Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **4**, July **2018**

BlindReader: An Assistant System to Support Text Reading for The Visually Impaired People

Fan Guo, Han-Dong He, Yan-Ru Xiao, Teng-Mu Fu, and Cheng Liao
School of Information Science and Engineering
Central South University
Changsha 410083, China
guofancsu@163.com; 631807642@qq.com; 568868065@qq.com;
771860992@qq.com; 3239892939@qq.com

Ping Tan*
Key Laboratory of Hunan Province for New Retail Virtual Reality Technology
School of Computer and Information Engineering
Hunan University of Commerce
Changsha 410205, China
*Corresponding author: tp2008@foxmail.com

Received March, 2017; revised February, 2018

Abstract. *Accessing printed text is a great challenge for the visually impaired people. Therefore, in this paper, we introduce a visual assistant system, BlindReader, that assists visually impaired people to read printed text. Some computer vision algorithms are introduced to adjust the intensity and orientation of captured document images in preprocessing stage. Using the system, the low-vision users can read the text more clearly just like the way people use the magnifying glass and the blind users can hear the recognized words synthesized to audio speech for the whole text or the single text line. Finally, the system is evaluated in terms of text recognition accuracy and user feedback to determine the usability of the BlindReader.*

Keywords: Visual impaired people, Assistant system, Layout analysis, Text recognition, Text reading

1. **Introduction.** There are numerous difficulties for visually impaired people to access printed text using existing technology. These difficulties include problems with alignment, focus, accuracy, and efficiency, etc. Besides, accessing text in non-ideal conditions (i.e. low lighting, unique page orientations, etc.) is another difficulty. All these factors make it a daunting task. Interviews we conducted with visually impaired people revealed that available technologies, such as screen readers, desktop scanners, and smartphone applications, are commonly under-utilized due to slow processing speeds or poor accuracy. Since the existing technological barriers in habit visually impaired people's abilities to gain more independence and learning opportunities, we present our work towards creating a device and the corresponding software that could overcome some issues that current technologies pose to visually impaired users. The proposed visual assistant system, BlindReader, can assist visually impaired people to read printed text. Some computer vision algorithms are introduced to adjust the intensity and orientation of captured document images in preprocessing stage. Using the system, the low-vision users can read the text more clearly just like the way people use the magnifying glass and the blind users can hear the recognized

words synthesized to audio speech for the whole text or the single text line. Therefore, it is very meaningful to the visually impaired people. The main contributions of this paper are as follows.

First, we conceptualize and implement BlindReader, an assistant system to support text reading for the visually impaired people. Using the system, the low-vision user can read the text more clearly just like the way people use the magnifying glass and the blind users can hear the recognized words synthesized to audio speech for the whole text or the single text line as well.

Second, our proposed method utilizes computer vision algorithms to adjust the intensity and orientation of captured document images in preprocessing stage. For effectively reading the printed text, audio output is realized by using the Flit Text-to-Speech (TTS) engine [1].

Third, we report findings from two evaluations: a technical evaluation to understand the text recognition accuracy, and user feedback sessions with visual impaired participants to assess the proposed system.

2. **Related work.** Researchers in both academia and industry exhibited a keen interest in aiding visually impaired people to read printed text. Many assistive text reading products appear in nowadays market. Besides, the core technologies of these products are text detection and recognition algorithms. Therefore, we discuss related work in two aspects: i) assistive text reading products, and ii) text detection and recognition algorithms.

2.1. **Assistive text reading products.** Nowadays, most assistive text reading products are handheld or mobile devices. For example, the VizWiz mobile assistive application takes a different approach by offloading the computation to humans, although it enables far more complex features than simply reading text, it lacks real time response [2]. The SYPOLE developed by Mancas-Thillou *et al.* consisted of a camera phone/PDA to recognize banknotes, barcodes and labels on various objects [3].

Meanwhile, mobile phone devices are very prolific in the community of blind users for their availability, connectivity and assistive operation modes, therefore many applications were built on top of them [4]: the kNFB kReader, Blindsight's Text Detective, ABBYY's Text Grabber, StandScan, SayText, ZoomReader and Prizmo. Meijer's vOICe for Android project is an algorithm that translates a scene to sound; recently they introduced OCR capabilities and enabling usage of Google Glass. ABiSee's EyePal ROL is a portable reading device. Since the device is quite large and heavy, thus OrCam's recent assistive eyeglasses or the Intel Reader present a more lightweight alternative.

