IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 13, 2019, accepted July 31, 2019, date of publication August 19, 2019, date of current version September 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936143

A Finger-Worn Device for Exploring Chinese Printed Text With Using CNN Algorithm on a Micro IoT Processor

YU-SHENG SU[1], CHIEN-HSING CHOU[2], YUNG-LONG CHU[2], AND ZHAO-YU YANG[2]
[1]Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung 20224, Taiwan
[2]Department of Electrical Engineering, Tamkang University, Taipei 25137, Taiwan

Corresponding author: Chien-Hsing Chou (chchou@mail.tku.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2511-H-019-002, Grant MOST 108-2511-H-019-003, Grant MOST 107-2221-E- 032-042, and Grant MOST 107-2511-H-019-003.

ABSTRACT This study designed a finger-worn device—named the Chinese FingerReader—that can be practically applied by visually impaired users for recognizing traditional Chinese characters on the micro internet of things (IoT) processor. The device is portable, easy to operate, and designed to be worn on the index finger. The Chinese FingerReader on the index finger contains a small camera and buttons. The small camera captures images by identifying the relative position of the index finger to the printed text, and the buttons are applied to capture an image by visually impaired users and provide the audio output of the corresponding Chinese character by a voice prompt. To recognize Chinese characters, English letters, and numbers, a robust Chinese optical character recognition (OCR) system was developed according to the training strategy of an augmented convolution neural network algorithm. The proposed Chinese OCR system can segment a single character from the captured image, and the system can accurately recognize rotated Chinese characters. The experimental results revealed that compared with the OCR application programming interfaces of Google and Microsoft, the proposed OCR system obtains 95% accuracy rate in dealing with rotated character images where the Google and Microsoft OCR APIs only obtain 65% and 34% accuracy rates. These results illustrate that the proposed OCR system was more suitable for the needs of visually impaired people in actual use. Finally, three usage scenarios were simulated, and the accuracy and operational performance of the system were tested. Field tests of this system were conducted for visually impaired users to verify its feasibility.

INDEX TERMS Assistive technology, Chinese OCR, IoT processor, convolution neural networks, wearable interface.

I. INTRODUCTION

Reading text is unavoidable in daily life. For example, reading books is crucial for gaining knowledge. People can learn about drug efficacy by reading the text on medicine bottles to avoid consuming the wrong drugs, which could be dangerous to human health. People can also operate home appliances correctly by reading the text and numbers on the buttons of appliances. Moreover, the current location, surrounding environment information, or key messages can be obtained by reading the text on maps or bulletin boards. However, visually impaired people cannot read such messages. Therefore, solving the reading problems of visually impaired people is a critical concern.

Visual impairment manifests in two forms: total blindness and amblyopia. The term "total blindness" refers to a complete lack of sight. People with total blindness cannot see anything and can only rely on sound or touch to perceive the real world. People with amblyopia still have some limited vision. They can walk slowly without assistance or vaguely see the outline and position of large or obvious objects. Nevertheless, these people cannot see the text on an object or read text on items such as bulletin boards. People that have either total blindness or amblyopia often require assistance from others to read a text and convey it through speech. The aforementioned issues are some examples of the daily problems faced by visually impaired people.

The associate editor coordinating the review of this article and approving it for publication was Chun-Wei Tsai.

The remainder of this paper is organized as follows. The related works are given in Section 2. The introduction of the system hardware and software is presented in Section 3. The software design of Chinese OCR system is introduced in Section 4. The construction of the OCR classifier with deep learning algorithm is given in Section 5. The evaluation results of the field trials are presented in Section 6 to demonstrate the effectiveness of the proposed method. The conclusion of the paper is discussed in Section 7.

II. RELATED WORKS

To assist people with visual impairment in reading printed text on signs or objects, several studies have been conducted, which can be divided into two categories [1]–[12]. The first category of research comprises a range of techniques that have designed wearable devices with a camera to assist visually impaired people [1]–[5]. After the acquisition of images containing text, some image preprocessing techniques have been used to detect the position of text. The text is then segmented for recognition by an optical character recognition (OCR) classifier. To help visually impaired people to read text from signage and objects in the hand, Yi and Tain set the camera on glasses to design a camera-based assistive system [1]. They also designed the text localization algorithm to localize text regions in images by learning gradient features of stroke orientations and distributions. Hanif and Prevost installed a micro camera on an intelligent glass to assist visually impaired people and facilitate navigation in the real world [2]. Their intelligent glass could recognize the text on signs and interact with vibration feedback. Mattar et al. developed a wearable system that was capable of detecting and recognizing signs in natural scenes [3]. The sign detector uses a wide array of features with a conditional random field classifier to find sign regions in the image. Ezaki et al. used a camera placed on the user's shoulder to acquire images and presented a scene-text detection module for visually impaired people [4]. Character extraction methods based on connected components have been proposed to detect text area from natural scene images.

The second category of research involves related techniques that have applied handheld and finger-worn devices to assist visually impaired people [6]–[12]. Shen et al. designed a prototype smartphone system that finds printed text in cluttered scenes and reads aloud text signs for visually impaired users [6]. They implemented an audio-tactile user interface on the Android platform. This interface assists the user in locating any text of interest and approaching it. Guo et al. presented an accessible mobile application (namely as VizLens) and a supporting backend that can robustly and interactively help visually impaired people use inaccessible interfaces in the real world [7]. The proposed approach helps visually impaired people recapture the interface in the field of the camera, and interactively describes the part of the interface beneath their finger. Kandemir et al. developed a smartphone application that can read aloud Turkish scene and book text [8]. Images acquired from the smartphone camera are processed to detect text regions, recognize the text, and synthesize speech in Turkish. Liu et al. developed Finger-eye, which is a portable and refreshable text reading system and a rapid OCR algorithm with a small camera [9]. The text is recognized to translate as audio or generate corresponding electrotactile stimulation on a visually impaired people's finger. Jiang et al. applied OCR and the text-to-speech (TTS) technique on an Android smartphone [10]. The proposed approach can recognize text and signs in the environment for navigating signs and can reconstruct sentences and convert them into speech. Shilkrot et al. designed a wearable device called the FingerReader [11], [12]. They proposed a computer vision algorithm for local-sequential text scanning, which enables the reading of single lines, reading of blocks of text, or skimming of the text with complementary multimodal feedback integrated with a finger-worn device. The FingerReader utilizes a TTS engine for enabling visually impaired people to listen to printed text.

