Journal of Applied Science and Engineering, Vol. 21, No. 3, pp. 475–484 (2018) *DOI: 10.6180/jase.201809_21(3).0019*

A Reading Assistant System Based on Restoring Warped Document Image

Kuo-Ming Hung[1]*, Chi-Hsiao Yih[2] and Cheng-Hsiang Yeh[2]

[1]*Department of Information Management, Kainan University, Taoyuan, Taiwan 338, R.O.C*
[2]*Department of Electrical and Computer Engineering, Tamkang University, Tamsui, Taiwan 251, R.O.C*

Abstract

A novel reading assistant system for blind, visually impaired people or children is developed in this paper. The user can choose the region to be read and then receive output via text to speech. We design a helpful system that users can choose the area (on the book) they want to listen when reading and the device can also speak out the entire article. Besides, users can read without distance perception by using this system. The step to construct this system is according to the following: first, this system is used depth images to detect the coordinates of the user's finger with Kinect, and the book is also input with Kinect. Next, we utilize text preprocessing to remove the border of the book and then take advantage of the common attribute of text height to eliminate images and tables. Beside, we also use a method to combine non-linear and linear compensation for correcting distortions of document images. Experimental results demonstrate that the proposed scheme has good performance and demonstrates robustness. It also provides an effective interactive platform.

Key Words: Assistant System, Restoring Warped, Text Recognition, Kinect, Depth Image

1. Introduction

According to the World Health Organization (WHO), 285 million people are visually impaired, of whom 39 million are blind and 246 million have low vision [1]. For them, many of the routine tasks associated with daily life can be frustrating and most of the information they need for daily life, which often exists in written or image form, is not easily accessible. Fortunately, assistive technology can help them with daily tasks and also help them independently perform certain activities, such as reading documents, communicating and searching for information on the Internet.

The Braille system was devised by Frenchman Louis Braille in 1821. It is a tactile writing system used for books, menus, signs and so on. Braille-users can read computer screens and other electronic support thanks to refreshable braille displays. However, learning this system requires repetition and training with a person who can help them through the learning process. Moreover, it is a very difficult task for elderly without assistance from their family. Araki et al. in [2] developed a spoken dialogue system allowing visually impaired individuals to self-learn Braille. Hira et al. [3] designed a wearable system that converts digital text files into tactile signals using USB port. This system will be based on MEMS (Micro-electromechanical System) actuators that would mimic these Braille dots. [4–7] used image processing technique to transform Braille cells into text or voice. Velazquez et al. in [8] developed a reading assistant device that is able to reproduce electronic books (eBooks) in portable electronic tactile displays. This system translated eBooks into Braille by computer and stored it in a regular USB flash drive inserting into the TactoBook consisting of a

*Corresponding author. E-mail: hkming@mail.knu.edu.tw

computer-based system allowing the user read eBooks. In this way, the user is able to read numerous eBooks comfortably without carrying the burdensome tactile print versions. Raisamo et al. [9] proposed three kinds of Braille presentation methods based on the Nokia 770 internet tablet without an additional displayed device. Nevertheless, the visually impaired couldn't read historical documents that are not in the e-document form with those methods. Fortunately, [10] described an embedded device dedicated to solving this problem. It was composed of three main separated stages: text detection, optical character recognition and a human-machine interface. Using a smartphone, taking a picture and hearing the text that exist in the picture have been reported in [11]. They used Tesseract framework and image filters in order to make a preprocessing of the images so that this system could get better results. After running several tests, it was shown that the combination of filters CIColorControls and CIColorMonochrome produced the best results. This system allowed the reading of texts, provided that the light conditions are the ideal for image recording and the equipment is properly directed to the text you want to listen to so that recognition, but the document cannot be warped. To further assist in this, a reading assistant software product has been developed [12], which has some advantages to this product, such as rapid processing and ease of operation. However, the system might display shadows if the document has warped images, resulting in word recognition error. Moreover, [13] propose a high resolution automatic page turning image recognition system. The author uses two high resolution cameras to extract the 3D shape of the book before recognizing. Although this system can recognize up to 300 pages per minute, it has its shortcomings as the equipment is too expensive, users can't assign the vocabulary they want to read and can't read comfortably.

To overcome such problems, we propose a reading system that can accept documents in either hard copy or warped document and the user needs only to point at the document with his finger. The rest of the paper is organized as follows. Section 2 presents the proposed reading assistant system. The experimental results are fully described in section 3. Section 4 concludes the paper by highlighting the main conclusions and future areas of study.

