Image processing technology for text recognition

Yen-Min Su[1], Hsing-Wei Peng[1], Ko-Wei Huang[1] and Chu-Sing Yang[2]
[1]Department of Electrical Engineering
National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan
[2]Institute of Computer and Communication Engineering, Department of Electrical Engineering,
National Cheng Kung University, Tainan, Taiwan.
1105304163@nkust.edu.tw, f107154130@nkust.edu.tw, elone.huang@nkust.edu.tw, csyang@ee.ncku.edu.tw

Abstract— **Image recognition and optical character recognition technologies have become an integral part of our everyday life due in part to the ever-increasing power of computing and the ubiquity of scanning devices. Printed documents can be quickly converted into digital text files through optical character recognition and then be edited by the user. Consequently, minimal time is required to digitize documents; this is particularly helpful when archiving volumes of printed materials. This study demonstrates how image-processing technologies can be used in combination with optical character recognition to improve recognition accuracy and to improve the efficiency of extracting text from images. Two software systems are developed and tested during this study: a character recognition system applied to cosmetic-related advertising images and a text detection and recognition system for natural scenes. The results of the experiment demonstrate that the proposed systems can accurately recognize text in images.**

Keywords-Image processing, optical character recognition, object detection component

I. INTRODUCTION

Attempts have long been made to design computer programs that can read printed documents with the objective of improving archiving efficiency by converting documents into electronic files in an automated manner. Systems capable of recognizing text in images and converting it into characters for editing on a computer are known as optical character recognition (OCR) systems [1]. OCR was first proposed by the German scientist Tausheck in 1929. Since the 1960s, scientists worldwide have sought to improve OCR using computers. Early OCR research was focused on identifying the numerals 0–9. The earliest research on recognition of printed Chinese characters was conducted by Casey and Nagy, who published their first paper on Chinese character recognition in 1966; this paper details the successful identification of 1,000 printed Chinese characters by using a template matching technique.

In this paper, we discuss our development of two OCR-based systems: a character recognition system for commercial advertising images and a text detection and recognition system for natural scenes. After the basic character recognition system is completed, we will integrate an improper words detection system to preemptively reduce the number of legal disputes that can arise from using inappropriate words in advertisements. The main purpose of the text recognition system to be applied to natural scenes is to assist managers in archiving documents. These two systems will both be installed on a Raspberry Pi development board.

The remainder of this paper is structured as follows. Section II provides an overview of related literature. Section III describes our research method. In Section IV, the analytical results of the study are presented. Finally, Section V presents our conclusions and points to recommended directions for future development.

II. LITERATURE REVIEW

A. Optical character recognition

As previously mentioned, efficient OCR systems have been an important aspect of computer program design for many years. A common tool currently used for character recognition is Tesseract-OCR, developed by Hewlett Packard (HP) between 1984 and 1994. Not long after it was released, HP discontinued its development and it released as open source in 2005; Google then obtained the source code and began to expand and optimize it. Google currently publishes Tesseract-OCR as an open source project as part of the Google Project.

B. Image preprocessing

Image quality of pictures as they are captured can be adversely affected by numerous external factors to include insufficient light, blurring, and insufficient resolution. Image preprocessing techniques are often employed to filter noise or enhance image quality. Additionally, although every image has a background and foreground [2], either can be highlighted as required by using image-preprocessing techniques [3][4]. Common image preprocessing techniques include grayscale, binarization, fuzzification, edge detection, and erosion and dilation.

C. Raspberry Pi

The Raspberry Pi development board was developed by the Raspberry Pi Foundation in the United Kingdom initially to promote basic computer literacy in schools. It was widely acclaimed and has been extensively applied in various fields since its launch. Raspbian is a Debian-based Linux distribution that provides a user-friendly interface for the Raspberry Pi [5], with all software suites specifically recompiled for the Raspberry Pi.

III. RESEARCH METHODS

This section is organized in two subsections. Section III-A introduces the first system proposed in this study, namely the character recognition system for cosmetic-related advertising

978-1-7281-4666-9/19/$31.00©2019 IEEE

images, and Section III-B describes the second system, the text detection and recognition system for natural scenes.

