Research Article

MyOcrTool: Visualization System for Generating Associative Images of Chinese Characters in Smart Devices

Laxmisha Rai and Hong Li

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Laxmisha Rai; laxmisha@ieee.org

Received 6 February 2021; Revised 12 March 2021; Accepted 16 April 2021; Published 11 May 2021

Academic Editor: Abd E.I.-Baset Hassanien

Copyright © 2021 Laxmisha Rai and Hong Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Majority of Chinese characters are pictographic characters with strong associative ability and when a character appears for Chinese readers, they usually associate with the objects, or actions related to the character immediately. Having this background, we propose a system to visualize the simplified Chinese characters, so that developing any skills of either reading or writing Chinese characters is not necessary. Considering the extensive use and application of mobile devices, automatic identification of Chinese characters and display of associative images are made possible in smart devices to facilitate quick overview of a Chinese text. This work is of practical significance considering the research and development of real-time Chinese text recognition, display of associative images and for such users who would like to visualize the text with only images. The proposed Chinese character recognition system and visualization tool is named as MyOcrTool and developed for Android platform. The application recognizes the Chinese characters through OCR engine, and uses the internal voice playback interface to realize the audio functions and display the visual images of Chinese characters in real-time.

1. Introduction

In recent years, increasing number of foreigners are studying Chinese as a second language around the world through institutions that promote Chinese language education. As many foreigners are coming to China, either to study, work, and travel—reading, speaking, listening and writing Chinese language are one of the essential requirements for managing the daily activities. Comparing to several other major foreign languages, Chinese languages is considered as one of the hardest language for learners [1]. To read a newspaper or an online article, one need to master at least over two thousand characters. As there are nearly several thousands of Characters in Chinese language, it is essential for a learner to master at least around two to three thousand characters to read or understand a sign board, public safety instructions or a food menu. In [2], a website lists nearly 4000 simplified Chinese characters, based on their frequency of their appearance in written documents. As per the website, a solid knowledge of all these characters makes a learner to read any

document written in simplified Chinese. However, for travelers who are interested in short-term stay in China, they may be more interested in understanding the meaning behind the text, rather than spending countless hours on studying Chinese or hiring a translator. Moreover, these days several software based applications, and mobile Apps are popular, which help the learners to understand the meaning of a Chinese character or a word. In the past years, there are several modern approaches to simplify learning Chinese using mobile devices, software tools, Apps, and electronic devices. A study on mobile assisted language learning (MALL) that emphasizes learner created content and contextualized creation of meanings is presented in [3]. While learning Chinese idioms, the tool let the students to take initiatives to use mobile phones to capture scenes that express the meaning of idioms, and to facilitate to construct sentences with them. This way of transforming idioms into photos can help students understand idioms in a more efficient way. You and Xu [4] evaluated the usability of a system named Xi-Zi-e-Bi-Tong (習-e-筆通), which is one of

the systems for writing Chinese characters used by the education ministry. The main focus of this study is to evaluate the efficacy of the system for foreign learners from different cultural backgrounds and ages. In summary, although Chinese non-native speakers can interact with the system, there are still problems exist, and there is scope for improvement. In [5], researchers designed and evaluated a software that facilitate users to learn Chinese through the use of mobile application. The results from the past literature shows that vast majority of foreign learners are satisfied with learning Chinese through electronic devices, and this also plays a huge role in learning Chinese.

To use the online or mobile applications, and language dictionaries related to Chinese language, a user need to aware of three things. (1) user should be able to know how to read a Chinese character, (2) user need to know how to write a Chinese character, mostly on phone screen by drawing different strokes and following stroke order and, (3) user need to be aware of usage of pinyin (*pinyin*). However, as mentioned earlier, as there are thousands of Chinese characters, and Chinese is a complex language, it is not easy for a non-native learner to be aware of all these.

Majority of Chinese characters represent some actions, events, animals, humans, or objects directly or indirectly. It is evident from Figure 1, that, the Chinese characters are evolved by following logical rules over the years. There are several stages of evolution of Chinese characters [6]. Some of these stages include oracle bone script, bronze script, small seal script, clerical script, standard script and simplified Chinese as shown in Figure 1. Today, simplified Chinese is the most common and widely used script for all official purposes in China. By merely grasping a Chinese character, its connotation and extension, it can produce endless associations. So, in theory, for a Chinese learner in the early stages, commonly used characters are presented as a painting or picture to provide quick association with that character. Earlier, a comprehensive analysis of the images generated by the simplified Chinese characters through the use of Internet and electronic devices, and emphasizing the understanding of text in the form of limited images is studied in [7]. This work is inspired by majority of opinions of scholars, where no matter how simple or complicated a character is, the Chinese character is still a picture. Here, the researchers, tried to understand how each character is represented as images in Internet usage, and in popular messaging tools. With this background, we try to investigate answers to the following research questions. (1) Considering our previous study [7], is it possible to visualize simplified Chinese characters in real-time with their associative images using smart devices? (2) How to develop the application for real-time visualization of Chinese characters extracted from different sources facilitating the evaluation of recognition rates?

The first research question is related to development of an visualization system for generating associative images of Chinese characters. Considering this aspect, there are several related works, which describe the research related to visual perception, virtual reality in applications of industrial domains, and Internet of Things (IoT) in the recent years. In [8]

authors have provided detailed account of applications of visual perceptions in different industrial fields. Three industrial fields include agriculture, manufacturing, and autonomous driving. In [9], importance of human visual system while acquiring different features of an image, and the impact of distortion distribution within an image is studied. In [10], a metric for evaluation of screen contents images for better visual quality is explored. In [11], researchers constructed an image processing and quality evaluation system using convolutional neural network (CNN), and IoT technology to investigate the applications of industrial visual perception in smart cities. The main goal is to provide experimental framework for future smart city visualization. In [12], as an security solution framework, an intrusion detection model of industrial control network is designed and simulated in virtual reality (VR) environment. In [13], the relevance of VR technology applications with consideration to IoT is discussed.