Although the researches in the aforementioned studies assist visually impaired people in reading text, some important problems must still be solved in practical applications. (1) Because visually impaired users cannot determine the exact location of text, they face difficulty in capturing useful text images. (2) When images are captured over a long distance by using a camera, the image quality may be easily affected by the light source, focal length, and reflection. (3) Visually impaired users cannot accurately capture images directly above text, which causes the captured text image to be skewed and rotated and thus increases the difficulty in OCR. (4) Because image preprocessing and OCR functions require a large amount of computation resources, some studies have relied on a notebook computer to accomplish these computations. However, carrying a notebook computer constantly is not convenient for visually impaired people. (5) Chinese OCR is difficult because there are many types of Chinese characters. Many categories have similar stokes, resulting in a few differences between Chinese characters (e.g., "人" and "入" or "大," "太," and "犬").

Recently, deep learning techniques have been used extensively in a wide range of fields. Convolutional neural networks (CNNs) [13] are applied to obtain the most accurate results in solving real-world problems, especially in image recognition. Several researchers have combined CNNs with other machine learning algorithms to recognize simplified Chinese characters [14]–[16]. Tang et al. applied CNN and transfer learning for historical Chinese character recognition [14]. Experiments were performed on 100 randomly selected classes of Chinese characters, and a satisfactory performance was obtained. Liu et al. developed a method of combining CNN with the support vector machine to identify similar handwritten Chinese characters [15]. Their method uses CNN to learn and extract features of simplified Chinese characters and then identifies 3000 samples of similar handwritten Chinese characters by applying the support vector machine algorithm. Song et al. used CNN and dataset expansion techniques to develop a handwritten Chinese character

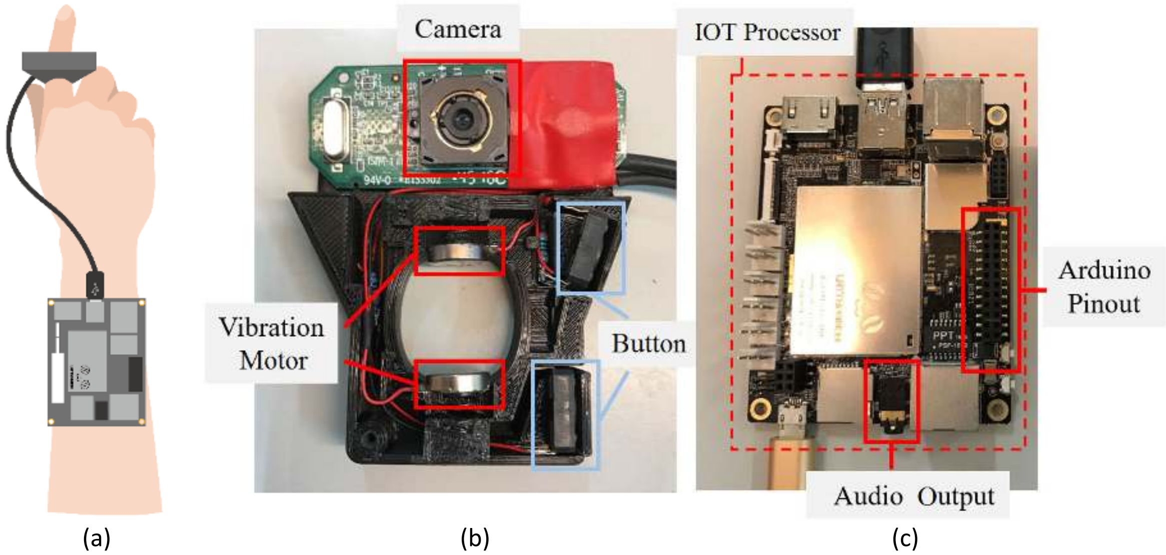


FIGURE 1. **(a) Wearing method of the developed device. (b) Hardware device of the FingerReader. (c) Hardware device of the micro IoT processor.**

FIGURE 2. **Button usage for operating software functions.**

recognition method [16]. They applied several dataset expansion techniques, including random rotation, shear transformation, and elastic deformation within a small range, to expand the scale of available samples. The proposed method was set to achieve high accuracy for testing 10 similar handwritten characters, with 300 samples for each character.

III. INTRODUCTION OF SYSTEM HARDWARE AND FUNCTIONS

To assist visually impaired people in their daily lives, a wearable device, namely a Chinese FingerReader, was designed to read Chinese printed text. The hardware of this device consists of the FingerReader that can be worn on the index finger and a micro IoT processor to be worn on the arm (Fig. 1(a)). The FingerReader on the index finger contains a small camera, two vibration motors, and two buttons (Fig. 1(b)), whereas the micro IoT processor (LattePanda) on the arm with built-in Arduino and audio output (Fig. 1(c)). The camera takes image information to identify the relative position between the index finger and the printed text, and the character image can be further segmented and captured. Two micro vibration motors are placed in the interior of the FingerReader. The micro IoT processor can instantly control the motor vibration and guide the index finger to the appropriate reading position through the vibration information provided. Moreover, two buttons are installed on the FingerReader so that visually impaired users can operate the software functions, as displayed in Fig. 2. In this study, a micro IOT processor was used as the core computation component. This micro IOT processor integrates a built-in microcontroller, Arduino Leonardo, to receive the button state and control the micro vibration motors of the FingerReader. This micro IoT processor is suitable for use as a core operating component and controller in this study because of its small size, low power consumption, and low cost. Finally, two usage scenarios were designed for this study: (a) article reading mode and (b) single-character reading mode. The article reading mode can be applied for reading a book or article under a simple background. The single-character reading mode can be used to identify an individual Chinese character when no

(a) (b) (c)

FIGURE 3. Notification when the user's finger (a) is at the beginning of the text line. (b) shifts upward or downward to a wrong position; (c) is at the end of the text line.

obvious horizontal information of the text line is available. For example, Chinese characters are printed on a label or an item such as a medicine bottle. The details are explained in the following section.

A. ARTICLE READING MODE

To help visually impaired people to read the text in books, the FingerReader includes two micro vibration motors. According to different usage scenarios, the Chinese OCR system guides a visually impaired user to move their finger by controlling the motor vibration. As presented in Fig. 3, three main guiding mechanisms were designed in this study: (a) if the user starts reading, our OCR system provides a notification through the vibration of the two motors when the user's finger is placed at the beginning of the line (Fig. 3(a)); (b) if the user's finger shifts upward or downward to a wrong position when reading (Fig. 3(b)), our system guides their finger back to the original text line according to the vibration guidance of the upper or lower motor; and (c) if the user's finger moves to the end of the text (Fig. 3(c)), our system provides a notification through the vibration of the two motors. The user is then guided to move their finger to the beginning of the next horizontal text line.