2. Proposed Method

Figure 1 shows the flowchart of the proposed framework, which is made up of three main blocks and is described as follows.

2.1 Document Preprocessing

To extract text from a page document, we remove the noisy black borders of each page and cut individual pages. After this stage, the pages can have text extracted in the following text extraction stage.

2.1.1 Page Segmentation

In this section, we modify the automatic border detection method for documents of Stamatopoulos et al. [14] to enhance performance. First, the adaptive image binarization [15] is processed to classify image pixels into foreground or background classes. Since the purpose of this stage is classification, we replace adaptive image binarization with the Otsu [16] method to enhance performance. The run length smoothing algorithm (RLSA) [17] is utilized to examine the pixel runs in the horizontal and vertical directions and these are eliminated by the empirical value. Subsequently, Connected Component Labeling [18] is carried out to make the text lines. The RLSA algorithm and CCL method are used for words to connect the text lines, so we perform the horizontal dilation operation, as shown in Figure 2, to enhance performance. The other procedures of page segmentation have already been described in detail [14]. Figure 3 shows the text regions after noisy black border detection and projections of the image. Figure 4 shows the result of removing the black border and cutting the page.

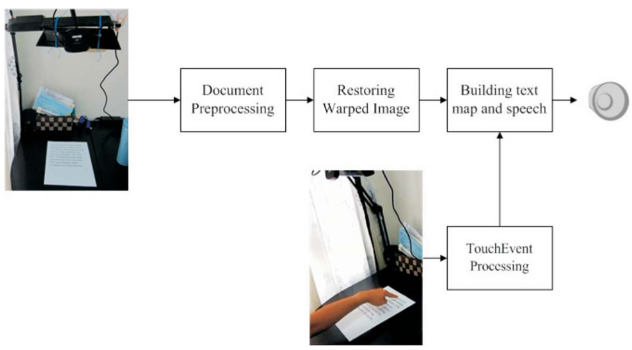


Figure 1. Flowchart of the proposed framework with Microsoft Kinect.

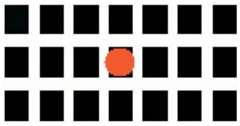


Figure 2. Mask of horizontal dilation.

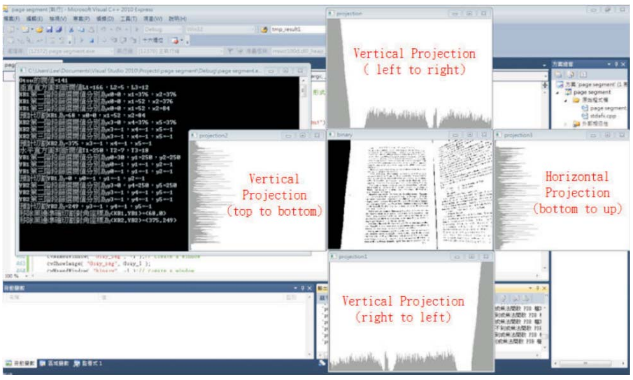


Figure 3. Projections of image and text regions detection.

(a) (b) (c) (d)

Figure 4. (a) and (c) Original document image; (b) and (d) Removing black border and cutting page result.

2.1.2 Text Extraction

• Removing graphs and noise

Before proceeding with text enhancement and binarization, we remove components such as pictures, graphs, and noise. Intuitively, the text has the following distinguishing characteristics:

(1) Text possesses a certain amount of frequency and orientation information.

(2) Text shows spatial cohesion – characters of the same text string (a word, or words in the same line) are of similar height, orientation, and spacing.

Thus, the most intuitive characteristics of text are its regularity. Although text extraction methods have been proposed [19,20], a faster and more effective method is needed. In this paper, we modify a previous text segmentation method [21] to reduce computational complexity and increase accuracy. The steps of the method for removing graphs and noise are shown in Figure 5. Specifically, the following steps describe the text segmentation:

Step 1: Transform the color image to a gray image after the page segmentation stage using Eq. (1).

$$I_{gary} = \frac{(I_R + I_G + I_B)}{3} \tag{1}$$

Step 2: Use Otsu's thresholding method after color to gray, as shown in Figure 6(b).

Step 3: To label components rapidly, use a horizontal dilation operation to obtain words to connect text lines, as shown in Figure 6(c).