Chinese characters are pictographic characters (or pictograms) with strong associative ability and when a character appears for Chinese readers, they usually associate with the objects, or actions related to the character immediately. Having this background, we propose a system to visualize the simplified Chinese characters so that non-native learners can understand the meaning of a character quickly without even typing or learning it. Considering the extensive use and application of mobile devices, automatic identification of Chinese characters and display of associative images are made possible in smart phones to facilitate quick overview of a Chinese text. This work is of practical significance considering the research and development of real-time Chinese text recognition and display of associative images for such users who has no background in Chinese writing or reading. The proposed Chinese character recognition system and visualization tool is named as MyOcrTool and is suitable for Android platform. The application recognizes the Chinese characters through optical character recognition (OCR) engine called Tesseract, and uses the internal voice playback interface to realize audio functions for character pronunciation and display the visual images of Chinese characters in real-time.

The main purpose of this study is to generate images for each Chinese character and then a representative image for the entire text, so that a user can able to obtain the approximate idea presented in the text. Moreover, a user is able to obtain the meaning, even without understanding *pinyin*, or romanization or any reading ability. So, the system is useful for anyone, who has no literacy on Chinese language, or someone who is unable to listen or speak.

Table 1 shows a list of examples of selected characters and their associated images. This kind of visual images for Chinese characters able to help non-Chinese to visualize the meaning behind Chinese characters rapidly. The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 describes model description by providing details of OCR, Tesseract open source engine, application overview, system design and implementation details. Section 4 presents details of experimental design and analysis of results. Finally, we conclude the paper in Section 5 with some pointers to future work.

FIGURE 1: Some stages of evolution of Chinese characters.

TABLE 1: Sample list of examples of characters and their associated images.

Characters	Meaning	Pinyin	Associative images
心	Heart	xīn
球	Ball	qiú
书	Book	shū
日	Sun	rì
树	Tree	shù
火	Fire	huǒ
车	Car	chē
手	Hand	shǒu

2. Related Work

The major process required to visualize a Chinese character as image is to scan the character using smart phone camera to extract the character within a text. After scanning the character, the subsequent steps such as character recognition, display of associative image, pinyin pronunciation are performed. There are several previous related studies on character recognition in different scenarios, where studies related to characters of languages other than Chinese are also involved [14–21]. From this, one can classify all the existing text extraction methods into three categories as region based, texture based and hybrid method [14]. There are

several works on text detection mentioned here, and they have proposed idea of text detection considering different factors and designed a suitable model. Kastelan et al. [15] presented the system for text extraction on image taken by grabbing the content of TV screen. An open source algorithm for OCR is used to read the text regions. After reading the text regions, comparison with the expected text is performed to make a final success or failure decision for test cases. The system successfully read the text from TV screen and used in a functional verification system. Considering the Chinese character recognition, in the past, several researchers focused their attention on recognition of printed Chinese characters [22, 23], handwritten Chinese characters [24, 25], characters in the vehicle's license plates [26, 27], and recognition of Chinese characters written in calligraphic styles [28].

In the recent years, the OCR technology is used in various applications, where recognitions of characters is the central requirement, such as in applications of e-commerce [29], and IoT [30]. Moreover, the importance of Tesseract engine in character retrieval from images, translation applications, and character recognition applications is widely popular [31, 32]. Ramiah et al. [16] developed an Android application by integrating Tesseract OCR engine, Bing translator and phones' built-in speech technology. By using this App, travelers who visit a foreign country able to understand messages described in different language. Chavre and Ghotkar [17] designed an Android application, which is a user-friendly application to assist the tourist navigation, while they are roaming in foreign countries. This application is able to extract text from an image, which is captured by a mobile phone camera. The extraction is performed using stroke width transform (SWT) approach and connected component analysis. The SWT a technique used to detect texts from natural images by eliminating the noise but preserving the text. Kongtahn et al. [18] presented a method for reading medical documents by using an Android smartphone that used techniques based on the Tesseract OCR Engine to extract the text contents from medical document images such as a physical examination report. The

following factors related to the document are considered: character font, text block size, and distance between the document and the camera on the phone. Dhakal and Rahnemoonfar [19] developed a mobile application for Android platform that allows a user to take a picture of the YSI Sonde monitor (an instrument used to measure water quality parameters such as pH, temperature, salinity, and dissolved oxygen), extract text from the image and store it in a file on the phone.

Nurzam and Luthfi [20] implemented Latin text translation from Bahasa Indonesia into Javanese text with Google Mobile Vision in real-time, also vice versa with Android mobile based application. The execution flow of this design is to first scan the text through the camera, then the recognized text is transmitted to the web services. Finally, the translated text is displayed in real-time on the mobile phone screen. This research uses Javanese language or Indonesian as the outcome of the language conversion. The purpose of this research is to design and implement real-time text-based translator application using Android mobile vision which includes a combination of mobile translator application architecture and web services applications. Yi and Tian [21] proposed a method of scene text recognition from detected text regions. They first designed a discriminative character descriptor by combining several advanced feature detectors and descriptors, and then shaped the structure of the character in each character class by designing the stroke configuration map. Android system is developed to show the effectiveness of their proposed method in extracting textual information from the scene. The results of the evaluation of test data shows that their proposed text recognition scheme has a positive recognition effect, which is comparable to the major existing methods.