B. SINGLE-CHARACTER READING MODE

The single-character reading mode is the best solution for dealing with situations that lack obvious horizontal information of the text line; examples of such situations include those in which the text is printed on a medicine bottle (Fig. 4) or the text is printed under complex background. For such usage scenarios, this study designed a guide label to guide visually impaired users to place their index finger at the correct reading position (Fig. 4). The guide label was pasted to the bottom of the text, and its thickness enabled visually impaired users to easily identify its position by touch. In addition, small vertical marks were placed directly under the guide label corresponding to individual characters. These marks can direct visually impaired users to place their index finger at the correct reading position. Whenever the visually impaired users find a vertical mark with the guide label, they press the button on the FingerReader with their thumb to capture the text image. Our Chinese OCR system then analyzes the captured image to recognize the content of the corresponding character at this reading position

FIGURE 4. Guide label and three vertical marks corresponding to the Chinese characters "維他命" ("vitamin" in English).

IV. SOFTWARE DESIGN OF THE CHINESE OCR SYSTEM

As the FingerReader transmits image information (in front of the fingertip) to the micro IoT processor, this study proposes the Chinese OCR system for identifying the Chinese characters inside the transmitted image. Fig. 5 illustrates the operational flow of the proposed Chinese OCR system. First, several image preprocessing techniques are used to detect the location of the fingertip and the skew angle to the text. The height of the text line is then evaluated. Next, a single-character image is segmented and recognized by the Chinese OCR classifier. Finally, Microsoft's TTS technology is applied to pronounce the Chinese character. This can help visually impaired users to listen to the corresponding voice of the printed Chinese text. The research methods and implementation steps of this system are introduced in the following sections.

A. FINGERTIP DETECTION

After the Chinese OCR system obtains the image captured by the FingerReader, it must first detect the position of the fingertip to extract the correct text according to the fingertip

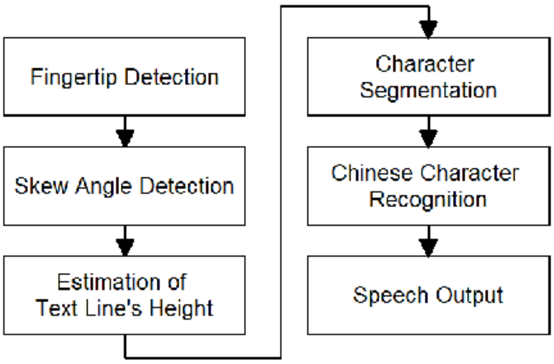


FIGURE 5. The operational flow of the proposed Chinese OCR system.

position (Fig. 6(a)). The first step is to detect the skin region in the captured image. Because the RGB color space is easily affected by the light source, thus causing detection errors, our system uses the YCbCr color space to determine skin color. The skin color range set is as follows:

$$\begin{cases} 98 \leq Cb \leq 142 \\ 133 \leq Cr \leq 177 \end{cases} \tag{1}$$

After skin color detection is used to perceive the skin regions, a size of 3 × 3 structure element is used for opening operations to remove relatively small skin regions and closing operations to fill the broken parts of the skin regions. The largest skin region is marked, and the top of the block is set as the position of the fingertip (Fig. 6(b)). The symbol "+" represents the position of the fingertip. The final detection result is depicted in Fig. 6(c).

B. SKEW ANGLE DETECTION

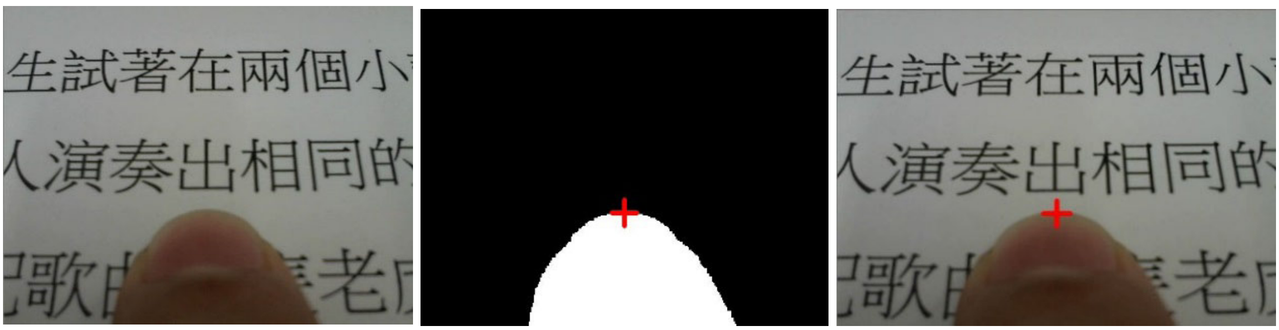
The captured image may appear skewed when a visually impaired user uses the FingerReader because the article or finger angle placement may be skewed. Therefore, after the position of the fingertip is detected, our system first removes the image region of the finger. Only the image above the fingertip is retained for image binarization and used for estimating the skew angle to the text (Fig. 7(a)). Image binarization is performed through Otsu's binarization algorithm. Subsequently, a size of 3 × 15 structure element is used to perform the closing operation of the image, and the text is connected to establish complete and compact horizontal line objects (Fig. 7(b)). Next, the bottom left and bottom right corners of each horizontal line object are identified, and the two positions are joined into a straight line. The skew angle of each straight line is calculated for estimating the skew angle to the text. For example, in Fig. 7(c), the skew angles of the two horizontal line objects are shown to be 4.6° and 5.2°. Finally, the average skew angle of the horizontal line objects is calculated (4.9°), and the image is then rotated by this average skew angle. If the system does not detect any horizontal line object in the captured image, the rotation process is skipped.

C. ESTIMATION OF THE TEXT LINE'S HEIGHT

To estimate the height of the text line, projection scanning is performed according to the horizontal direction of the rotated image (Fig. 8(a)). The gray area in Fig. 8(b) presents text-line blocks identified during the horizontal projection process. The nearest complete horizontal text block is then identified by searching upward from the fingertip position, and the height of this block is set as the height of the text line.

D. CHARACTER SEGMENTATION

In an actual test, the application of the global binarization method for the entire image is easily affected by the light source and other factors, which causes a poor binarization quality for the character images and affects the accuracy of character segmentation and OCR classification. Therefore, local binarization was applied to obtain a suitable binarization quality for the character images. Otsu's binarization algorithm was applied for the red rectangular area in Fig. 9(a), which was directly above the fingertip. The height and width of the rectangular area were twice and four times the height of the text line, respectively. Figure 9(b) illustrates the image binarization results for the red rectangular image. In this rectangular image, each individual Chinese character was segmented using the vertical projection method (Fig. 9(c)). The Chinese character "同" ("same" in English) nearest to the fingertip position was then recognized by the



(a) (b) (c)

FIGURE 6. Fingertip detection. (a) Image captured by the FingerReader. (b) The symbol "+" represents.