All the above products need mobile devices, which are usually hard to manipulate for visually impaired people. In contrast, our system only requires imaging the text use a special stand, and then the system will automatically transform the printed text to audio speech. Therefore, a more easy-use reading experience can be created.

2.2. **Text detection and recognition algorithms.** The existing text detection methods can be categorized into two groups: region-based and connected components-based methods [5]. The region-based method usually apply sliding windows to scan original image and obtain a set of small patches, which will be classified into text and non-text groups based on their textural features such as histogram of oriented gradient (HOG), discrete cosine transform (DCT) wavelet transform and so on. The connected components-based method divides text detection work into three sub tasks: 1) extracting connected components from the original image using pixel properties; 2) producing text string from the obtained connected components; and 3) identifying candidates' text string to obtain final text detection results. In contrast to the region-based methods, the components-based methods have better results and consume less time. Thus, we present a connected components-based method to detect each text line of the input document images.

Once the text is detected, it can be recognized by using the Optical Character Recognition (OCR) technology. There are many OCR engines can be used for text recognition, such as Tesseract [6], OCRopus etc. we can directly use the mature OCR engine to effectively recognize printed text. Therefore, text detection and text recognition can be closely related problems since good text detection results ensure a high text recognition rate.

3. **System overview.** The basic idea of our system is to detect each text line based on the layout analysis of captured document images, and then recognize the text using OCR engine. Finally, the system outputs the OCR results in the form of audio speech.

The system is divided into four major modules as shown in Fig. 1. The first module is to capture document images by using a digital camera.

The second module is to adopt computer vision algorithms to adjust the intensity and orientation of captured document images. As the image preprocessing stage, the module provides ideal text visual effects for the following operations.

In the third module, some functions, such as foreground/background (F/B) color transform, magnifying glass effect, are developed for low-vision users. For the purposes of discussion, here we define the text characters in the document image as the foreground and other parts of the image as the background. The color of the image foreground and background is transformed to make the text characters more obvious for low-vision users. Similarly, the magnifying glass effect is realized as a useful tool to help low-vision user read text more clearly.

In the last module, the ideal text image obtained by the image preprocessing is first adopted to extract each text line through the layout analysis, and then OCR engine is used to automatically recognize these text lines. Based on the OCR results, the blind users can hear the recognized words synthesized to audio speech for the whole text or the single text line, as shown in Fig. 1. Evaluation results demonstrate that good text recognition accuracy and user feedback can be obtained by using our system.

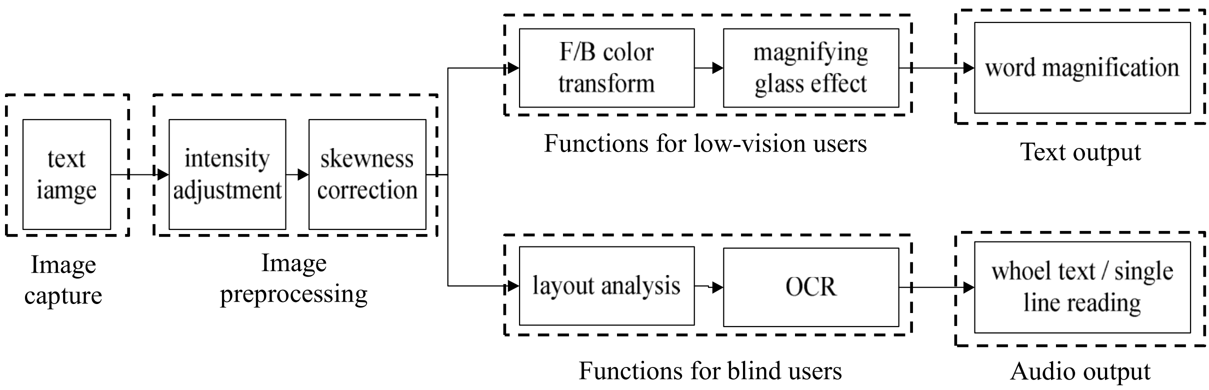
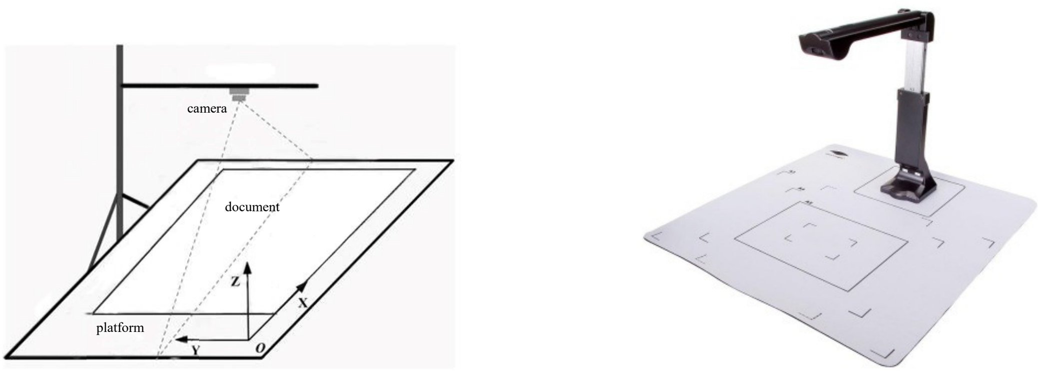