Moreover, other than the character recognition, there are several work which are focused on the conversion from text-to-speech (TTS). Celaschi et al. [33] integrated a set of image capturing and processing framework, such as OCR and TTS synthesis. Their work include integration of selected components and several control functions of the application: CPU through the camera to capture images; image preprocessing; OCR framework for text recognition; finally, the speech synthesis process is performed for Portuguese rather than Chinese. This design includes two versions, a preliminary desktop version designed under the Windows operating system, and a mobile device version developed as an application for Android devices. Chomchalerm et al. [34] designed an Android based App called Braille Dict that runs on smart phones. This application was developed for the blind, converting Braille input into English letters and translating them into Thai, and displaying a list of words related to the input words by retrieving them from dictionary database. One of the most significant function of this system is that, the program uses TTS function to output Thai as speech, which provides a more comfortable way for the blind to use the dictionary. In addition, several works in the past focused on OCR in android applications [35, 36], real-time OCR [37], character readability on smart phones [38], character recognition models suitable for handheld devices [39], and App to recognize food items in Chinese

menu [40]. Considering several related work, it is evident that, none of the previous research focused on developing a method to visually understand the text only by scanning. So, in this paper we propose a novel method to facilitate the users to visualize the Chinese text with only by scanning it, rather than typing or entering the text into the electronic devices. Summary of these existing studies is shown in Table 2. As shown in Table 2, most of the research focuses on the OCR technology that only recognizes characters. Only three studies include text-to-speech functions. None of the studies proposes to display the visual images of Chinese characters in real-time. Therefore, this application still has its unique and innovative compared to the studies listed in the table above.

3. Model Description

3.1. OCR Technology. Considering the related works mentioned earlier, most of the earlier implementations are focused on such languages where there are limited characters in a language. However, the character extraction, and recognition is challenging especially in Chinese language considering thousands of complex Chinese characters. In this section, the problems associated with the extraction and recognition of text within an image in various scenarios is considered. Therefore, the OCR method is studied, where rapid extraction of text information from images is possible. The basic operating principle of OCR technology is to convert the information presented in documents into an image file of black and white dot matrix using camera, scanner and other optical equipments. After this process, the characters within the image are converted into editable text through the OCR engine for further information processing [41]. In recent years, OCR technology has been a hot research topic in several disciplines. The concept of OCR was first proposed by Austrian scientist Gustav Tauschek in 1929. Later, American scientist Paul Handel also proposed the idea of using technology to identify words. The earliest research on the recognition of printed Chinese characters was proposed by Casey and Nagy in 1966, where they worked on Chinese character recognition, which used template matching to identify 1000 printed Chinese characters [42]. Research work on OCR technology in China started much later. In 1970s, research on the identification of numbers, English letters and symbols began. In late 1970s, research on Chinese character recognition has started. By 1986, the study of Chinese character recognition entered a substantive stage and many research centers have successively launched Chinese OCR products. Early OCR software failed to meet actual requirements due to various factors such as recognition rate and building them as actual products.

Simultaneously, products have not reached to a level to use in practical applications due to poor execution speed and expensive hardware equipments. After 1986, China's OCR research has made substantial progress and there are several innovations on Chinese character modelling and recognition methods. The developed applications have displayed fruitful results and many centers successively launched Chinese OCR products.

TABLE 2: Comparison of existing character visualization systems on mobile devices.

Authors, Publication year	Mobile OCR Engine	OS	Display feature of visual images	Text-to-Speech feature
Kastelan et al., 2012	Not stated	Not stated	—	—
Kongtala et al., 2014	Tesseract	Android	—	—
Yi and Tian, 2014	Not stated	Android	—	—
Chomchalerm et al., 2014	Not stated	Android	—	Yes
Ramiah et al., 2015	Tesseract	Android	—	Yes
Dhaka and Rahmemonfar, 2015	Not stated	Android	—	—
Chavre and Ghotkar, 2016	Tesseract	Android	—	—
Celaschi et al., 2017	Not stated	Android	—	Yes
Nurzam and Luthfi, 2018	Not stated	Android	—	—

3.2. Tesseract Open Source Engine. The OCR technology used in this work is based on the Tesseract open source engine, which was originally developed by Hewlett-Packard (HP) between 1985 and 1994, and additional changes were made in 1996 to make it compatible with Windows [43]. In 2005, HP made available the Tesseract as open source software. Since 2006, it is developed by Google. The Tesseract engine is powerful and can be broadly divided into two parts: (1) picture layout analysis, and (2) character segmentation and recognition.

The design goal of Tesseract is character segmentation and recognition. Smith et al. [44] described the efforts to adapt the Tesseract open source OCR engine for multiple scripts and languages in 2009. They also presented the top-level block diagram of Tesseract. Real-time display of visual images associated with Chinese character is accomplished using the Android's *RecyclerView* control [45]. When the characters are recognized, it displays the visual images of respective character. In addition, the voice broadcast function use the Android's built-in TTS control [46], which does not require permission to read text and do not require Internet connection. This feature can facilitate the specified text to read aloud providing voice broadcast option to users.

3.3. Overview of the Proposed System. To answer the first research question, designing a mobile intelligent system based on platform such as Android is essential. The main function of this system is to recognize the text contained in a scanned image and display the associated image of Chinese character in real-time, and provide options for other features such as audio for character pronunciation, and pinyin display. Figure 2 shows the screenshots displaying the MyOcrTool in practical scenarios where a user using it to visualize a Chinese text. The Figure 2(a) shows a scenario, where the user trying to visualize the Chinese text in a public sign board. The Figure 2(b) shows another scenario, where the user try to visualize the restaurant menu. To operate the tool developed, a user has to follow the following steps:

- (1) Open MyOcrTool and select the recognized language (Chinese, or English).
- (2) Open the smart phone camera and point the scan frame to text area to be recognized.

- (3) Identify the selected text area. The OCR will automatically identify the scanned text and extract valid string information.
- (4) Real-time display of associated pictures of Chinese characters is performed after the above steps. When a word is recognized, the associated image is displayed in real-time on the mobile phone interface.
- (5) Use the voice playback feature to listen to the text recognized.

3.4. System Design and Implementation. This section introduces the overview of the system architecture and implementation details. The sequence of steps are divided into several processes. They are: (1) scanning using camera to obtain image, (2) image graying, (3) text region binarization, (4) text recognition, (5) displaying of visual images in real-time, and (6) implementation of voice broadcast feature.

3.4.1. Scanning to Obtain Images. Zxing is a Google open source library based on various 1D/2D barcode processing. It is powerful for bar code scanning and decoding via a mobile phone camera and is now commonly used to scan and decode QR codes or barcodes [47, 48]. In this work, Zxing is used to customize the scanning interface of MyOcrTool. The customization process is a three step process, which include: (1) adding the Zxing dependency packages to project, (2) configuring the permission to use the camera in manifest file, and (3) setting the scan interface and scan box.