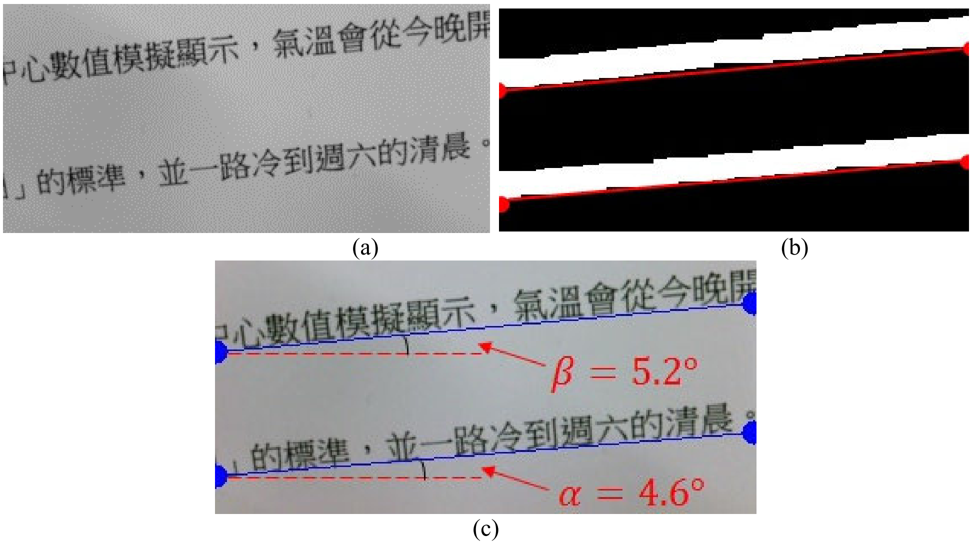


FIGURE 7. (a) The image above the fingertip is used for estimating the skew angles; (b) text lines are connected to establish horizontal line objects; (c) the skew angles of the horizontal line objects are calculated.

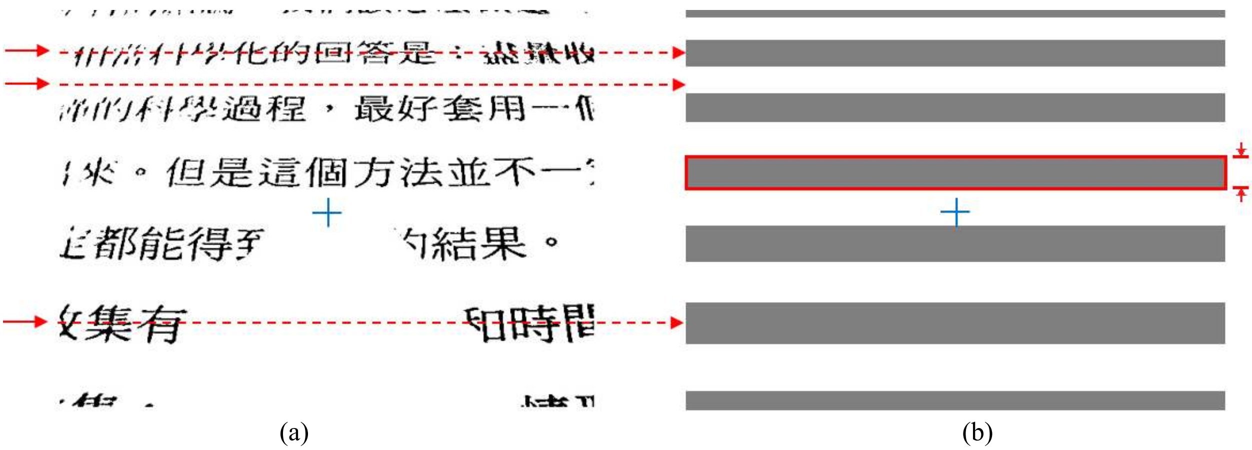


FIGURE 8. (a) Detection of text-line blocks by projection scanning from the horizontal direction; (b) estimating the height of the text line.

OCR classifier. Because Chinese characters generally have a square appearance, the height of the text line was referred to when the Chinese character was segmented.

E. CHINESE CHARACTER RECOGNITION AND SPEECH OUTPUT

In this study, the training strategy of an ACNN algorithm was applied to train the OCR classifier to identify Chinese printed characters. This training strategy can reduce training time and address the problem of skewed characters. The training dataset contained 532,597 Chinese character images, including 6,000 classes of various fonts in traditional Chinese characters and punctuation marks. A testing accuracy of 99.2% was obtained. The training strategy of the ACNN algorithm is explained in the following section. After the identification of the Chinese characters, Microsoft's TTS technology was used to pronounce the Chinese character. This enabled visually impaired people to listen to the corresponding voice of the printed Chinese text.

V. CONSTRUCTION OF THE OCR CLASSIFIER WITH THE ACNN ALGORITHM

In this study, the ACNN algorithm was used for training the OCR classifier. The CNN algorithm was implemented by applying Google's TensorFlow [18], an open-source program provided by Google Brain. This program is supposed to be flexible for research purposes while also allowing its models to be deployed productively. The training process was divided into two stages. In the first training stage, the dataset contained 532,597 Chinese character images for constructing the first-stage CNN classifier. The image size was 28×28 pixels. In the second training stage, 2,734,850 Chinese character images with different rotation angles were applied to fine tune the parameters of the first-stage CNN classifier. Figure 10 shows the architecture of the proposed CNN classifier. The first input layer had a size of 28×28. The C2 and C4 convolutional layers used 32 and 64 kernel filters of size 5×5, respectively. The max pooling (P3 and P5) layers applied in this study took the output of the convolution and

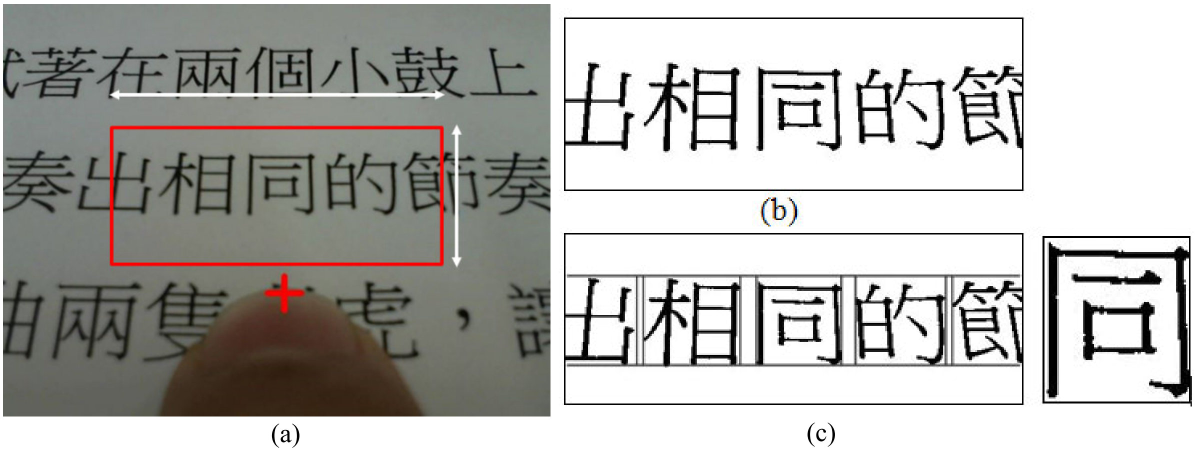


FIGURE 9. **(a) Rectangular image directly above the fingertip; (b) binarization results of this rectangular image; (c) segmentation of the Chinese character "同" nearest to the fingertip position.**