FIGURE 1. Our system framework

4. **BlindReader: An assistant system.** In this section, we will introduce the details of our system, including hardware details and software details.

4.1. **Hardware details.** We designed a system to acquire document images. The system is illustrated in Fig. 2(a), which consists of a high-resolution digital camera. The camera is fixed on the top bracket and has a fixed focus, which is used to capture high-resolution images of the document page.



(a) (b)

FIGURE 2. BlindReader prototypes

Using a high-resolution digital camera to capture document images is straightforward. To begin with, we need to control the angle and height of the digital camera. Once the camera is fixed, the position to place the document should be considered. Therefore, a board with some position marks is used for this purpose, as shown in Fig. 2(b). By using the BlindReader hardware (see Fig. 2b), a text image with relatively good quality can be captured for the BlindReader software.

4.2. **Software details.** We developed a software system that includes four modules: image capture, image preprocessing, functions for low-vision users, and functions for blind users, as shown in Fig. 1. In this section, we will discuss the four modules in details.

4.2.1. *Image capture.* The image capture module automatically detects printed text and captures the text image by the BlindReader hardware, as shown in Fig. 2. Once the module detects the text page is updates, it will automatically capture the text image. Otherwise, the camera will not capture images since the text page stay the same.

TABLE 1. The pseudo-code of the document page update detection

Algorithm 1: Document page update detection

```
Input: current frame, background image
Output: true or false
Frame = getFrame;
While 1
   Frame_pre = Frame;
   Frame = getFrame;
   If similarity(Frame, Frame_pre) > θ1 & similarity(Frame, Background_img) < θ2
      Return true
   End
End
```

Specifically, the basic idea of update detection algorithm of document page is based on the similarity measurement between previous frame and current frame of the captured image sequence. If the similarity is greater than θ_1, the text page can be regarded as remain unchanged. However, we find that when no document is placed under the camera, the similarity also satisfy the condition. Thus, the similarity between current frame and background image is also computed. If this similarity is less than θ_2, the former situation can be excluded. The algorithm of the document page update detection is presented in Tab. 1.

Using the algorithm, the text image can be automatically captured with no user interaction. However, the quality of the captured images is often not good due to the low luminance, light reflection, document place position, etc. Therefore, image preprocessing is very necessary to remove these disturbing factors.