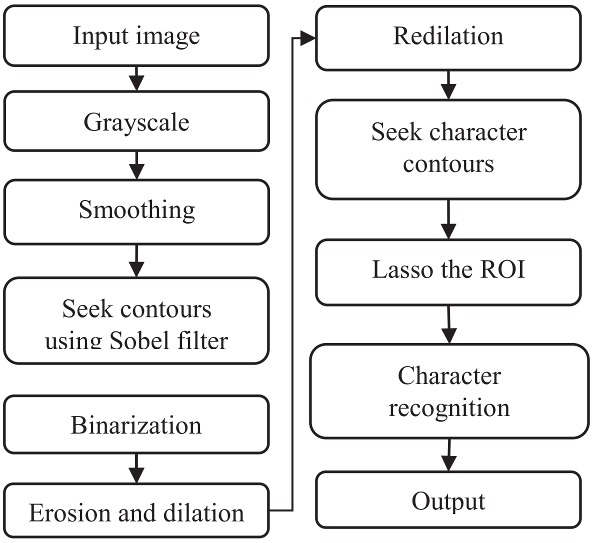

Figure 1. Schematic diagram of the first system

A. *Character recognition system for cosmetic-related advertisement images*

A schematic of the first system is displayed in Figure 1. Figure 2 shows the advertising image selected as the test image for our experiment. To identify the text in the image, the system first extracts all text from the image through image preprocessing, after which the recognition process begins.

Figure 2. Test image (www.senshido.com.tw/)

1) *Image preprocessing*

The image preprocessing techniques employed in this study are edge detection, binarization, and erosion and dilation, detailed as follows.

a) Edge detection

Edge detection is a commonly used technique for image recognition and computer vision. It is used to detect meaningful and discontinuous spots in the grayscale of an image. Through edge detection, the image is sliced into several discrete regions [6]. Common edge detection algorithms include Sobel filtering, Laplacian filtering, and Canny edge detection filtering [7]. We employed the Sobel filtering algorithm in this study. Figure 3 displays the result of binarization after performing edge detection using Sobel filtering.

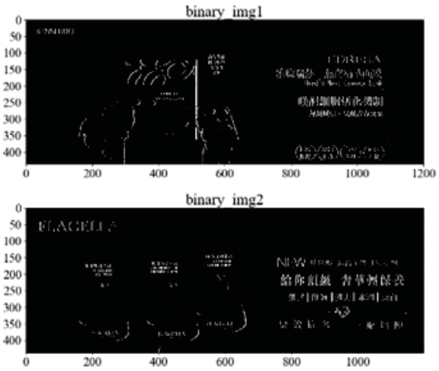

Figure 3. Image after edge detection and binarization

b) Erosion and dilation

Erosion and dilation are two techniques often employed for morphological image processing to process and analyze the shapes in an image. Through erosion and dilation, subtle noise in the image can be removed while retaining the outer contour of the text area. Figure 3-4 shows the noise removal result after applying erosion and dilation.

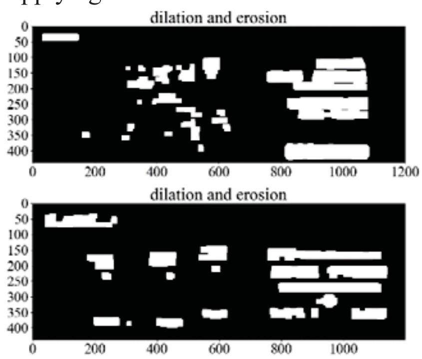

Figure 4. Image after erosion and dilation

2) *Establishing contours and lassoing the region of interest (ROI)*

After the image preprocessing has been completed, the contours of the text area are identified, and the ROI is lassoed. Figure 5 illustrates the results of ROI lassoing after image preprocessing.

Figure 5. Image after lassoing the ROI

3) Character recognition

Upon completing the pre-operations, character recognition is performed on the text area lassoed in the test image. Character recognition is the process of recognizing individual words from an image. This study used Tesseract-OCR, an open source tool released by Google Project, in combination with Python for text recognition.

B. Text detection and recognition system for natural scenes

The second system proposed in this study detects text in natural scenes by using a camera connected to a Raspberry Pi. The detected text is automatically input into a database, replacing the manual key-in of text by workers. This reduces the labor required for the job. A schematic of the processes involved in the second system is presented in Figure 6.