3.4.2. Image Graying. In order for the open source engine Tesseract to better recognize the image text, some preliminary processing is needed for the image. Gray-scale is the most basic and commonly used to perform this step [49]. In the RGB model, if the values of *R* (red), *G* (green) and *B* (blue) are equal, then color represents a grayscale color. Moreover, the value is called grayscale value. Therefore, each pixel of grayscale image only needs one byte to store the grayscale value (also known as intensity value and brightness value), and the grayscale range is 0–255. There are four methods to gray color images: component method, maximum method, average method and weighted average method [50]. In this paper, the weighted average method is

FIGURE 2: The screenshots displaying the MyOcrTool in practical scenarios where a user using it to visualize a Chinese text (a) in a public sign board, and (b) in a restaurant menu.

used to gray the image to obtain the image Y , and the formula is shown in equation (1). The sequence of steps and implementation details involved in picture graying is shown in Program Listing 1.

$$Y = 0.3R + 0.59G + 0.11B. \quad (1)$$

3.4.3. Text Region Binarization. In order to facilitate recognizing the text within images, binary processing of grayscale images is required [51]. Binary processing is mainly applied for the convenience of image information extraction and this can increase the recognition efficiency. Binary image refers to an image whose pixel is either black or white, and whose gray value has no intermediate transition. The most commonly used method for binarization of images is to set a threshold value T , which is used to divide the image data into two parts. The pixel groups greater than T , and groups smaller than T which are represented by 1 and 0 respectively. Considering the input grayscale image function be $f(x, y)$ and the output binary image $g(x, y)$ can be expressed by .

$$g(x, y) = \begin{cases} 0, & f(x, y) < T, \\ 255, & f(x, y) \geq T. \end{cases} \quad (2)$$

Threshold is a measurement to distinguish the target from the background. Selecting an appropriate threshold is not only necessary to save image information as much as possible, but also to minimize the interference of background and noise, which is the principle of behind threshold selection. To accomplish this, the program use the iterative method to find the threshold [52], and this iterative method is a global binarization method. It requires the image segmentation threshold algorithm based on the approximation strategy. Firstly, an approximate threshold is selected as the initial value of the estimated value, then segmentation is performed to generate a sub-image. Following this, a new threshold is selected according to the characteristics of the sub-image and, a new threshold is utilized. Secondly, the image is divided, after several iterations, minimizing the

number of incorrectly segmented image pixels. This procedure performs better than the effect of directly segmenting the image with the initial threshold. The specific algorithmic steps are as follows:

- (1) Find the minimum and the maximum gray value in the image which are denoted as $Z_{\min}$ and $Z_{\max}$ respectively, then obtain the initial value of the threshold.

$$T_0 = Z_{\min} + Z_{\max}. \quad (3)$$

- (2) According to threshold value T_k , the image is divided into two parts, target and background, and average gray values Z_0 and Z_1 of the two parts are obtained.
- (3) Find the new threshold T_1

$$T_1 = \frac{(Z_0 + Z_1)}{2}. \quad (4)$$

- (4) If $T_0 = T_1$ then the current T is the optimal threshold, otherwise the value of T_1 is assigned to T_0 , and the calculation restarts from step (2).

The implementation details of the iterative method for calculating the threshold is shown in Program Listing 2.

3.4.4. Chinese Text Recognition. After the image has been pre-processed, the processed image will be used for character recognition and open source engine Tesseract is used as a tool for recognizing characters. Android Studio is used for writing programs, and programming requires Tesseract's third-party JAR package as additional support. In addition, the language package "`<language>.traineddata`" required to be placed in the mobile phone's secure digital (SD) card root directory [53]. The language packs can be downloaded directly from the Tesseract website, or its own trained language packs. This design also uses trained language packs and uses its own language library which are suitable for identifying at correct rate and speed. The flow diagram representing principle involved in character recognition is shown in Figure 3.

```

Input: original image
Output: grayImage private static Bitmap getGrayImg(){
    int alpha=0xFF << 24;
    // Set transparency
    for (int i=0; i < imgHeight; i++) {
        for (int j=0; j < imgWidth; j++) {
            int grey = imgPixels[imgWidth*i + j];
            // Get the jth pixel of the i-th row
            int red = ((grey & 0x00FF0000) >> 16);
            // Get red gray value
            int green = ((grey & 0x0000FF00) >> 8);
            // Get green gray value
            int blue = (grey & 0x000000FF);
            // Get blue gray value
            grey = (int) ((float) red*0.3 + (float) green*0.59 + (float) blue*0.11);
            // obtain grayscale color values
            grey = alpha | (grey << 16) | (grey << 8) | grey; imgPixels[imgWidth*i + j] = grey;
        }
    }
    Bitmap result = Bitmap.createBitmap(imgWidth, imgHeight, Config.RGB_565);
    result.setPixels(imgPixels, 0, imgWidth, 0, 0, imgWidth, imgHeight);
    return result;
}

```

PROGRAM LISTING 1: Implementation of picture graying procedure.

```

Input: grayImage
Output: threshold
private static int getIterationHresholdValue (int minGrayValue, int maxGrayValue) {
    int T1;
    int T2 = (maxGrayValue + minGrayValue)/2;
    do {
        T1 = T2;
        double s = 0, l = 0, cs = 0, cl = 0;
        for (int i = 0; i < imgHeight; i++) {
            for (int j = 0; j < imgWidth; j++) {
                int gray = imgPixels[imgWidth * i + j];
                if (gray < T1) {
                    s += gray;
                    cs++;
                }
                if (gray > T1) {
                    l += gray;
                    cl++;
                }
            }
        }
        T2 = (int) (s / cs + l / cl) / 2;
    }
    while (T1 != T2);
    return T1;
}

```

PROGRAM LISTING 2: The program listing of iterative method for calculating the threshold.

