Matlab - Based Reading System for the Blind

Yue Wang[1], Wei Li[2]

[1]Institute of Electronic Science and Technology, Xiamen University, Xiamen, Fujian Province, 361102, People's Republic of China
[2]Sino-British College, University of Shanghai for Science and Technology, Shanghai, 200031, People's Republic of China

Abstract: As a minority group, blind people have always faced many inconveniences in their life. In particular, despite the popularity of e-books, the number of physical books is still huge, which poses a serious obstacle to the reading of the blind. Therefore, assisted reading has become the focus of the current action to care for the blind. There are audio books, text-to-speech and other related technologies on the market. However, these mature technologies are not efficient in dealing with the fragmentation of physical text. This paper gives a implementation which based on MATLAB, it will converts text of a picture to speech playback. Through template matching method and text-to-speech technology, the function of picture text to speech is realized, and the gap from picture to speech is solved. By using this system, pictures and words can be converted into speech playback in real time, which provides a method for helping the blind to read. In addition, the related algorithms used in this system can be extended to intelligent vision monitoring, image retrieval, image matching and other aspects.

Keywords: Template matching; Text segmentation; TTS

1. Introduction

Blind people lose their sight, they face great reading difficulties compared to the general reading population. Living in the dark for a long time, blind people have some difficulty in communicating with ordinary people, and reading is even more difficult. Without the help and care of ordinary people, reading barriers and difficulties will increase, which is the main cause of reading disability for the blind [1].

In recent years, the rapid development of image processing technology has brought up image to text technology. OCR technology has solved the word gap in the Internet era: entity -- electronics. At the same time, the technology related to voice is also constantly mature, realizing the multi-directional development of voice in the Internet level, convenient for our life.

However, the technology of image text to speech on the market needs to be supplemented. Converting visual images into audio can not only facilitate our lives, but also help the blind to solve the problem of dyslexia caused by fragmented text. The combination of the two technologies can solves the connection between text and speech.

Existing OCR techniques mainly include the feature extraction and convolution networks, but this kind of algorithm based on AI needs a large number of data sets and the training set. At the same time, a large amount of data is easy to cause system operating burden. However, this algorithm does not need a large amount of data of libraries, it just need to put the character templates into the template library.

This paper will propose a solution based on MATLAB, which can convert pictures and words into audio: First of all, the image is gray procedure and binarized to transform the image into white characters on black background (0,1 matrix), and then the pre-processed image is segmented. The text is matched with the text in the text library by template matching method, and the audio in the corresponding audio library is played through logical operation. By using the method proposed in this paper, the text information in the picture can be effectively converted into speech information in real time, and the real time connection from image to audio is realized, and the gap between image and audio conversion is overcome. At the same time, the text segmentation algorithm can retain the original attributes of the text image and make a direct comparison with the image in the text database. This system can convert the picture text to the speech playback, provides a new method for the blind reading, and solves the inconvenience of the blind reading text.

2. System design

The system is mainly composed of image preprocessing module, text segmentation module, text library generation module, audio library generation module and text and audio matching module. First of all, the system uses gray transformation and image binarization to preprocess the image. Then, the generated binarization image is divided and the text in the picture is divided into single words. At the same time, the text library and audio library are established. The segmented text is matched with the text in the text library. When the matching is completed, the corresponding audio is played to realize the function of transferring pictures to audio. Each module is described in detail below.
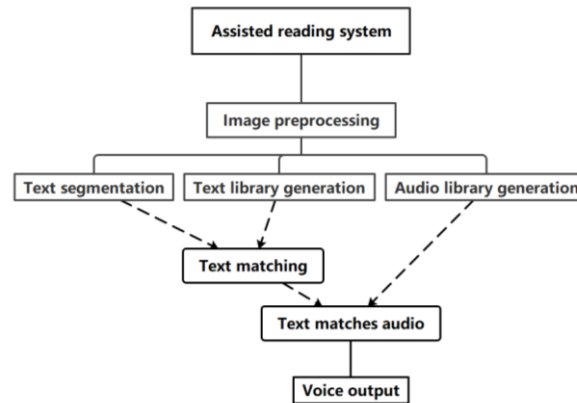