Step 4: Use labeling to make it easier to classify the objects. After observations, this function can be terminated, and then the performance is better. The connected component method (CCM) and the region growth method [22] are the most common methods of labeling. The connected component method is used for a 2-D binary image. It scans an image, pixel-by-pixel (from top to bottom and left to right), in order to identify connected pixel regions, i.e., regions of adjacent pixels, that share the same set of intensity values. CCM can use either a 4-connected component or 8-connected component for two dimensions. The connected component method can be a 6-con-

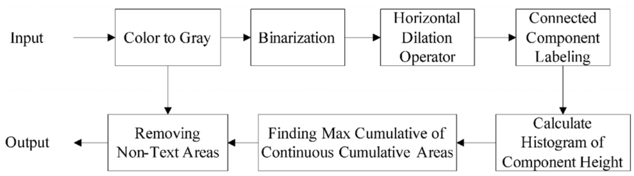


Figure 5. Flow char of text extraction.

(a) (b) (c) (d) (e)

Figure 6. Removing graphs and noises process. (a) Original image. (b) Binarization result. (c) Horizontal dilation result. (d) Result of removing graphs and noises from binarization image. (e) Result of removing graphs and noises from gray image.

nected neighborhood, an 18-connected neighborhood, or a 26-connected neighborhood for three dimensions. The disadvantage of the connected component method, however, is that it is time-consuming. Thus, we use our previous method [23] to solve this disadvantage of region-growing techniques.

Step 5: Stamatopoulos et al. [21] accorded the maximum value of the histogram to remove the connected components. When the stage of binarization processing generates excessive noise, the estimated height of the histogram is not the text height because the noise of the connected components amounts to more than the texts of the connected components. Thus, character height (CH) is denoted as the maximum value of the section, as shown in Figure 7. Then remove connected components which satisfy the following condition:

$$h > 3 \times CH \text{ or } h < \frac{CH}{4} \text{ or } w < \frac{CH}{4} \tag{2}$$

where h and w denote the connected component's height and width, respectively. Figure 6(d) shows the result of removing the graphs and noise from the binarized image.

Step 6: Remove the graphs and noise according to the result of removing the graphs and noise of the binarized image in the gray image, as shown in Figure 6(e), and output this image result.

• Text enhancement and binarization

Specifically in the document binarization field, this local intensity variation appears quite often and results from various factors such as uneven illumination, stains, and the texture of paper. Therefore, we apply a preprocessing step to extract the binarized text from original distorted image with a different light source. The flowchart of the algorithm is shown in Figure 8 and is described as follows:

Figure 7. Histogram of components height.

Step 1: Estimate the background using a continuous-scale morphological closing operator.

Step 2: Obtain the foreground by subtracting the background from the image of the previous stage.

Step 3: Use Otsu's thresholding method to extract text.

Comparing our method with the Otsu method, as shown in Figure 9, indicates that our method is more robust than that of Otsu under different light.

2.2 Restoring the Warped Image

In this session, we use our previous method [24] for restoring warped images, relying on 2-D information in a two-stage process

• Non-linear compensation

Step 1: Perform the dilated operation making the text block.

Step 2: Use the fast connected component labeling method [23].

Step 3: Perform fast connected component cutting with a certain vertical intervals, and the point of the baseline of the text line is cut out.

Step 4: The natural cubic curve of the baseline obtained using the least squares method, as follows:

$$y_{im} = a_i x_m^3 + b_i x_m^2 + c_i x_m + d_i \tag{3}$$

where the y_i represents each line in the document and i = 1, 2, 3 … n. The a_i, b_i, c_i and d_i in formula (3) respectively represent the same text line equation coefficients,

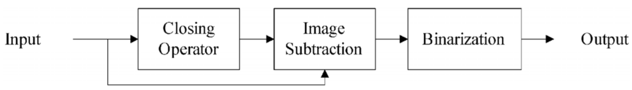


Figure 8. Text enhancement and binarization process.