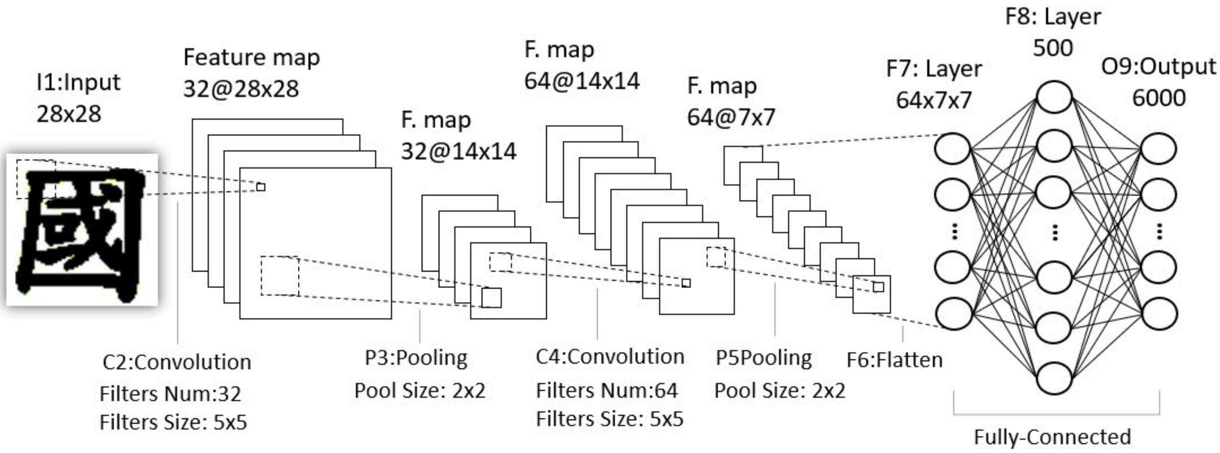


FIGURE 10. **The architecture of the CNN classifier applied in this study.**

split it up into tiles. The tiles were selected to have a size of 2 × 2 pixels with stride 2, and the largest value from each tile was used in the next layer of the network. After flattening (F6), the fully connected neural network contained 3,136 neurons in the F7 input layer and 500 neurons in the F8 hidden layer. We applied cross entropy as the loss function and the Adam optimizer to minimize the loss function. The batch size was set as 1,000. The initial learning rate was set as 0.01, and decreased by multiplying 0.1 after every 400 training iterations.

In the first training stage, the dataset contained 532,597 Chinese character images, including 6,000 classes of various fonts in traditional Chinese characters and punctuation marks. Figure 11 presents some examples of Chinese character images. The data patterns of the first training stage were further partitioned into a training dataset and testing dataset. A total of 481,557 and 51,400 character images were used as the training and testing datasets, respectively. The training and testing processes were performed on a personal computer with a GTX 1080Ti graphics card using the GPU version of TensorFlow. The total computation time for training the first-stage CNN classifier was 13.5 h, and the first-stage CNN classifier achieved accuracy rates of 99.9% and 99.2% for the training and testing datasets, respectively.

To further test the accuracy of the OCR classifier application with the Chinese FingerReader, the FingerReader was used to capture 400 Chinese characters (Fig. 12). This FingerReader testing dataset (FRT dataset) comprised four fonts with 100 classes of Chinese characters. The FRT dataset was tested by our CNN classifier with an accuracy rate of 96.75%. For comparison, the FRT dataset was also tested using the OCR APIs of Google and Microsoft, with accuracy rates of 93.42% and 96.25% being obtained, respectively (Table I). The testing results indicate that our CNN classifier outperformed the OCR APIs of Microsoft and Google.

In actual use, visually impaired people can not accurately target the text for capture due to their visual impairment, which causes the captured character images to be rotated and skewed. Therefore, the previously trained CNN classifier could not correctly identify the classes of characters. To solve this problem, the original image dataset was augmented, and each character image was rotated by +10°, +20°, −10°, and

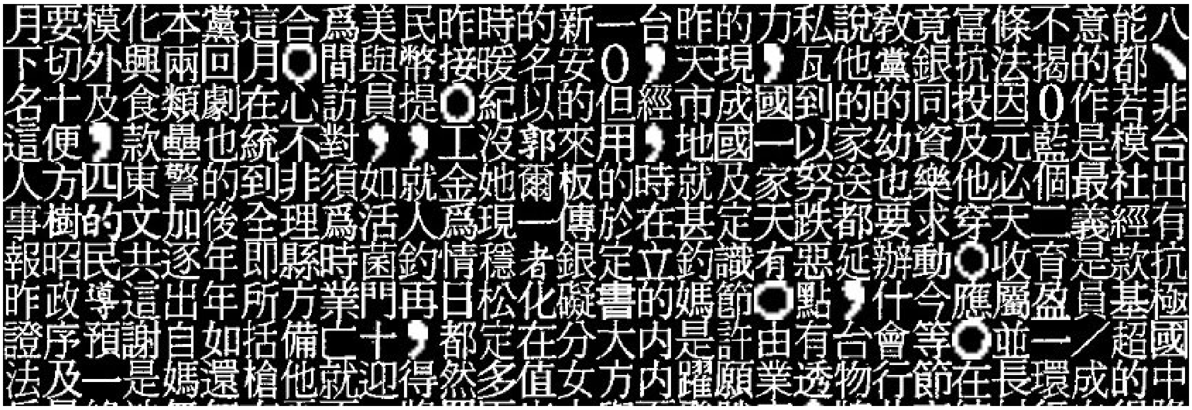


FIGURE 11. Some examples of Chinese character images in this study.

FIGURE 12. Some examples of the FRT dataset.

TABLE 1. Classification accuracy rates of three OCR classifiers for the FRT dataset.

	ACCURACY RATE FOR TESTING FRT DATASET
Our CNN Classifier	96.75%
Google OCR API	93.42%
Microsoft OCR API	96.25%

−20° (Fig. 13). Finally, a total of 2,734,850 rotated character images were created. Subsequently, total of 2,477,850 and 257,000 rotated character images were used as the training and testing dataset. In this second training stage, the created character images with different rotation angles were used to fine tune the parameters of the first-stage CNN classifier. The learning rate was set to an initial value of 0.001 and gradually decreased during training. The batch size was set to 10,000, and the total training time was 28 h. This second-stage CNN classifier was named the ACNN OCR classifier. The ACNN OCR classifier achieved accuracy rates of 99.8% and 98.8% for the training and testing datasets, respectively.