Figure 1: System Architecture Diagram

2.1. Image preprocessing

Due to the different quality of the original image, it is easy to be affected by external factors such as light, brightness and Angle. In order to achieve the effect of the subsequent image text recognition and segmentation, it is necessary to preprocess the original image. Image preprocessing technology mainly includes image binarization, denoising, tilt correction, image open and closed operation. Firstly, most of the interference signals in the image are eliminated by image denoising, and then the image text is corrected by tilt correction algorithm. After experiments and debugging, this paper first carries out grayscale processing on the image. Grayscale processing can reduce the matrix dimension while preserving the image pixel gradient information, so as to improve the operation speed. Binarization processing of the image is to set the gray value of the points on the image to 0 or 255, so that the image presents an obvious black and white effect, that is, to obtain a black and white image [5]. Binarization is shown in Figure 2(a).

Image preprocessing finally changes the pixel matrix into 0 and 1 elements, which lays a foundation for the subsequent realization of text segmentation, improves the running efficiency of the program, and thus simplifies the algorithm. After the original image is preprocessed, the text part pixel value is set to 1, and the background pixel value is set to 0.
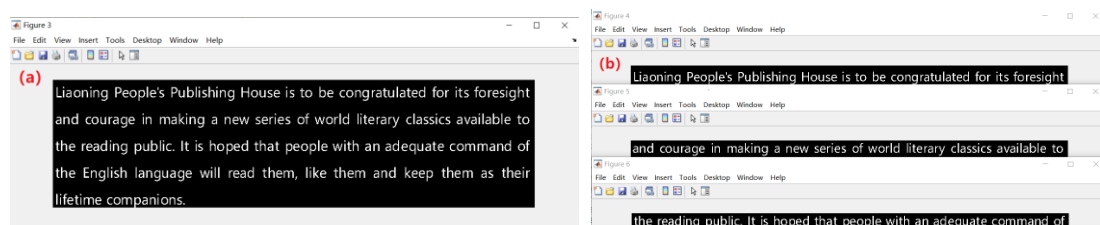

Figure 2: (a) Image binarization processing and (b) Branch results

2.2. Text Segmentation

At present, the text segmentation techniques mainly include threshold segmentation, unsupervised clustering segmentation and image text segmentation based on statistical model. Threshold segmentation method is mainly used in text document analysis, namely the background of the text region are usually not very complex, at the same time the text area background pixel brightness is lower than text brightness,

so you can set the brightness of the reasonable threshold, the background pixels in the image and text area pixel segmentation, the resulting text collection of pixels [4]. Unsupervised clustering segmentation mainly uses k-means algorithm, which is usually used to process color images. The image text segmentation method based on statistical model establishes a probability analysis model for the pixels in the digital image, conducts special processing for the pixels in the text region, and realizes the segmentation of the text region according to the relevant statistical results [4].

In this paper, the threshold segmentation method is used for text segmentation. Firstly, the paragraph branch is processed, and then the row is divided into a single word. By adjusting the threshold value, it will achieve the best effect. The result of branch processing is shown in Figure 2(b), and the result of word segmentation is shown in Figure 3. The method of branch and hyphenation is the same.

First of all, sum the row pixels together and get a row number * 1 matrix. The matrix storage the sum of each row of pixels. By finding the minimum value, we will use it as a threshold for branch operation. However, the output result is not ideal, so we need to experiment on the threshold to find a reasonable threshold as a branch of the threshold. By recording the starting line of pixel and the line reaching the threshold value respectively as the output basis of each line of text, the result of each branch is output by using the circular structure. The principle of hyphenation is the same, just transpose rows and columns.