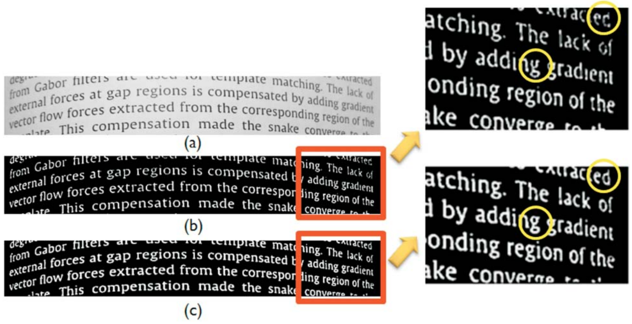


Figure 9. (a) Original image. (b) Using Otsu's method. (c) Using our method.

and are obtained by the least squares method. In addition, m represents the x-axis coordinates of each connected component as $x_m = x_0, x_1, x_2 \dots x_{L-1}$, where L is the length of each connected component. We use formula (3) to calculate the height position y_{i0} of the first point of the left and also calculate the height difference of the other points y_{im} to the position and then use the text vertical rectification operation. The rectification rule is:

$$y'_{im} = y_{im} + \Delta y \quad \text{where } \Delta y = y_{i0} - y_{im} \tag{4}$$

• Linear compensation

The height of the document have already been adjusted by non-linear compensation method, but it still need to modify the width of the document to improve readability and text recognition rate. First, use the vertical projection method to extract the words. Then the width of the word can be calculatedas follows:

$$x'_{ij} = x_{ij} + dx_{ij} \quad \text{where } dx_{ij} = x'_{ij} \times \frac{\Delta x}{Word\ length \times \cos\theta} \tag{5}$$

where the x_ij is the $i_$th row of the $j_$th word of the text line. Δx = Word length-Word length × cos θ.

2.3 TouchEvent

In this session, we use our previous method [25] for TouchEvent with Microsoft Kinect, and is described as follows. First, a skin detection method is used to find the user's hand. Then the boundaries of the hand are extracted. After extracting the hand boundary, the angle value is chosen for experimental purposes to detect the fingers. Finally, we use the depth of a user's finger and the depth of tabletop at the moment to detect a finger when the user wishes to read a document, as shown in Figure 10.

2.4 Building Text Map and Speech

After the stage of restoring the warped images, we build a text map that is output to the user by text to speech. This section is divided into three stages, described as follows.

• Building the text coordinate index

Step 1: To build a text coordinate index, first segment restorations of the warped image by horizontal projection profile.

Step 2: Use a horizontal dilation operator to connect the letters into words.

Step 3: Segment the sentence into words by using a vertical projection profile operator.

Step 4: Character recognition uses Tesseract [26], Google's OCR engine.

Step 5: Build a sentence map and word map according to the coordinates of the horizontal and the vertical projection profile operator, respectively.

Figure 11 shows the processes of building a text coordinate index.

• Getting coordinate form user and text to speech

After building the coordinate databases of the sentences and word maps, the user needs to touch the word on a Kinect or a Touchscreen, and then the system will search for words in the coordinate databases. Finally, the system produces output via text to speech [27].

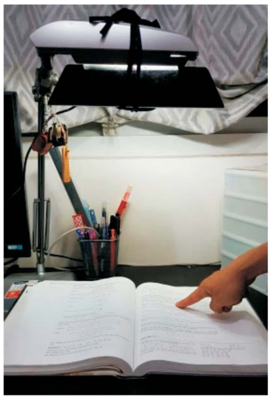
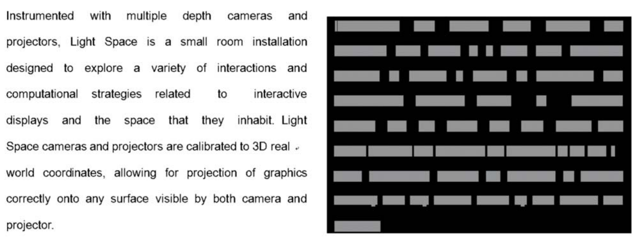


Figure 10. Example of TouchEvent with Kinect.



(a) Original image (b) The result

Figure 11. The processes of building text coordinate index.

3. Experimental Results

To test the effectiveness of our method, paper documents with curved surfaces were used in the experiment. We took pictures of the document with Kinect as shown in Figure 12. The resolution of each image was 1920 × 1080 pixels. Our documents contained figures, equations, tables and warped surface shapes. On the reading aspect, our system not only can speak out words when users tap somewhere on the physical book, but also can read aloud the full text. Therefore, it can facilitate the blind when they read it.