FIGURE 13. Some examples of rotated Chinese character images.

For comparison, each character image of the FRT dataset was also rotated by ±10° and ±20° (Fig. 14). Finally, the new FRT dataset contained 2,000 rotated Chinese character images for testing. The accuracy rates of the ACNN OCR classifier and the Google and Microsoft OCR APIs were then compared again. Table II lists the accuracy rates of the three classifiers. The accuracy rate of the ACNN OCR classifier decreased marginally to 95.25%, whereas the accuracy rates of the Google and Microsoft OCR APIs were 64.74% and 34%, respectively. This finding reveals that compared with the Google and Microsoft OCR APIs, the ACNN algorithm achieved better performance in dealing with rotated character images, demonstrating its superior conformity to the needs of visually impaired people in actual use.

To further analyze the performance of the ACNN OCR classifier, the character images of the FRT dataset were rotated by different angles in the range of −25° to +25°. The testing results under various rotation angles are presented in Table 3. All the observed accuracy rates

FIGURE 14. Each character image of the FRT dataset was rotated by ±10° and ±20°.

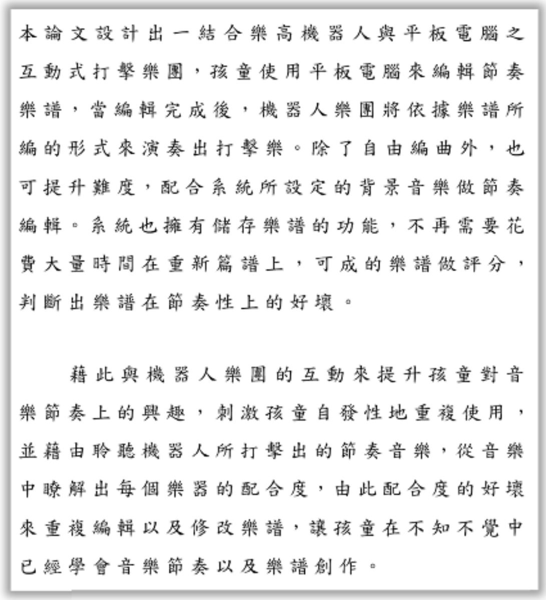
TABLE 2. Classification accuracy rates of three OCR classifiers for testing the rotated character images in new FRT dataset.

	Accuracy Rate for Testing New FRT Dataset
ACNN OCR Classifier	95.25%
Google OCR API	64.74%
Microsoft OCR API	34%

TABLE 3. The accuracy rate of the augmented CNN classifier for FRT dataset under various rotated angles.

Rotation Angle	Accuracy Rate	Rotation Angle	Accuracy Rate
0°	97.00%		
1°	97.00%	**-1°**	96.00%
2°	97.00%	**-2°**	96.75%
3°	97.50%	**-3°**	97.00%
4°	96.75%	**-4°**	96.75%
5°	97.50%	**-5°**	97.50%
6°	97.50%	**-6°**	96.75%
7°	97.50%	**-7°**	96.50%
8°	97.75%	**-8°**	97.25%
9°	98.00%	**-9°**	96.75%
10°	96.75%	**-10°**	96.00%
11°	97.50%	**-11°**	95.50%
12°	97.50%	**-12°**	96.50%
13°	97.75%	**-13°**	96.50%
14°	97.25%	**-14°**	96.00%
15°	97.25%	**-15°**	96.25%
16°	97.75%	**-16°**	95.00%
17°	97.25%	**-17°**	95.00%
18°	97.25%	**-18°**	95.50%
19°	96.75%	**-19°**	95.00%
20°	96.00%	**-20°**	95.75%
21°	94.75%	**-21°**	93.75%
22°	95.50%	**-22°**	92.50%
23°	95.00%	**-23°**	91.75%
24°	95.25%	**-24°**	89.50%
25°	94.50%	**-25°**	87.0%

exceeded 87% (Table III). This experimental result verifies the robustness of the proposed method in addressing rotated character images.



本論文設計出一結合樂高機器人與平板電腦之互動式打擊樂團，孩童使用平板電腦來編輯節奏樂譜，當編輯完成後，機器人樂團將依據樂譜所編的形式來演奏出打擊樂。除了自由編曲外，也可提升難度，配合系統所設定的背景音樂做節奏編輯。系統也擁有儲存樂譜的功能，不再需要花費大量時間在重新篇譜上，可成的樂譜做評分，判斷出樂譜在節奏性上的好壞。

藉此與機器人樂團的互動來提升孩童對音樂節奏上的興趣，刺激孩童自發性地重複使用，並藉由聆聽機器人所打擊出的節奏音樂，從音樂中瞭解出每個樂器的配合度，由此配合度的好壞來重複編輯以及修改樂譜，讓孩童在不知不覺中已經學會音樂節奏以及樂譜創作。

FIGURE 15. Document used for the field test.

VI. EXPERIMENT AND FIELD TEST

In this study, the experimental evaluation of the proposed approach was assisted by the Resource Center for the Visually Impaired at Tamkang University. This center provides academic, vocational, emotional, and life assistance to visually impaired students. The center also develops and promotes comprehensive information systems for such students. The field test was conducted on a 22-year-old visually impaired male student. The informed consent was signed by the visually impaired student. After discussions with visually impaired people, this study compiled three scenarios in which visually impaired people would require a Chinese Finger-Reader in their daily lives.

(1) Article reading mode: This mode can assist visually impaired people in reading newspapers, magazines, textbooks, lecture notes, or exam questions.
(2) Single-character reading mode: People with total blindness can read the text on medicine bottles or products with the help of guide labels. In addition, people with amblyopia or partial vision can identify text using this mode for activities such as interpreting announcements on bulletin boards or searching for books in the library.
(3) English letters and numeral reading mode: English character and number recognition functions were also included in the Chinese FingerReader. The test results of the three usage scenarios are provided in the following sections.