Due to the different sizes of the segmented text images, it is necessary to operate the segmented images with a uniform size, in order to facilitate the logical operation of pixel points during the subsequent text matching. The imresize function is used to unify the size of Chinese characters to 40×40, and the agreed height of English characters to 40.
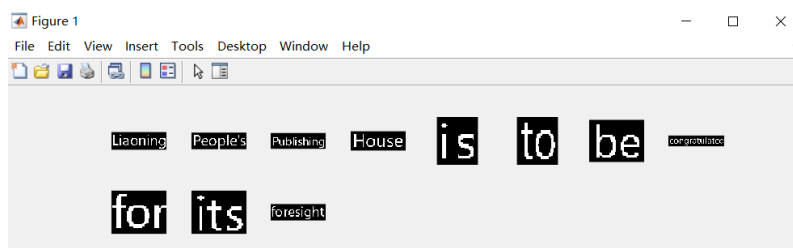

Figure 3: Split result (take the first act as an example)

2.3. Text library generation

Chinese and English word databases are established apart by using the above word segmentation methods, and each segmented word image is stored in the folder of the word database according to the generation sequence for subsequent word matching. The unified format of the name of the text is generated, and the circular naming is used to facilitate the design and matching of subsequent algorithms. It is worth noting that the text image here should be the same size as the image after the text segmentation, otherwise the subsequent text matching algorithm will not be applicable. Part of the text library generation effect is shown in Figure 4.
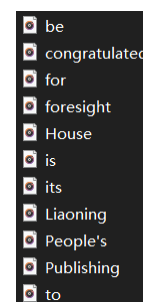

Figure 4: Text library effect   *Figure 5: Audio library*

2.4. Generation of audio library

Record the audio corresponding to the text in the text library, unified named format stored in the audio library folder. All the files are stored in MP3 format, and the audio files are read by using MATLAB's own Audioread function. Part of the audio library generation effect is shown in Figure 5.

2.5. Match the text and play the audio

Firstly, the image text is extracted, and then the circulation algorithm is used to search for the most matching text in the text library. The results are matched with the audio in the audio library, and the sound function of MATLAB is used to play the audio. First of all, the use of TXT file to generate text matching part of the experiment, that is, when the corresponding text matching, write the matching text into TXT file, in order to test the matching results. Take a photo of the first line of text and run the program. The running result is shown in Figure 6. The matching effect is good. Change the part of writing in the TXT file to play the corresponding audio. Audio playback effect because the program has a certain execution time, there is a certain delay between the words, for the auditory effect has a certain impact, but within the acceptable range.
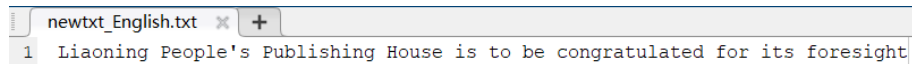
newtxt_English.txt

1 Liaoning People's Publishing House is to be congratulated for its foresight

Figure 6: Text matching experiment results

Both text matching and text segmentation retain all the attributes of the text image, which overcomes the distortion of the original attributes caused by the direct processing of the image text by the traditional algorithm. Combining image processing technology with statistical analysis method, the corresponding percentage of text matching degree is obtained. When the text matching degree percentage is greater than 85%, it will output.

3. Debugging process

The debugging process mainly focuses on the threshold part of image preprocessing and text segmentation.

3.1. Image preprocessing debugging

The image pretreatment of this paper mainly uses: open operation, gray, binarization, edge recognition. Open operation is corrosion first and then expansion: image expansion will enlarge the image components, and image corrosion will reduce the image components. - For open operation: after the image is corroded, the noise can be reduced and the image can be compressed to reduce the amount of subsequent program computation. The result of image opening operation is shown in Figure 7. Experimental results show that this method is not suitable for text template matching, which seriously affects the accuracy of matching results.
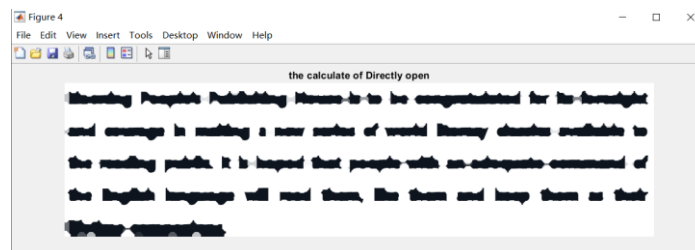

Figure 7: Image open operation

Edge recognition technology can be used for image recognition and text feature extraction. Edge detection is widely used in object and background segmentation, object and object segmentation, and is one of the important means of image segmentation [2]. The function of bwperim of matlab is used to find binary image edges for edge recognition. The result is shown in Figure 8.
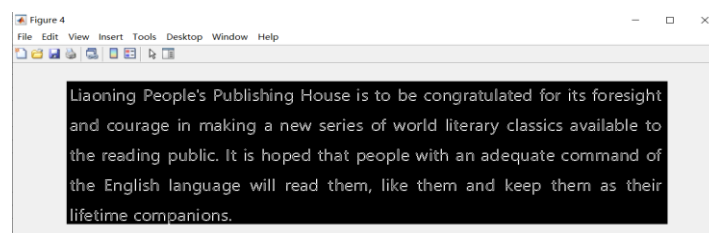

Figure 8: Edge recognition effect

The experimental results show that the pre-processing method of gray scale and then binarization is most suitable for this system.