3.1 The Evaluation of Recognition

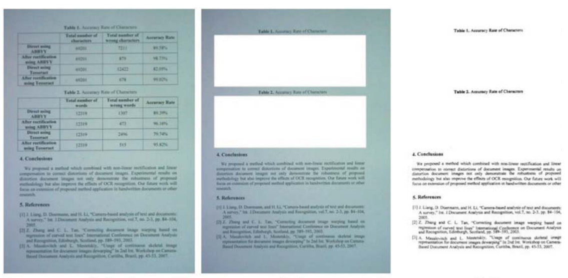
We used a dataset of 40 English document images and 32 Chinese document images, containing both modern and historical printed document images in English or Chinese. The commercial software Tesseract character recognition processing is applied. Figures 13 and 14 show that our system can effectively remove tables and graphs from documents to be read by the user. The correction results are illustrated in Figure 15, in which (a) the distorted image of the book pages, (b) distorted document images containing mathematical formula, (c) the document images containing different language. The corrected results show that the proposed method can achieve effective rectification as long as each text line is a continuous line. For the evaluation of rectification techniques, OCR is widely used as a means of indirect evaluation [19]. Character and word accuracy, as defined below, are used as the evaluation metrics for the OCR results:

$$Character\ Accuracy = \frac{(characters - errors)}{characters} \tag{6}$$

$$Word\ Accuracy = \frac{(words - misrecognized\_words)}{words} \tag{7}$$

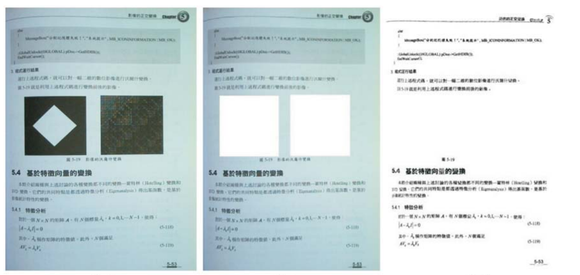
Table 1 illustrates the average character accuracy results. Table 2 illustrates the average word accuracy. In Table 1 and Table 2 show the character accuracy and word accuracy for English by Tesseract are 82.05% and 79.74% respectively. The character accuracy and word accuracy for English by dewarping using Tesseract are 99.02% and 95.82% respectively. Our proposed rectification method improves accuracy by 17% and 16%. In addition, in Table 3 and Table 4 show the character accuracy and word accuracy for Chinese by Tesseract are 77.95% and 76.40% respectively. The character accuracy and word accuracy for Chinese by dewarping using Tesseract are 97.61% and 92.01% respectively. Our proposed rectification method improves accuracy by 19.66% and 16.61%. The experimental results obtained show using restoring warped technique as preprocessing can enhance recognition accuracy, effectively. Our

Figure 12. Kinect.



(a) (b) (c)

Figure 13. (a) Original image. (b) The result of removing tables. (c) The correction result of document images.



(a) (b) (c)

Figure 14. (a) Original image. (b) The result of removing graphs. (c) The correction result of document images.

(a) (b) (c) (d)

Figure 15. (a) and (c) Original images. (b) and (d) The correction result of document images.

Table 1. Average character accuracy on 40 document images (English)

	Characters	Errors	Character accuracy
Without dewarping using Tesseract	69201	12422	82.05%
Dewarping using Tesseract	69201	678	99.02%

Table 2. Average word accuracy on 40 document images (English)

	Words	Misrecognized word	Word accuracy
Without dewarping using Tesseract	12319	2496	79.74%
Dewarping using Tesseract	12319	515	95.82%

Table 3. Average character accuracy on 32 document images (Chinese)

	Characters	Errors	Character accuracy
Without dewarping using Tesseract	51360	11321	77.95%
Dewarping using Tesseract	51360	1223	97.61%

Table 4. Average word accuracy on 32 document images (Chinese)

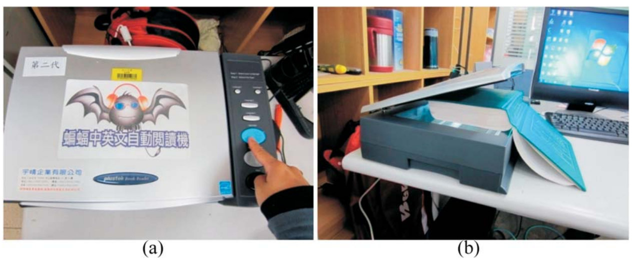
	Words	Misrecognized word	Word accuracy
Without dewarping using Tesseract	10555	2490	76.40%
Dewarping using Tesseract	10555	843	92.01%

experimental results demonstrate the robustness of our reading assistant system, which achieved remarkable improvements in OCR accuracy in different environments. Our approach requires an average of 10.74 seconds to process one page.