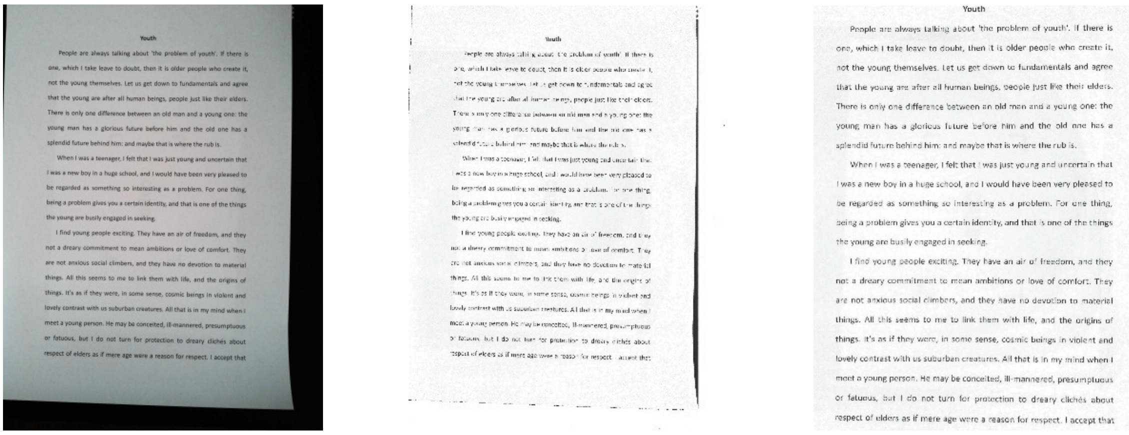
4.2.2. *Image preprocessing.* In our experiment we find that even though the directly captured image is enough to display the text contents, the quality of the text image is not good enough. There are many reasons accounting for it: i) the light level is hard to control, and it is sometimes too high or too low; ii) the text image is usually deflection. In the image preprocessing stage, the intensity adjustment and skewness correction are performed to solve above problems.

For the intensity adjustment of input text image, the nonparametric illumination correction for document images via convex hulls [7] is adopted here. The method comes from two observations: that the shading surface of most text pages is quasi-concave and that the document contents are usually printed on a sheet of plain and bright paper. Based on these observations, a shading image can be accurately extracted via convex hulls-based image reconstruction. The method proves to be surprisingly effective for image shading correction and dark borders removal. It can restore a desired shading-free image and meanwhile yield an illumination surface of high quality. More importantly, the method is nonparametric and thus does not involve any user interactions or parameter fine-tuning. This would make it very appealing to our application. Fig. 3 shows the results of shading correction for a real-captured document page.

For the skewness correction of input text image, a skew estimation method of document images using bagging [8] is adopted to adjust the orientation of the text image. Rather than to derive a skew angle merely from text lines, the method exploits various types of visual cues of image skew available in local image regions. The visual cues are extracted by Radon transform and then outliers of them are iteratively rejected through a floating cascade. A bagging (bootstrap aggregating) estimator is finally employed to combine the estimations on the local image blocks. Note that the method obtains the skewness angle *ang* under the assumption that the document is placed vertically. However, in our application the document is placed horizontally under the camera, as shown in Fig. 2. Thus, the final skewness angle $ang2$ can ultimately be expressed by

$$ang2 = \begin{cases} -ang & \text{if } ang <= 90^\circ \\ 180 - ang & \text{if } ang > 90^\circ \text{ and } ang <= 180^\circ \end{cases} \quad (1)$$

According to the value of $ang2$, the document skewness can be corrected. An illustrative example is shown in Fig. 3. As can be seen in Fig. 3(b), the convex hulls based method [7] yields quite desirable outputs for shading extraction and correction. The method can correctly remove dark border noises around the captured documents and the extracted shading images are also of high quality, which could be further used as an input to skewness correction. As can be seen in Fig. 3(c), once the final skewness angle is obtained, the bagging based method [8] can provide an accurate skew estimation and correction result.



Youth

People are always talking about 'the problem of youth'. If there is one, which I take leave to doubt, then it is older people who create it, not the young themselves. Let us get down to fundamentals and agree that the young are after all human beings, people just like their elders. There is only one difference between an old man and a young one: the young man has a glorious future before him and the old one has a splendid future behind him: and maybe that is where the rub is.

