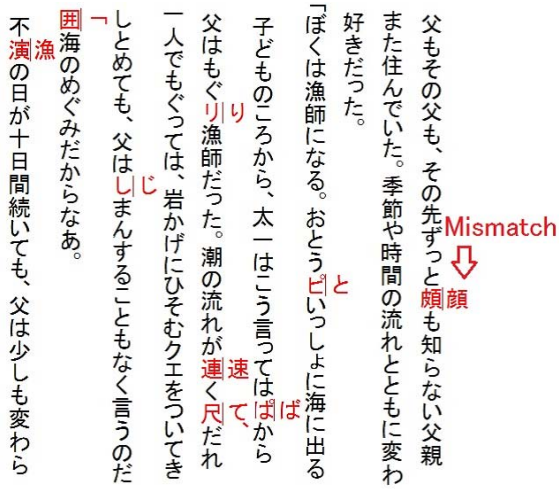

Fig. 10. Experimental results

REFERENCES

[1] Y. Komatsubara, Y. Yamauchi, S. Kamata (2008), Recognizing and reading for characters on books (in Japanese). ITE Annual conference 2008

[2] Y. Komatsubara, Y. Yamauchi, S. Kamata (2009), Character segmentation and reading from Japanese character (in Japanese). ITE Annual conference 2009

[3] T. Kawabuchi, Y. Yamauchi, S. Kamata (2010), Technique for Reading documents with multiple better type (in Japanese), SICE Kyushu Annual conference 2010, 203A-4

[4] S. Akram, M. Dar, A. Quyoum (2010), Document image processing, International Journal of Computer Apprlications, Vol.10, No.5:35-40

[5] Cheng-Lin Liu (2007), Normalization-cooperated gradient feature extraction for handwritten character recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.29, No.8:1465-1469

[6] S. Imamoto, Y.Yamauchi (2011), Head and stem extraction from printed music scores, International Workshop on Target Recognition and Tracking 2011, pp.72

[7] N. Habu, Y.Yamauchi (2011), Printed Japanese characters recognition system, International Workshop on Target Recognition and Tracking 2011, pp.73

[8] T. Hakui, Y. Yamauchi (2011), Two-finger robotic hand for a touch panel system, International Workshop on Target Recognition and Tracking 2011, pp.74