3.4.5. Real-Time Display of Associative Images. The function of displaying visual images in real-time is performed using Android's own control *RecyclerView*. *RecyclerView* is a container for displaying huge data sets that displays large

volume of data in a limited window and simplifies the presentation and processing of data [45]. While using *RecyclerView*, we must specify an Adapter and a *LayoutManager*. The main function of the Adapter is to bind the

FIGURE 3: Flow diagram representing the principle involved in character recognition.

data to the control. The *LayoutManager* can control the layout of the *Item*. The functions of real-time display of visual images introduced in this paper are mainly to bind Chinese characters, visual pictures and edit boxes that display recognized Chinese characters. When Chinese characters are identified and displayed in the edit box, simultaneously visual images of respective characters are displayed on the mobile phone screen.

3.4.6. Implementation of Pronunciation Playback Feature.

The voice playback function presented in this paper use the TTS engine that comes with Android, and it is new and important function in Android 1.6. It can be easily embedded into the application to convert the specified text into different language audio output to enhance the user experience. The role of this implementation is to play the recognized words of the Chinese text by clicking the voice button, so that the user not only able to understand the meaning of Chinese characters, but also could hear the pronunciation. The implementation details for the voice playback function is shown in Program Listing 3.

4. Results and Discussion

4.1. Overview of the Experiment. The entire application is tested on two brands of Android based mobile phones. After initially selecting the tool, we have to select the recognition language type, open the camera for scanning, and then align the scan box with the text area to be scanned. The scan frame size set by the system is of minimum width of 200 dp, a maximum width of 250 dp and a height of 80 dp, which is physically 1.5 cm wide and 0.5 cm high. The main purpose of using dp (device-independent pixel) unit is to adapt the UI layout of application to display devices of various resolutions. Finally, the recognition results and the image will appear on the mobile phone display interface as shown in Table 3, which sufficiently provide the answer to the first research question posed. The testing is carried out by considering the different font size, distance between the phone camera and text, and the text from different sources such as books, warning signs and restaurant menus. This kind of abilities to evaluate recognition rate of characters extracted from different sources answers the second research question posed in this work.

As shown in Figure 4, we have shown three general cases for displaying the images for characters and words. In

case (a), where a signboard related to water conservation is translated as an image, and in case (b) for the Chinese word “中国” in a book, shows the map of China, because the word “中国” means the name of country China. In the case (c), for the word “花生米”, the picture of peanut dish is displayed, because meaning of “花生米” in Chinese is peanut. We have tested for these three scenarios, with general presumption that non-native learners interact more frequently with signboards, restaurant menus, and tourist guidebooks.

4.2. Testing for Recognition Stability. While testing the system, we have considered the several factors as the main test criteria to evaluate the stability and recognition rate of the system. The recognition rate is defined as the ratio between the number of successfully recognized characters and the total number of characters in the test image. In Table 3, we have presented the results obtained for different kinds characters which represent animals, objects, and actions by taking sixty characters as test samples. All these characters generated independent single and unambiguous image to represent the characters and these results are considered 100% acceptable. Two reasons can be identified for this success. Traditionally these characters represent the same meaning of objects, animals and actions. Moreover, even though they are used in different contexts, and communication scenarios today, the original meaning of characters can be still possible to interpret with traditional meaning.

However, as mentioned in [6, 7], it is nearly impossible to find an exact image for each Chinese character, especially within a word because of contextual differences and usage. For example, considering an example of object “tree”, some users may expect tree of bigger size, other may think tree with only few leaves, and of smaller size. So, the fundamental approach here is to provide the image which is widely accepted by users. We have also followed the similar steps as presented in [7] to collect images for testing. In some cases, several characters in a word has the same meaning, so a single image is enough to represent a word or several characters. Table 4 shows an example with several Chinese characters, their corresponding *pinyins*, and general meaning of these characters. As shown, 12 characters (如,何,吗,因,由,认,谁,思,怎,想,若,难) may share a single image because they all have similar meanings (if, why, question?, reason, because of, how?, to recognize, difficult, who?, to think etc.).


```

Input: text
Output: speech
private static ImageButton yuyinButton;
private TextToSpeech textToSpeech;
@Override
protected void onCreate (Bundle savedInstanceState) {
    super.onCreate (savedInstanceState);
    setContentView (R.layout.my_scan);
    yuyinButton = (ImageButton) findViewById (R.id.yuyinButton);
    textToSpeech = new TextToSpeech (this, new TextToSpeech.OnInitListener() {
    @Override
    public void onInit(int status) {
        if (status == textToSpeech.SUCCESS) {
            int result = textToSpeech.setLanguage(Locale.CHINA);
            if (result != TextToSpeech.LANG_COUNTRY_AVAILABLE&& result !=
                TextToSpeech.LANG_AVAILABLE){
                Toast.LENGTH_SHORT).show();
            }
        }
    }
    });
    yuyinButton.setOnClickListener (new View.OnClickListener() {
    @Override
    public void onClick (View view) {
        textToSpeech.speak (status_view_tv_result.getText().toString(),
            TextToSpeech.QUEUE_ADD, null);
        textToSpeech.setSpeechRate (0.5f);
        textToSpeech.setPitch (0.1f);
    }
    });
}

```

PROGRAM LISTING 3: The program listing of voice playback function.

TABLE 3: Chinese characters and generated images with 100% acceptance rate.

Objects	球	日	月	火	树	星	锁	船	车	药

	伞	杯	包	云	花	耳	鞋	脚	眼	鼻

Animals	狗	虎	兔	蛇	羊	马	牛	猴	猪	鸡

	鼠	龙	鸟	狼	象	鹰	熊	龟	猫	鱼

Actions	走	跑	看	打	唱	敲	想	听	踩	按

	喝	咳	笑	哭	读	切	吃	抱	坐	睡

FIGURE 4: MyOcrTool's is illustrated with application's screenshots with different cases such as scanning (a) signboard, (b) book, and (c) restaurant menu.

TABLE 4: A representative image which is shared by many characters having similar meanings.