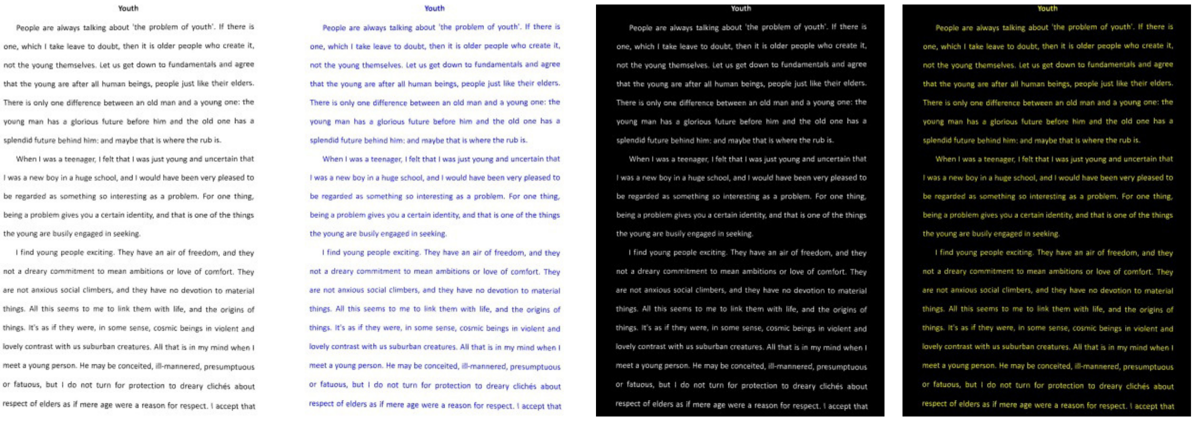
When I was a teenager, I felt that I was just young and uncertain that I was a new boy in a huge school, and I would have been very pleased to be regarded as something so interesting as a problem. For one thing, being a problem gives you a certain identity, and that is one of the things the young are busily engaged in seeking.

I find young people exciting. They have an air of freedom, and they not a dreary commitment to mean ambitions or love of comfort. They are not anxious social climbers, and they have no devotion to material things. All this seems to me to link them with life, and the origins of things. It's as if they were, in some sense, cosmic beings in violent and lovely contrast with us suburban creatures. All that is in my mind when I meet a young person. He may be conceited, ill-mannered, presumptuous or fatuous, but I do not turn for protection to dreary clichés about respect of elders as if mere age were a reason for respect. I accept that

(a) (b) (c)

Figure 3. Example of shading correction and skewness correction for a real-captured document page. (a) captured document page. (b) shading correction result. (c) skewness correction result.

4.2.3. *Functions for low-vision users.* In our system, we developed two useful functions for low-vision users: foreground/ background (F/B) color transform and magnifying glass effect. For F/B color transform, the main purpose of this operation is to make the text characters more obvious for low-vision users. Since the text page is relatively clear and suitable for image binarization after image preprocessing, it is not very hard to achieve this goal. Fig. 4 displays some results of text color transform.



Youth

People are always talking about 'the problem of youth'. If there is one, which I take leave to doubt, then it is older people who create it, not the young themselves. Let us get down to fundamentals and agree that the young are after all human beings, people just like their elders. There is only one difference between an old man and a young one: the young man has a glorious future before him and the old one has a splendid future behind him: and maybe that is where the rub is.

When I was a teenager, I felt that I was just young and uncertain that I was a new boy in a huge school, and I would have been very pleased to be regarded as something so interesting as a problem. For one thing, being a problem gives you a certain identity, and that is one of the things the young are busily engaged in seeking.

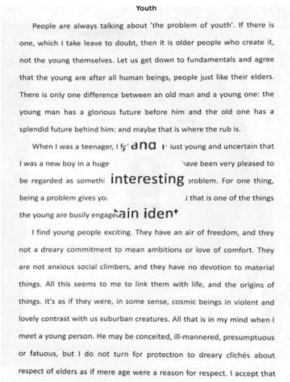
I find young people exciting. They have an air of freedom, and they not a dreary commitment to mean ambitions or love of comfort. They are not anxious social climbers, and they have no devotion to material things. All this seems to me to link them with life, and the origins of things. It's as if they were, in some sense, cosmic beings in violent and lovely contrast with us suburban creatures. All that is in my mind when I meet a young person. He may be conceited, ill-mannered, presumptuous or fatuous, but I do not turn for protection to dreary clichés about respect of elders as if mere age were a reason for respect. I accept that

Figure 4. Results of text color transform

To better help low-vision user read text, another useful function is magnifying glass effect. The virtual visual effect is created by defining the shape of the magnifier as a smoothed circular window and updating the position of the magnifying glass 20 times per second. Fig. 5 shows some examples of magnifying glass effect. One can clearly see that the words or characters become much larger than before, so the low-vision users can read the text more easily.