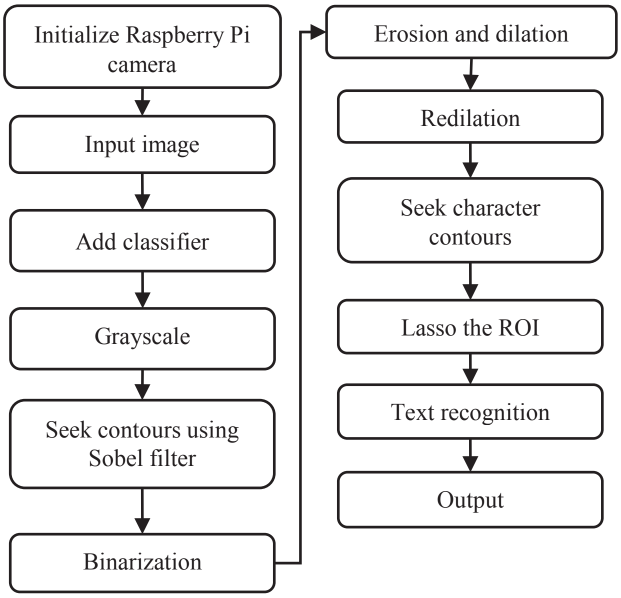

Figure 6. Schematic diagram of the second system

1) Operating the Raspberry Pi camera and detecting the target object containing text

The Cascade Classifier provided by OpenCV [8] was used to detect target objects, in our study the object were texts. The classifier has a pretrained feature model that can detect faces, eyes, and pedestrians, as illustrated in Figure 7. The feature model is established through training on hundreds of positive samples with identical attributes (e.g., a human face or car). These samples are resized to specific dimensions (i.e., 20 × 20). Additionally, the same operations are performed on negative samples—images not containing the target attributes. The greater the difference between the positive and negative samples, the better. Once the training is complete, the classifier can be used to detect objects identical to the positive samples. The detected object is identified in a marquee. This study uses a trained classifier to perform object detection, as illustrated in Figure 8.

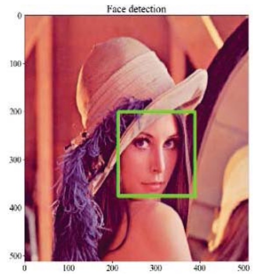

Figure 7. Detection result

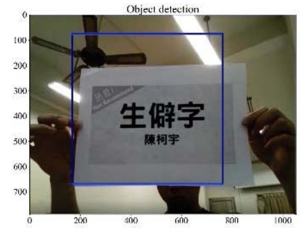

Figure 8. Detection result

2) Image preprocessing

After detecting an object containing text, image preprocessing is performed as described in Sections III-A.1 through III-A.3. The results are presented in the results and analysis section.

IV. RESULTS AND ANALYSIS

This study used Python, Tesseract OCR, and a Raspberry PI 3 for software development. Advertising images and images from the ICDAR Robust Reading Competition were used as test images. In Section IV-A, we first discuss the results and analysis of the pretest of the core system. In Section IV-B, the research results and analysis of the first system are presented. The results of the second system are discussed in Section IV-C.

A. System pretest

This section describes the use of images from the ICDAR Robust Reading Competition [9] as test images for system pretesting. Figure 9 shows the images obtained from the ICDAR Robust Reading Competition. Table I presents the results of recognition using the method proposed in this study.

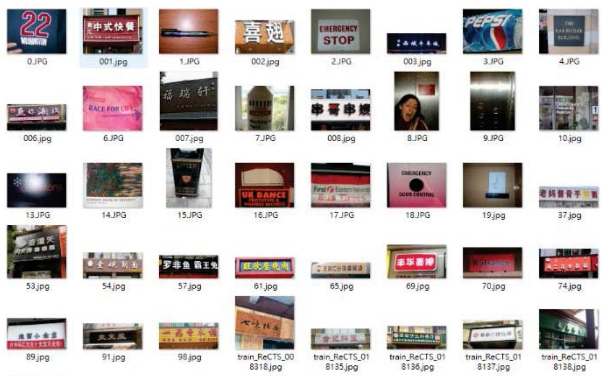

Figure 9. Test images

TABLE I. RECOGNITION RESULTS

	Test image 1	*Test image 2*
Original image		
Lassoed ROI		
Recognition results	EMERGENCY DOOR CONTROL	EMERGENCY STOP

To understand the effectiveness of the preprocessing, the system was assessed in terms of its precision, recall, and F-measure. The results of the assessment are presented in Table II.

TABLE II.
RECOGNITION RESULTS OBTAINED USING DIFFERENT SYSTEMS

Method used	*Accuracy*	*Recall*	*F1-measure*
This study	93.44%	79.16%	85.71%
Zhong et al[10]	82.45%	84.23%	83.33%
TextFlow[11]	78.40%	84.70%	81.40%
Yin et al[12]	82.60%	68.50%	74.60%

The method proposed in this study resulted in satisfactory recognition. Consequently, we extended the method to develop a character recognition system for use with commercial advertising images and a text detection and recognition system for natural scenes. Further description of these systems is provided in Sections IV-B and IV-C, respectively.