3.2 The Evaluation of the System by Blind and Blindfolded Participants

Two blind university students, two visually impaired university students and twenty university students were used to evaluate the system. This system is not meant to take the place of traditional braille system or other assistant systems, but to improve convenience and intuition speaking out the area and the entire article on the book when they want to listen. Our system has some limitations. First, the blind user can only use to speak out the entire article on the book, because he does not look at the reading area. Second, this system cannot recognize successfully when the light is more insufficient. Third, this system cannot recognize successfully when the font is smaller and this depends on the resolution of the Kinect lens. Traditional braille system have often used on disabled space and this system has an advantage which cost of maintenance is lower. In addition, others reading assistant system also have an advantage which speaking outthe entire article quickly.

The participants consisted of two blind university students (blind participants: BP), two visually impaired university students (visually impaired participants: VIP) and twenty university students (students participants: SP). The participants used our system and automatic reading machine [12] as shown in Figure 16 to do satisfaction survey. These items of the score include intuition, performance and convenience, respectively. Intuition is defined that there's no clear evidence one way or the other and you just have to operate this system on intuition when users use the system for the first time. Performance is defined as the system executes the reactionin using this system. Convenience is defined as the degree of convenience in using this system. The Likert Scale 5 level is used in this statistical data of satisfaction. Figures 17 and 18 show that the results of participants for our system and automatic reading machine after scoring. The format of a five-level Likert item for satisfaction as: largely disagree, disagree, nature, agree and largely agree. Satisfaction calculation formula = (largely disagree × 0 point + disagree × 1 point + nature × 2 point + agree × 3 point +



(a) (b)

Figure 16. Automatic reading machine [12].

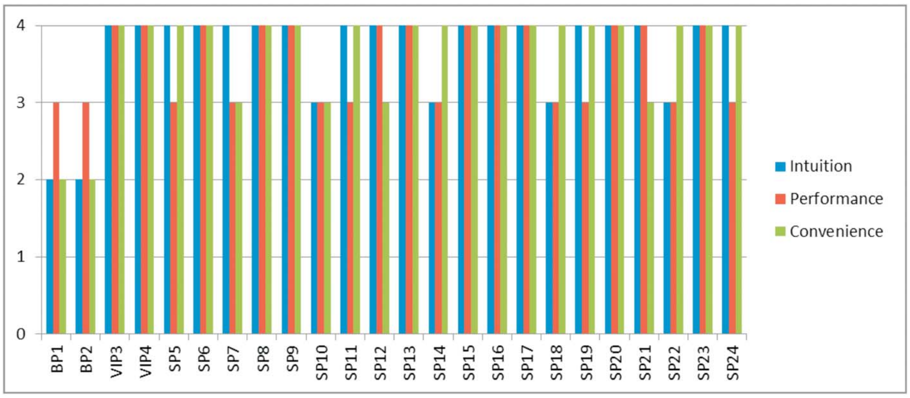


Figure 17. The statistical data of satisfaction (our system).

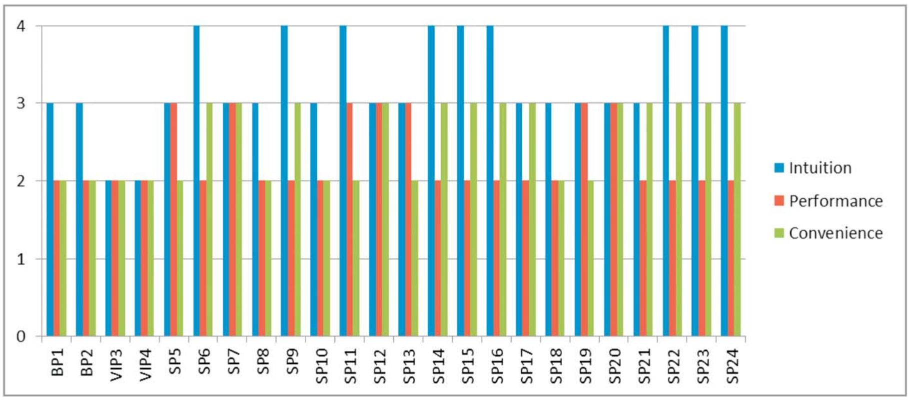


Figure 18. The statistical data of satisfaction [12].

largely agree × 4 point) / total count). The experimental results obtained show it is not more intuition and convenience for blind, because they cannot look at that area on the book in using our system. Besides, visually impaired university students and university students can get good results of intuition, performance and convenience in using our system. Referring to Figure 19, it can be observed that comparison results can reflect the advantage of our scheme.