4.2.4. *Functions for blind users.* The most important function in our system is the text reading for blind users. The steps to realize this function consist in: text line segmentation, OCR for text line, and TTS output.

For the first step text line segmentation, we segment the whole text according to the line of the printed text. Specifically, suppose $g(i, j)$ denotes the whole text image, which

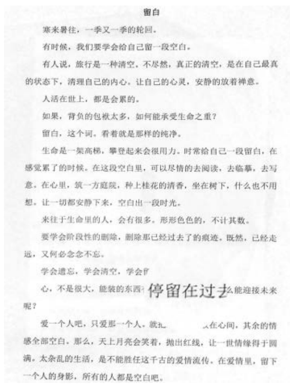


Youth

People are always talking about 'the problem of youth'. If there is one, which I take leave to doubt, then it is older people who create it, not the young themselves. Let us get down to fundamentals and agree that the young are after all human beings, people just like their elders. There is only one difference between an old man and a young one: the young man has a glorious future before him and the old one has a splendid future behind him: and maybe that is where the rub is.

When I was a teenager, I f and just young and uncertain that I was a new boy in a huge have been very pleased to be regarded as somethi interesting problem. For one thing, being a problem gives yo that is one of the things the young are busily engage ain iden

I find young people exciting. They have an air of freedom, and they not a dreary commitment to mean ambitions or love of comfort. They are not anxious social climbers, and they have no devotion to material things. All this seems to me to link them with life, and the origins of things. It's as if they were, in some sense, cosmic beings in violent and lovely contrast with us suburban creatures. All that is in my mind when I meet a young person. He may be conceited, ill-mannered, presumptuous or fatuous, but I do not turn for protection to dreary clichés about respect of elders as if mere age were a reason for respect. I accept that



留白

寒来暑往，一季又一季的轮回。

有时候，我们要学会给自己留一段空白。

有人说，旅行是一种清空。不尽然，真正的清空，是在自己最真的状态下，清理自己的内心。让自己的心灵，安静的放着禅意。

人活在世上，都是会累的。

如果，背负的包袱太多，如何能承受生命之重？

留白，这个词。看着就是那样的纯净。

生命是一架高梯，攀登起来会很用力。时常给自己一段留白，在感觉累了的时候。在这段空白里，可以尽情的去阅读，去临摹，去写意。在心里，筑一方庭院，种上桂花的清香，坐在树下，什么也不用想。让一切都安静下来，空白出一段时光。

来往于生命里的人，会有很多。形形色色的，不计其数。

要学会阶段性的删除，删除那已经过去了的痕迹。既然，已经走远，又何必念念不忘。

学会遗忘，学会清空，学会

心，不是很大，能装的东西 停留在过去 么能迎接未来呢？

爱一个人吧，只爱那一个人。就 在心间，其余的情感全部空白。那么，天上月亮会笑着，抛出红线，让一世情缘得于圆满。太杂乱的生活，是不能胜任这千古的爱情流传。在爱情里，留下一个人的身影，所有的人都是空白吧。

FIGURE 5. Examples of magnifying glass effect

defines all the pixels of the image at the position $(i,\ j)$, and L denotes the height of the image. Thus, the horizontal integral projection $F(i)$ of the ith row in the text image can be written as

$$F(i) = \sum_{i=1}^{L} g(i,j) \tag{2}$$

Then, we calculate the average value of the projection $F(i)$. Suppose N denotes the number of lines in the text image, the average value *Average* can thus be defined as

$$Average = \sum_{i=1}^{N} F(i)/N \tag{3}$$

Once the average value is obtained by using Eq. (3), some constraints can be given to find the region of each line. Two conditions that used for the text line segmentation are defined as follows:

- **First condition:** If there is a continuous n lines before the ith line that satisfying

$$\left(F(i) > \frac{Average}{p}\right) \cap \left(F(i+1) > \frac{Average}{p}\right) \cap \cdots \cap \left(F(i+n-1) > \frac{Average}{p}\right) \tag{4}$$

and the ith line satisfies:

$$F(i) > r \tag{5}$$

Then, the ith line is the upper edge of the text.