3.2. Text threshold Debugging

At first, the minimum value of each row (or column) was selected for the segmentation operation, but there were many rows that were not segmented, so the threshold value needed to be limited to a reasonable range, and the rule was found by displaying the histogram of the total pixel value of each row. The bar chart is shown in Figure 9. As can be seen from the figure, pixel values of the text part are concentrated in [50,300], so pixel values below 50 are further screened, the proportion of pixel values of each row in the whole image is counted, and the results are arranged in order. By comparing with the total pixel value, the system calculates the corresponding proportion and converts it into threshold value. Finally, the segmented text is matched with the text database template.
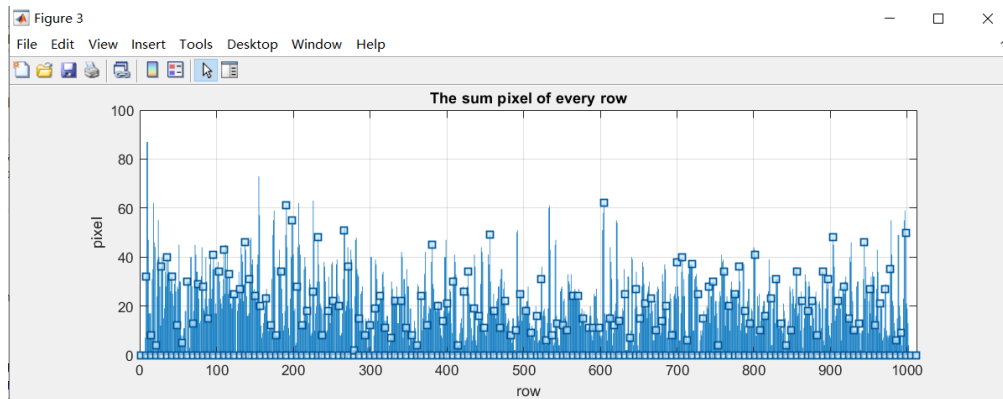

Figure 9: Sum of pixel values of each row

4. Conclusion

This paper proposes an image text to speech system based on MATLAB, through image correlation processing algorithm, template matching method and audio playback to achieve the connection and compatibility of each module. Through the debugging threshold, the text matching effect can reach the best, and finally generate the voice playback function. The text extracted and identified can be saved for further analysis and processing in the back end, which can be widely used in intelligent visual monitoring, image retrieval, image supplement and other aspects. This algorithm can be used as the algorithm support of APP development, and can also realize the interactive interface with users by using MATLAB GUI. The function of voice playback can not only help the blind to achieve fragmented reading, but also make the content of images more widely used in all aspects of scientific research.

References

[1] J. L. Nie. "The blind reading difficulties and safeguard measures of research [J],". Friends of the editor, 2018 (9): 11-15. DOI: 10.13786/j.carol carroll nki/g2.2018.9.002 cn14-1066.
[2] X. B. Yin. "License Plate Recognition System Based on MATLAB [D],". Zhejiang University of Technology. 2017.
[3] Y. Chen. "Based on the Matlab image character recognition research [J],". Journal of electronics, 2018 (14): 34-35 + 49. DOI: 10.16589/j.carol carroll nki cn11-3571/tn. 2018.14.016.
[4] W. Qin. "Research on image text location and extraction technology [D],". Harbin Engineering University. 2015.
[5] P. Yang. "Fruit recognition system based on matlab GUI [J],". Journal of electronics, 2021 (12): 38 and 39, DOI: 10.16589/j.carol carroll nki cn11-3571/tn. 2021.12.015.