A. FIELD TEST OF THE ARTICLE READING MODE

The article reading mode was mainly developed to help visually impaired people read the text in a book. In this experiment, the Chinese FingerReader was used to identify

(a)

(b)

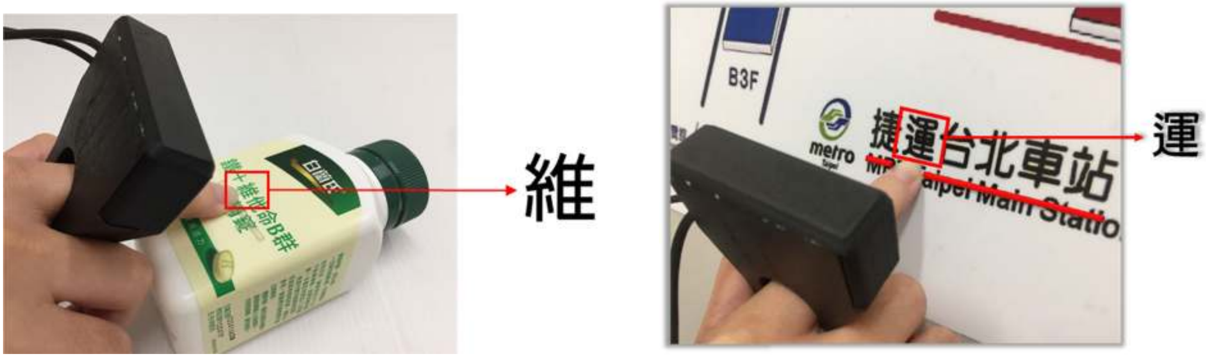
FIGURE 16. **Images with (a) a complex background and (b) vertical lines.**

each of the characters in the article depicted in Fig. 15. This article contained 277 Chinese characters, and the total identification time was 1,110 s (i.e., approximately 4 s to recognize a Chinese character). The OCR accuracy rate was 99.3%. The identification time included the time required for finger movement, the computational time of the Chinese OCR system, and the time required for speech feedback. The average computational time, including the time required for detecting the fingertip and skew angle, segmenting the character, and character recognition, of the micro IoT processor was 0.4 s. In this scenario, the ability to accurately detect the position of the text line and estimate the height of the text line was the most crucial factor affecting the identification performance. The height of the text line was used to segment the character image. It also helped the system to determine whether the index fingertip position was maintained at the bottom of the text line.

During testing, a user had to constantly move their finger to capture the character image. However, the camera lens might be severely shaken during finger movement, which might result in the lens capturing blurred image characters, which in turn might affect the accuracy of the ACNN OCR classifier. Furthermore, our system was completely unusable in some scenarios. An article with a complex background (Fig. 16(a)) can cause substantial noise in the background image after image binarization. For such an article, our method could not separate the text and background images; that is, the text line could not be detected and the character image could not be segmented. Furthermore, as displayed in Fig. 16(b), although the image background was not complex, some vertical lines appeared in the image after binarization, which led to an inability to accurately detect the position of the text line when the system performed projection scanning in a horizontal direction. Thus, our system was rendered unusable in this scenario.

B. FIELD TEST OF THE SINGLE-CHARACTER READING MODE

The single-character reading mode was mainly developed to help visually impaired people read the text on medicine bottles or products with the help of guide labels. In addition, people with amblyopia or partial vision can identify text using this mode for activities such as interpreting announcements on bulletin boards or searching for books in the library. Because visually impaired people are unaware of a character's position, guide labels were designed to guide the users to move their fingers for accurately capturing the character image (Fig. 17). When visually impaired people wanted to read the text on medicine bottles, they could easily locate the position of the text by touching the guide label and placing their finger on the position of the character, as displayed in Fig. 17(a). Moreover, guide labels are useful for application to complex image background usage scenarios, such as guide maps at transport stations. In general, most of these maps are not designed for visually impaired people. A transparent guide label was pasted onto the position of the red line in the guide map, as presented in Fig. 17(b), so that visually impaired people could read the text on the map along the guide label. People with amblyopia or partial vision might not require assistance with guide labels. Such people can simply place their fingers under the text and press the button to directly recognize the content of the corresponding character at the reading position. Although the guide label system was convenient to use, classification errors occurred when the size of the character exceeded the camera capture range. In this situation, our system could not capture complete



(a) (b)

FIGURE 17. **Reading characters according to guide labels. (a) Text on a medicine bottle. (b) Text on a guide map.**

FIGURE 18. **Oversized character.**

FIGURE 19. **Reading the numbers and Chinese characters on microwave buttons.**

character images for recognition (yellow rectangular area in Fig. 18).

C. FIELD TEST FOR THE ENGLISH LETTERS AND NUMERAL READING MODE

The text printed on many electrical products used in daily life is in English or in the form of Arabic numerals. Thus, English letter and numeral recognition functions were also included in the Chinese FingerReader to allow visually impaired people to identify English letters and numbers. As displayed in Fig. 19, guide labels were pasted (red line position) under the text of microwave buttons to notify visually impaired people of the text position. This enabled the visually impaired people to use the Chinese FingerReader to read the text and numbers on a microwave and helped them to avoid the danger caused by the incorrect operation of electrical appliances.

D. COMPARISON WITH RELATED STUDIES

The related researches applied dataset expansion techniques, however, these researches only focused on handwritten Chinese character recognition under the small-scale datasets [14]–[16]. By contrast, we focused on identifying Chinese printed text. Moreover, in a field trial, visually impaired people could not accurately target text intended for capture due to their visual impairment, which could cause the captured character images to be rotated and skewed. Therefore, in this study, we proposed the ACNN algorithm to construct OCR classifier instead of SVM to achieve best performance. The training strategy ACNN can reduce the training time and address the problem of skewed characters. Furthermore, this study used a large-scale dataset, which contained 2,734,850 Chinese character images, including 6000 classes of various fonts in traditional Chinese characters and punctuation marks in the training process. In additions, the field trials were given for three scenarios in which visually impaired people would require a Chinese FingerReader in their daily lives. The experimental results in Section 5 also revealed that compared with the OCR application programming interfaces (APIs) of Google and Microsoft, our OCR classifier could provide better performances for dealing with rotated character images, demonstrating its superior conformity to the needs of visually impaired people in actual use.

E. ACTUAL FIELD TEST OF VISUALLY IMPAIRED PEOPLE WITH TOTAL BLINDNESS

To verify the practical feasibility of the Chinese Finger-Reader, a field test was conducted on a 22-year-old visually impaired male student, who served as the participant, and two special education experts. The opinions of the visually impaired user and special education experts are organized as follows:

(a) In the single-character reading mode, the participant could smoothly read the Chinese characters on a medicine bottle by touching and following the guide label. In addition, a trademark with a guide label attached was designed, and the participant could successfully identify the product name on the trademark by using this mode.

(b) The participant stated that the Chinese FingerReader was useful for reading text while outdoors. For example, the participant could purchase products at a store or search for a book in a library without the assistance of others.

(c) The participant had previously attempted to use OCR software on a mobile phone. However, using the software with poor vision resulted in images being captured through an inappropriate focal length or the text being skewed, thus rendering the text unreadable by the software. The skew correction function of the Chinese FingerReader could solve this reading problem experienced by visually impaired people.