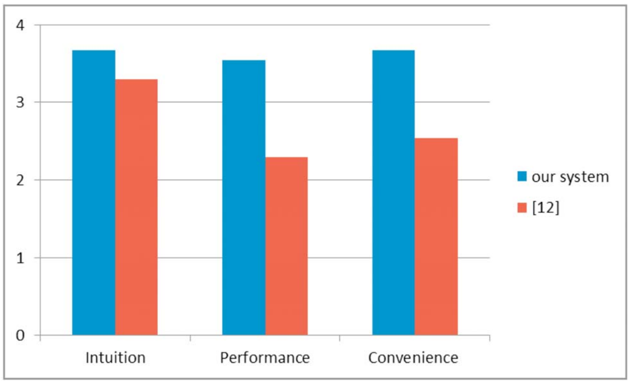


Figure 19. Comparison our system with [12].

4. Conclusions and Future Work

This paper presents a reading assistant system for blind, visually impaired people or children, considering document preprocessing, warped image and TouchEvent, to create an application that is gradually improved and refined over the process. First, the noisy black borders of a page are removed and the pages are cut to extract text from each page document. Then components such as pictures, graphs, and noise, are removed. Document imaging is acquired by scanners or cameras, which often lead to a variety of different distortions because of the object's volume or environment. Furthermore, local intensity variation appears quite often, due to various factors such as uneven illumination, stains, and the texture of paper. Thus we adopt the enhancement technique and the method for restoring warped images to solve the above problems. In addition, we also present a reading assistance which uses depth images to detect the coordinates of a user's finger with Kinect. Experimental results on several document images, including warped surface, graphs, and tables, show the proposed methods perform well and are robust. Moreover, the interactive platforms allow users to point to a word with their finger on the Document or Touchscreen, providing an effective interactive platform. Our future research will focus on optical character recognition and speech synthesis.

Acknowledgments

This work was supported by the National Science Council under grant number NSC 101-2410-H-032-054-.

References

[1] WHO Report: Visual impairment and blindness (2014).

[2] Araki, M., Shibahara, K. and Mizukami, Y., "Spoken Dialogue System for Learning Braille," 35th IEEE Annual Computer Software and Applications Conference, pp. 152–156 (2011). doi: 10.1109/COMPSAC.2011.27

[3] Arshad, H., Khan, U. S. and Izhar, U., "MEMS Based Braille System," 2015 IEEE 15th International Conference on Nanotechnology, pp. 1103–1106 (2015). doi: 10.1109/NANO.2015.7388815

[4] Matsuda, Y., Isomura, T., Sakuma, I., Kobayashi, E., Jimbo, Y. and Arafune, T., "Finger Braille Teaching System for People who Communicate with Deafblind People," Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, pp. 3202–3207 (2007). doi: 10.1109/ICMA.2007.4304074

[5] Al-Shamma, S. D. and Fathi, S., "Arabic Braille Recognition and Transcription into Text and Voice," 5th Cairo International Biomedical Engineering Conference Cairo, pp. 16–18 (2010). doi: 10.1109/CIBEC.2010.5716095

[6] Abdul Malik, S. A. S., Ali, E. Z., Yousef, A. S., Khaled, A. H. and Abdul Aziz, O. A. Q., "An Efficient Braille Cells Recognition," 6th International Conference on Wireless Communications Networking and Mobile Computing (WICOM), pp. 1–4 (2010). doi: 10.1109/WICOM.2010.5601020

[7] Onur, K., "Braille-2 Otomatik Yorumlama Sistemi," Signal Processing and Communications Applications Conference (SIU), pp. 1562–1565 (2015). doi: 10.1109/SIU.2015.7130146

[8] Velazquez, R., Preza, E. and Hernandez, H., "Making eBooks Accessible to Blind Braille Readers," IEEE International Workshop on Haptic Audio Visual Environments and Their Applications, pp. 25–29 (2008). doi: 10.1109/HAVE.2008.4685293

[9] Rantala, J., Raisamo, R., Lylykangas, J., Surakka, V., Raisamo, J., Salminen, K., Pakkanen, T. and Hippula, A., "Methods for Presenting Braille Characters on a Mobile Device with a Touchscreen and Tactile Feedback," *IEEE Transactions on Haptics*, Vol. 2, No. 1, pp. 28–39 (2009). doi: 10.1109/TOH.2009.3

[10] Gaudissart, V., Ferreira, S., Thillou, C. and Gosselin, B., "Mobile Reading Assistant for Blind People," Proceedings of European Signal Processing Conference, pp. 538–544 (2005).