Character	Pinyin	Meaning	Associative image
何	hé	Why	
吗	má	Question?	
因	yīn	Reason	
由	yóu	Because of	
认	rèn	To recognize	
谁	shéi	Who?	
思	sī	To think	
怎	zěn	How?	
想	xiǎng	To think	
若	ruò	If	
难	nán	Difficult	
如	rú	If	

However, Table 5 lists such characters, where it is not possible for some characters to generate an accurate image to represent them. Because, there are characters without direct or independent image and are difficult to be visualized. Some characters with medium acceptance rate and others with low

acceptance rate are shown. This is because, some of the low acceptance rate characters belong to the grammatical structure or modal particles of Chinese writing. Interestingly, these characters are found less in sign boards and restaurants menus. So, failure to generate them accurately

TABLE 5: Chinese characters and generated images with medium and low acceptance rate.

	可	嘻	飞	气	比	希	或	吗	孤	喊
Medium acceptance rate										
	了	的	幅	科	苦	斯	展	闻	裹	片
Low acceptance rate										

TABLE 6: Evaluation of most suitable distance between the camera and characters.

Scan frame size (cm) (width × height)	Font size	Distance (cm) (vertical direction)
1.5 × 0.5	48	7.1
	26	12.9
	14	19.5

cannot be considered as a major limitation of the system developed. Another solution to this kind of problems is to use the translated captions next to pictures of recognized characters, so that users could perceive the intending meaning of pictures as suggested in [54].

4.3. Testing Based on Different Fonts, Font Sizes, and Varying Distance. While testing the Chinese text which is written in song typeface, bold-face, regular script and imitation song, and their recognition rates were measured. We found that font type has no significant influence on recognition rate. However, if scan box size is fixed, the font size and distance are the two factors that affect each other. In order to measure this, we divided the character size into three levels: large, medium, and small. Characters with large fonts had font size of 48, medium characters had a size of 26, and small characters had a size of 14. After setting the scan box size and font size, we can determine the most appropriate distance required between the camera and the character. The measurement results are shown in Table 6.

4.4. Text from Different Sources. We have tested the accuracy and stability of MyOcrTool in different scenarios, by scanning the text from different sources such as books, warning signs, and restaurant menus. From the test results, we found that MyOcrTool has nearly the same accuracy and stability in these different scenarios. The text recognition rates of software in three different scenarios is shown in Figure 5, where both single word, and long sentences are considered.

Considering the results obtained in the above two cases, the system is showing acceptable performance and can provide better support for Chinese learners to understand the meaning of Chinese characters and text from the perspective of visual information association. The system has a high accuracy rate of about 88%, which can meet the daily learning needs of Chinese learners, but further strengthening of the recognition ability is also necessary.

While testing in two brands of mobile phones (Oppo-R7SM, and Vivo-Y66) it is found that, the average time

FIGURE 5: The histogram of recognition rate of MyOcrTool in different text sources.

required to identify a text is 7.8 seconds. The software execution speed is depends on many factors. Firstly, it can be determined according to the specific words. The frequently used words are recognized faster, and if the words are not used frequently the speed will be slower. Secondly, the execution speed is depends on the font type, and number of strokes. It is founds that, if the font design is complicated, the stroke recognition will be slower. For example, comparing two different characters such as “翻” and “天”, the latter character is faster than the former. Thirdly, execution speed is also depends on the resolution of camera lens, and higher the resolution, the faster the recognition speed. As the application is developed for handheld devices, there is no need of any Internet access to generate Images, and in the current system all the images are packaged into the program itself. However, having Internet provides future possibilities to integrate with several tools, which we have not considered yet. Moreover, we have also not considered memory space required in the mobile devices, because, as images are of small in size, the entire visualization system takes less system space.

5. Conclusion and Future Work

In this paper, an Android based system named MyOcrTool to capture Chinese characters from different text sources and displaying the visual images associated with them in

real-time along with audio options to listen to the script is developed. MyOcrTool displays the visual image related to a Chinese character in real-time after recognizing the text. With this learners from almost all backgrounds are able to visualize the Chinese text only through scanning in Android based devices. This proves, that, we can conclusively answer the research first research question mentioned in the Introduction section. Moreover, learners do not need to develop the skills such as remembering *pinyin*, or stroke sequences. They can also use this system without even reading or writing Chinese characters, and entering any information to the device to obtain the meaning is absolutely unnecessary. The proposed system is designed for such learners who would like to visualize the daily life Chinese texts for rapid assimilation of meanings behind them. After the experimental evaluation, it is found that, the text recognition rate of MyOcrTool reaches nearly 90%, and the time delay between text recognition, and display of visual image in real-time is less than half seconds. The recognition results obtained conclusively prove that, it is possible to evaluate recognition rate of characters which were extracted from different sources. This answers the second research question.

However, we can also list some of the limitations of this work, and there is scope for further research. Firstly, considering the sources such as text from newspaper articles, as they are written with particular context, and generating image for a sentence is beyond the scope of this work. As the sentences becomes longer, we find that, the characters which can be shown with 100% accurate image are diminishing, because there would be more Chinese pronouns. So, we found testing such features of long sentences, and providing exact recognition considered as part of future work. Secondly, there is scope for improving the text recognition speed through applying better recognition algorithms, and image processing methods. It is also possible to show multiple sequence of images for single character based on context with using pictures using GIF (Graphics Interchange Format) animation to avoid ambiguities in their visualization meanings. In addition, displaying corresponding pictures with translation functions could solve the problems with ambiguous words or pictures. Similarly, in-depth study on handwritten Chinese text recognition, and associative image generation is also necessary. Thirdly, the developed MyOcrTool unable to process the noisy background around a scanned text, which makes the system difficult to identify characters in text sources with messy background. This limitation also reduce the recognition rate and processing speed significantly. Finally, regarding the voice playback feature, more sophisticated and advanced playback engine can be used to make the text-to-speech sound more user friendly, and error-free.