(d) Although the Chinese FingerReader required 0.4 s to identify a Chinese character (including the time required for detecting the fingertip and skew angle, segmenting the character, and character recognition), the time is still within an acceptable range. In addition, the approach of guiding visually impaired people to read articles through hints from two vibration motors was found to be feasible and practical.

(e) During the test, the participant commented that the Chinese FingerReader and micro IoT processor were not excessively heavy. The weight of the Chinese FingerReader and micro IoT processor did not affect the participant's willingness to use the device.

(f) When using the Chinese FingerReader at the beginning, the participant tended to bend his index finger, preventing the FingerReader from being able to capture and detect his fingertip. After an explanation was provided, the participant was able to master the use of the FingerReader immediately after straightening his index finger.

VII. CONCLUSION

In this study, the Chinese FingerReader is developed for visually impaired people to recognize Chinese printed text. The device is portable, easy to operate, and designed to be worn on the index finger. A micro IoT processor serves as the computing core to recognize Chinese characters, English letters, and numbers. The training strategy of an ACNN algorithm was applied to train the OCR classifier to identify Chinese printed characters, which could reduce the training time and address the problem of skewed characters. The experimental results revealed that compared with the Google and Microsoft OCR APIs, the ACNN algorithm achieved better performance in dealing with rotated character images, demonstrating its superior conformity to the needs of visually impaired people in actual use. Finally, three usage scenarios were simulated, and the field test illustrated that the Chinese FingerReader was useful for reading text while outdoors. For example, the participant could purchase products at a store or search for a book in a library without the assistance of others.

ACKNOWLEDGMENT

The authors would like to thank the Resource Center for the Visually Impaired of Tamkang Universityfor field trials and valuable suggestions during developing the prototype.

REFERENCES

[1] C. Yi and Y. Tian, "Assistive text reading from complex background for blind persons," in *Camera-Based Document Analysis and Recognition*. Beijing, China: Springer, 2012, pp. 15–28.

[2] S. M. Hanif and L. Prevost, "Texture based text detection in natural scene images—A help to blind and visually impaired persons," in *Proc. CVHI*, 2007, pp. 1–6.

[3] M. Mattar, A. Hanson, and E. Learned-Miller, "Sign classification using local and meta-features," in *Proc. IEEE CVPR Workshops*, Sep. 2005, p. 26.

[4] N. Ezaki, M. Bulacu, and L. Schomaker, "Text detection from natural scene images: Towards a system for visually impaired persons," in *Proc. ICPR*, vol. 2, Aug. 2004, pp. 683–686.

[5] J.-P. Peters, C. Thillou, and S. Ferreira, "Embedded reading device for blind people: A user-centered design," in *Proc. IEEE ISIT*, Oct. 2004, pp. 217–222.

[6] H. Shen and J. M. Coughlan, "Towards a real-time system for finding and reading signs for visually impaired users," in *Proc. ICCHP*, 2012, pp. 41–47.

[7] A. Guo, X. A. Chen, H. Qi, S. White, S. Ghosh, C. Asakawa, and J. P. Bigham, "VizLens: A robust and interactive screen reader for interfaces in the real world," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.*, 2016, pp. 651–664.

[8] H. Kandemir, B. Cantürk, and M. Ba tan, "Mobile reader: Turkish scene text reader for the visually impaired," in *Proc. 24th Signal Process. Commun. Appl. Conf.*, May 2016, pp. 1857–1860.

[9] Z. Liu, Y. Luo, J. Cordero, N. Zhao, and Y. Shen, "Finger-eye: A wearable text reading assistive system for the blind and visually impaired," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot.*, Jun. 2016, pp. 123–128.

[10] H. Jiang, T. Gonnot, W.-E. Yi, and J. Saniie, "Computer vision and text recognition for assisting visually impaired people using Android smartphone," in *Proc. IEEE Int. Conf. Electro Inf. Technol.*, May 2017, pp. 350–353.

[11] R. Shilkrot, J. Huber, C. Liu, P. Maes, and S. C. Nanayakkara, "Finger-Reader: A wearable device to support text reading on the go," in *Proc. ACM CHI*, 2014, pp. 2359–2364.

[12] R. Shilkrot, J. Huber, M. E. Wong, P. Maes, and S. C. Nanayakkara, "FingerReader: A wearable device to explore printed text on the go," in *Proc. ACM CHI*, 2015, pp. 2363–2372.

[13] Y. LeCun, L. D. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Müller, E. Sackinger, P. Simard, and V. Vapnik, "Learning algorithms for classification: A comparison on handwritten digit recognition," *Neural Netw., Stat. Mech. Perspective*, vol. 261, p. 276, Jan. 1995.

[14] Y. Tang, L. Peng, Q. Xu, Y. Wang, and A. Furuhata, "CNN based transfer learning for historical Chinese character recognition," in *Proc. 12th IAPR Workshop Document Anal. Syst. (DAS)*, vol. 8, Apr. 2016, pp. 25–29.

[15] L. Liu, P.-L. Yang, W.-W. Sun, and J.-W. Ma, "Similar handwritten chinese character recognition based on CNN-SVM," in *Proc. Int. Conf. Graph. Signal Process.*, 2017, pp. 16–20.

[16] X. Song, X. Gao, Y. Ding, and Z. Wang, "A handwritten Chinese characters recognition method based on sample set expansion and CNN," in *Proc. 3rd Int. Conf. Syst. Inform. (ICSAI)*, Shanghai, China, Jan. 2017, pp. 843–849.
[17] Y. L. Chu, H. E. Hsieh, W. H. Lin, H. J. Chen, and C. H. Chou, "Chinese FingerReader: A wearable device to explore Chinese printed text," in *Proc. ACM SIGGRAPH*, 2017, Art. no. 54.
[18] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467

YU-SHENG SU received the Ph.D. degree from the Department of Computer Science and Information Engineering, National Central University, Taiwan, in 2010. He is currently an Assistant Professor with the Department of Computer Science and Engineering, National Taiwan Ocean University, Taiwan. His research interests include AIoT applications, social media mining, document analysis and recognition, cloud computing, and computational thinking.

CHIEN-HSING CHOU received the B.S., M.S., and Ph.D. degrees from the Department of Electrical Engineering, Tamkang University, Taiwan, in 1997, 1999, and 2003, respectively, where he is currently an Associate Professor of electrical engineering. His research interests include image analysis and recognition, mobile phone programming, machine learning, document analysis and recognition, and clustering analysis.

YUNG-LONG CHU received the M.S. degree from the Department of Electrical Engineering, Tamkang University, Taiwan, in 2014, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include image analysis and recognition, and mobile phone programming.

ZHAO-YU YANG received the M.S. degree from the Department of Electrical Engineering, Tamkang University, Taiwan, in 2018. His research interests include image analysis and recognition, and machine learning.