[11] Neto, R. and Fonseca, N., "Camera Reading for Blind People," *Procedia Technology*, Vol. 16, pp. 1200–1209 (2014). doi: 10.1016/j.protcy.2014.10.135

[12] http://www.u-tran.com/index.php, visited in May (2016).

[13] Noguchi, S. and Yamada, M., "Real-time 3D Page Tracking and Book Status Recognition for High-speed Book Digitization Based on Adaptive Capturing," 2014 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 24–26 (2014).

[14] Stamatopoulos, N., Gatos, B. and Kesidis, A., "Automatic Borders Detection of Camera Document Images," Int. Workshop Camera-Based Document Anal. Recognition Conf. (CDBAR), pp. 71–78 (2007).

[15] Gatos, B., Pratikakis, I. and Perantonis, S. J., "Adaptive Degraded Document Image Binarization," *Pattern Recognition*, Vol. 39, pp. 317–327 (2006). doi: 10.1016/j.patcog.2005.09.010

[16] Otsu, N., "A Threshold Selection Method from Gray-level Histogram," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, pp. 62–66 (1976). doi: 10.1109/TSMC.1979.4310076

[17] Wahl, F. M., Wong, K. Y. and Casey, R. G., "Block Segmentation and Text Extraction in Mixed Text/Image Documents", *Computer Graphics and Image Processing*, Vol. 20, pp. 375–390 (1982). doi: 10.1016/0146-664X(82)90059-4

[18] Chang, F., Chen, C. J. and Lu, C. J., "A Linear-time

Component-labeling Algorithm Using Contour Tracing Technique," *Computer Vision and Image Understanding*, Vol. 93, No. 2, pp. 206–220 (2004). doi: 10.1016/j.cviu.2003.09.002
[19] Jiang, H., "Research on the Document Image Segmentation Based on the LDA Model," *Advances in Information Sciences and Service Sciences (AISS)*, Vol. 4, No. 3, pp. 12–18 (2012). doi: 10.4156/aiss.vol4.issue3.2
[20] Xiaoying, Z., "Study on Document Image Segmentation Techniques Based on Improved Partial Differential Equations," *Journal of Convergence Information Technology (JCIT)*, Vol. 8, No. 5, pp. 821–831 (2013). doi: 10.4156/jcit.vol8.issue5.96
[21] Stamatopoulos, N., Gatos, B., Pratikakis, I. and Perantonis, S. J., "Goal-oriented Rectification of Camera-based Document Images," *IEEE Trans. on Image Processing*, Vol. 20, No. 4, pp. 910–920 (2011). doi: 10.1109/TIP.2010.2080280
[22] Gao, Y., Ai, X., Rarity, J. and Dahnoun, N., "Obstacle Detection with 3D Camera Using U-V Disparity," Proceedings of the 2011 7th International Workshop on Systems, Signal Processing and Their Applications (WOSSPA), pp. 239–242 (2011). doi: 10.1109/WOSSPA.2011.5931462
[23] Huang, H. C., Hsieh, C. T. and Yeh, C. H., "An Indoor Obstacle Detection System Using Depth Information," *Sensors (Basel)*, Vol. 15, No. 10, pp. 27116–27141 (2015). doi: 10.3390/s151027116
[24] Hsieh, C. T., Lee, S. C. and Yeh, C. H., "Restoring Warped Document Image Based on Text Line Correction," 2013 9th International Conference on Computing Technology and Information Management (ICCM 2013), Vol. 14, pp. 459–464 (2013).
[25] Hsieh, C. T., Yeh, C. H., Liu, T. T. and Huang, K. C., "Non-visual Document Recognition for Blind Reading Assistant System," 2013 9th International Conference on Computing Technology and Information Management (ICCM 2013), Vol. 14, pp. 453–458 (2013).
[26] Tesseract: https://github.com/tesseract-ocr, visited in May (2016).
[27] Text-to-speech: http://msdn.microsoft.com/en-us/library/ms720163.aspx, visited in May (2016).

Manuscript Received: Dec. 25, 2017
Accepted: Jan. 18, 2018