Data Availability

No raw data is required to reproduce the work other than few representative images as shown in Tables 1, 3–5. Three Program Listings are included within the manuscript itself,

so that programming support is enclosed to reproduce the application. The representative images are collected by following the earlier work proposed in Reference [7], and Chinese characters are collected from link provided in Reference [2]. The software used to develop the proposed system are obtained from the links provided in References [45, 48].

Conflicts of Interest

The authors declares that there is no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China.

References

- [1] X. Yan, Y. Fan, Z. Di, S. Havlin, and J. Wu, "Efficient learning strategy of Chinese characters based on network approach," *PLoS One*, vol. 8, no. 8, Article ID e69745, 2013.
- [2] "Learn to read and write simplified Chinese characters," 2018, <http://www.learnchineseez.com/read-write/simplified/index.php>.
- [3] L. H. Wong, C. K. Chin, C. L. Tan, and M. Liu, "Students' personal and social meaning making in a Chinese idiom mobile learning environment," *Educational Technology & Society*, vol. 13, no. 4, pp. 15–26, 2010.
- [4] M. You and Y. J. Xu, "A usability testing of Chinese character writing system for foreign learner," *Lecture Notes in Computer Science*, vol. 8013, no. 2, pp. 149–157, 2013.
- [5] F. Rosell-Aguilar and K. Qian, "Design and user evaluation of a mobile application to teach Chinese characters," *The JALT CALL Journal*, vol. 11, no. 1, pp. 19–40, 2015.
- [6] Omniglot, "Evolution of Chinese characters," 2018, <https://www.omniglot.com/chinese/evolution.htm>.
- [7] L. Rai, T. Yang, Z. Yue, N. Sun, and R. Shadiev, "Visualizing characters as images: understanding Chinese through Internet usage," in *Proceedings Of the 17th IEEE International Conference On Advanced Learning Technologies (ICALT)*, Timisoara, Romania, July 2017.
- [8] J. Yang, C. Wang, B. Jiang, H. Song, and Q. Meng, "Visual perception enabled industry intelligence: state of the art, challenges and prospects," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2204–2219, 2021.
- [9] K. Sim, J. Yang, W. Lu, and X. Gao, "MaD-DLS: Mean and Deviation of Deep and Local Similarity for image quality assessment," *IEEE Transactions On Multimedia*, pp. 1–12, 2021.
- [10] J. Yang, Y. Zhao, J. Liu et al., "No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions," *IEEE Transactions on Cybernetics*, pp. 1–13, 2020.
- [11] Z. Lv and D. Chen, "Industrial visual perception technology in Smart City," *Image and Vision Computing*, vol. 105, 2021.
- [12] Z. Lv, D. Chen, R. Lou, and H. Song, "Industrial security solution for virtual reality," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6273–6281, 2021.