B. *Results obtained using the first system*

An advertising image was employed as a test image for experimentation with the first system. Table III presents the recognition results obtained using the method proposed herein.

TABLE III. RECOGNITION RESULTS OF THE FIRST SYSTEM

Images	*1*
Original image	
ROI Regions	
Recognition results	SENSHIDO EDRESA 內含多醞植物斛峭及肥窩馴睢 雅德瑞莎 燕窩精華面膜 Bird's Nest Essence Mask 喚醒細胞活化機制 舒緩賦活 ' 呈現青春光采 鎖水有癱 延緩保濕

C. *Results of the second system*

The second system was employed to recognize text in natural scenes. The Cascade Classifier provided by OpenCV was trained using the image set downloaded from ImageNET. then detection was performed for the target object. The detection results are displayed in Figures 10 and Figure 11.

Figure 10. Detection result

Figure 11. Detection result

Finally, the captured images were extracted using edge detection, binarization, and erosion and expansion. The extracted text area then underwent text recognition. The results of this recognition are presented in Table IV.

TABLE IV. RECOGNITION RESULTS OF THE SECOND SYSTEM

	Test image 1	***Test image 2***
Original image		
Lassoed ROI		
Characters recognized	白話深度學習 與 TensorFlow	生僻字 陳柯宇

V. CONCLUSIONS AND FUTURE RESEARCH

We developed two recognition systems based on image processing and OCR technology. An image set from the ICDAR Robust Reading Competition and photographs taken of our surroundings were used as our test data set. The image-preprocessing algorithm was used to test the data set and highlight the text areas in the test data; subsequently, Tesseract-OCR was employed to recognize text. The recognition results reveal that the proposed method was highly effective at extracting text from images and had a 93% accuracy rate. The accuracy rate was higher when recognition was performed on numbers and English letters.

In future research, we expect to continually optimize the character recognition system for commercial advertising images, increasing its recognition rate and accuracy. Furthermore, by integrating the system with an improper word detection system, we aim to develop the system into a timesaving, easy-to-use, and inclusive improper word detection system that can detect improper words not only in text-based documents but also within images.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Science and Technology, Taiwan, R.O.C., under grants MOST 107-2218-E-992-303-., MOST 107-2221-E-006-073-, and MOST 108-2221-E-006-111-

REFERENCES

[1] Berchmans, D., & Kumar, S. S. (2014, July). Optical character recognition: an overview and an insight. In 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) (pp. 1361-1365). IEEE.

[2] Fu, K. S. (1976). Pattern recognition and image processing. IEEE transactions on computers, 100(12), 1336-1346.

[3] Garcia, C., & Apostolidis, X. (2000). Text detection and segmentation in complex color images. In 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100) (Vol. 4, pp. 2326-2329). IEEE.

[4] Chowdhury, S. P., Dhar, S., Das, A. K., Chanda, B., & McMenemy, K. (2009, July). Robust extraction of text from camera images. In 2009 10th International Conference on Document Analysis and Recognition (pp. 1280-1284). IEEE.

[5] [online] Available: www.raspberrypi.org.

[6] Li, S., & Kwok, J. T. (2004, October). Text extraction using edge detection and morphological dilation. In Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004. (pp. 330-333). IEEE.

[7] Canny, J. (1987). A computational approach to edge detection. In Readings in computer vision (pp. 184-203). Morgan Kaufmann.

[8] Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.".

[9] Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., ... & Shafait, F. (2015, August). ICDAR 2015 competition on robust reading. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1156-1160). IEEE.

[10] Zhong, Z., Sun, L., & Huo, Q. (2017, November). Improved localization accuracy by locnet for faster r-cnn based text detection. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 923-928). IEEE.

[11] Tian, S., Pan, Y., Huang, C., Lu, S., Yu, K., & Lim Tan, C. (2015). Text flow: A unified text detection system in natural scene images. In Proceedings of the IEEE international conference on computer vision (pp. 4651-4659).

[12] Yin, X. C., Yin, X., Huang, K., & Hao, H. W. (2013). Robust text detection in natural scene images. IEEE transactions on pattern analysis and machine intelligence, 36(5), 970-983.