- [13] Z. Lv, "Virtual reality in the context of Internet of things," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9593–9602, 2020.
- [14] K. Jung, K. In Kim, and A. K. Jain, "Text information extraction in images and video: a survey," *Pattern Recognition*, vol. 37, no. 5, pp. 977–997, 2004.
- [15] I. Kastelan, S. Kukolj, V. Pekovic, V. Marinkovic, and Z. Marceta, "Extraction of text on TV screen using optical character recognition," in *Proceedings Of the 10th IEEE Jubilee International Symposium On Intelligent Systems And Informatics*, pp. 153–156, Subotica, Serbia, September 2012.
- [16] S. Ramiah, T. Y. Liong, and M. Jayabalan, "Detecting text based image with optical character recognition for English translation and speech using Android," in *Proceedings Of the Student IEEE Conference On Research And Development (SCORED)*, pp. 272–277, Kuala Lumpur, Malaysia, December 2015.
- [17] P. Chavre and A. Ghotkar, "Scene text extraction using stroke width transform for tourist translator on Android platform," in *Proceedings Of the International Conference On Automatic Control And Dynamic Optimization Techniques (ICADOT)*, pp. 301–306, Pune, India, September 2016.
- [18] A. Kongtahn, S. Minsakorn, L. Yodchaloemkul, S. Boontarak, and S. Phongsuphap, "Medical document reader on android smartphone," in *Proceedings of the 3rd ICT international senior project conference (ICT-ISPC)*, pp. 65–68, Nakhon Pathom, Thailand, October 2014.
- [19] S. Dhakal and M. Rahnemoonfar, "Mobile-based text recognition from water quality devices," in *Proceedings of the Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications*, vol. 9411, San Francisco, CA, USA, March 2015.
- [20] F. D. Nurzam and E. T. Luthfi, "Implementation of real-time scanner Java language text with mobile vision Android based," in *Proceedings Of the International Conference On Information And Communications Technology (ICOIACT)*, pp. 724–729, Yogyakarta, Indonesia, April 2018.
- [21] C. Yi and Y. Tian, "Scene text recognition in mobile applications by character descriptor and structure configuration," *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 2972–2982, 2014.
- [22] J. Tan, X. H. Xie, W. H. Zheng, and J. H. Lai, "Radical extraction using affine sparse matrix factorization for printed Chinese characters recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no. 3, pp. 211–226, 2012.
- [23] X. Wang, H. Du, and X. Wen, "Research on segmentation and recognition of printed Chinese characters," *Journal of Physics: Conference Series*, vol. 1237, 022011 pages, 2019.
- [24] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing Chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.
- [25] Z. Cao, J. Lu, S. Cui, and C. Zhang, "Zero-shot handwritten Chinese character recognition with hierarchical decomposition embedding," *Pattern Recognition*, vol. 107, 2020.
- [26] H. Chen, B. Hu, X. Yang, M. Yu, and J. Chen, "Chinese character recognition for LPR application," *Optik*, vol. 125, no. 18, pp. 5295–5302, 2014.
- [27] M. Zhang, F. Xie, J. Zhao, R. Sun, L. Zhang, and Y. Zhang, "Chinese license plates recognition method based on A robust and efficient feature extraction and BPNN algorithm," *Journal of Physics: Conference Series*, vol. 1004, no. 1, 012022 pages, 2018.
- [28] P. Gao, J. Wu, Y. Lin, Y. Xia, and T. Mao, "Fast Chinese calligraphic character recognition with large-scale data," *Multimedia Tools and Applications*, vol. 74, no. 17, pp. 7221–7238, 2015.
- [29] Y. Lu, J. Li, H. Zhang, and S. Lin, "Chinese character recognition of e-commerce platform pictures," in *Proceedings Of the 13th International Computer Conference On Wavelet Active Media Technology And Information Processing*, pp. 28–31, Chongqing, China, March 2017.
- [30] Y. Yin, W. Zhang, S. Hong, J. Yang, J. Xiong, and G. Gui, "Deep learning-aided OCR techniques for Chinese uppercase characters in the application of Internet of things," *IEEE Access*, vol. 7, pp. 47043–47049, 2019.
- [31] P. Sun, Q. Xie, Z. Wu, X. Feng, J. Cai, and Y. Jiang, "Yi characters recognition based on tesseract-OCR," in *Proceedings Of IEEE 3rd Advanced Information Management, Communicates, Electronic And Automation Control Conference*, pp. 102–106, Chongqing, China, October 2019.
- [32] B. Wang, Y. W. Ma, and H. T. Hu, "Hybrid model for Chinese character recognition based on Tesseract-OCR," *International Journal of Internet Protocol Technology*, vol. 13, no. 2, pp. 102–108, 2020.
- [33] S. Celaschi, M. S. Castro, and S. P. Da Cunha, "Read it aloud to me," *Universal Access in Human-Computer Interaction. Designing Novel Interactions*, vol. 10278, pp. 260–268, 2017.
- [34] G. Chomchalerm, J. Rattanakajornsak, U. Samsrisook, D. Wongsawang, and W. Kusakunniran, "Braille Dict: dictionary application for the blind on android smartphone," in *Proceedings of the 3rd ICT international senior project conference (ICT-ISPC)*, pp. 143–146, Nakhon Pathom, Thailand, October 2014.
- [35] G. A. Robby, A. Tandra, I. Susanto, J. Harefa, and A. Chowanda, "Implementation of optical character recognition using Tesseract with the Javanese script target in android application," *Procedia Computer Science*, vol. 157, pp. 499–505, 2019.
- [36] G. B. Holanda, J. W. M. Souza, D. A. Lima et al., "Development of OCR system on android platforms to aid reading with a refreshable braille display in real time," *Measurement*, vol. 120, pp. 150–168, 2018.
- [37] Y. G. Lee, "Novel video stabilization for real-time optical character recognition applications," *Journal of Visual Communication and Image Representation*, vol. 44, pp. 148–155, 2017.
- [38] S.-M. Huang, "Effects of font size and font style of Traditional Chinese characters on readability on smartphones," *International Journal of Industrial Ergonomics*, vol. 69, pp. 66–72, 2019.
- [39] A. D. Cheok, Z. Jian, and E. S. Chng, "Efficient mobile phone Chinese optical character recognition systems by use of heuristic fuzzy rules and bigram Markov language models," *Applied Soft Computing*, vol. 8, no. 2, pp. 1005–1017, 2008.
- [40] M. C. Lee, S. Y. Chiu, and J. W. Chang, "A deep convolutional neural network based Chinese menu recognition App," *Information Processing Letters*, vol. 128, pp. 14–20, 2017.
- [41] M. Koga, R. Mine, T. Kameyama, and T. Takahashi, "Camera-based Kanji OCR for mobile-phones: practical issues," in *Proceedings of the International Conference On Document Analysis And Recognition (ICDAR)*, pp. 635–639, Seoul, South Korea, January 2006.
- [42] R. Casey and G. Nagy, "Recognition of printed Chinese characters," *IEEE Transactions on Electronic Computers*, vol. 15, no. 1, pp. 91–101, 1996.

- [43] R. Smith, "An overview of the Tesseract OCR engine," in *Proceedings Of the International Conference On Document Analysis And Recognition (ICDAR)*, Parana, Brazil, November 2007.
- [44] R. Smith, D. Antonova, and D. S. Lee, "Adapting the Tesseract open source OCR engine for multilingual OCR," in *Proceedings of International Workshop on Multilingual OCR*, Barcelona, Spain, July 2009.
- [45] "Using the recyclerview," 2021, <https://guides.codepath.com/android/using-the-recyclerview>.
- [46] H. Jiang, T. Gonnot, W. J. Yi, and J. Saniie, "Computer vision and text recognition for assisting visually impaired people using Android smartphone," in *Proceedings Of the IEEE International Conference On Electro Information Technology*, pp. 350–353, Lincoln, NE, USA, October 2017.
- [47] Y. Yi, "The design of 2D bar code recognition software on Android," *Advanced Materials Research*, vol. 442, pp. 453–457, 2012.
- [48] Zxing, "Google open source," 2018, <https://opensource.google.com/projects/zxing>.
- [49] G. Chen, "Application of processing techniques from color image to grey image," in *Proceedings Of the 2nd International Conference On Software Technology And Engineering (ICSTE)*, San Juan, PR, USA, October 2010.
- [50] C. Kanan and C. W. Cottrell, "Color-to-grayscale: does the method matter in image recognition?," *PLOS One*, vol. 7, no. 1, Article ID e29740, 2012.
- [51] S. S. Lokhande and N. A. Dawande, "A survey on document image binarization techniques," in *Proceedings Of the 1st International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, pp. 742–746, Pune, India, July 2015.
- [52] S. H. Shaikh, A. Maiti, and N. Chaki, "Image binarization using iterative partitioning: a global thresholding approach," in *Proceedings Of the International Conference On Recent Trends In Information Systems(ReTIS)*, pp. 281–286, Kolkata, India, December 2011.
- [53] C. Clausner, A. Antonacopoulos, and S. Pletschacher, "Efficient and effective OCR engine training," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 23, no. 1, pp. 73–88, 2020.
- [54] R. Shadiev, T.-T. Wu, and Y.-M. Huang, "Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts," *ReCALL*, vol. 32, no. 2, pp. 195–212, 2020.