- **Second condition:** If there is a continuous m lines before the ith line that satisfying

$$(F(i) < r) \cap (F(i+1) < r) \cap \cdots \cap (F(i+m-1) < r) \tag{6}$$

and the ith line satisfies:

$$F(i) > \frac{Average}{p} \tag{7}$$

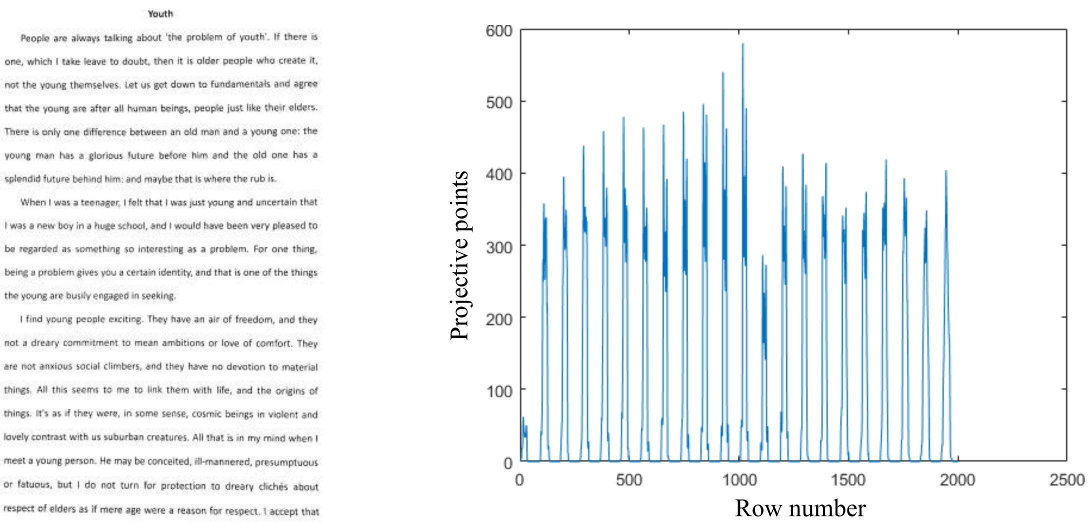
Then, the ith line is the lower edge of the text. The algorithm of text line segmentation is presented in Tab. 2.

TABLE 2. The algorithm of text line segmentation

Algorithm 2: Text line segmentation
Step 1: scanning the text image pixels from bottom to top, and calculating $F(i)$;
Step 2: checking if $F(i)$ satisfies the first condition or the second condition;
Step 3: continue the second step until finish all the text image.

In our experiment, we find that when $n = 5$, $p = 1$, $m = 3$ and $r = 0.5$, the text line segmentation algorithm can achieve a relatively accurate segmentation result. An illustrative example for the text line segmentation is shown in Fig. 6. As can be seen in Fig. 6(b), the line projection result has 22 peaks, which means the printed text shown in Fig. 6(a) has 22 lines (including text title). This conclusion is confirmed with our observations on Fig. 6(a). Tests on other printed text images also verify the conclusion.

Once each line of the text image is detected, word extraction can be performed by using the Tesseract engine. The OCR engine is instructed to extract each word from the detected text line, and it returns: the word string of each text line. Then, the string is synthesized to audio speech by using the Flite Text-to-Speech (TTS) [1]. The TTS engine is developed at CMU and designed as an alternative synthesis engine for voices built using the FestVox suite of voice building tools. By combining the OCR engine with the TTS engine, the printed text can be read once and for all or line by line. Fig. 7 displays the user interface of our software in the midst of reading for the whole text reading and single line reading.


Youth

People are always talking about 'the problem of youth'. If there is one, which I take leave to doubt, then it is older people who create it, not the young themselves. Let us get down to fundamentals and agree that the young are after all human beings, people just like their elders. There is only one difference between an old man and a young one: the young man has a glorious future before him and the old one has a splendid future behind him: and maybe that is where the rub is.

When I was a teenager, I felt that I was just young and uncertain that I was a new boy in a huge school, and I would have been very pleased to be regarded as something so interesting as a problem. For one thing, being a problem gives you a certain identity, and that is one of the things the young are busily engaged in seeking.

I find young people exciting. They have an air of freedom, and they not a dreary commitment to mean ambitions or love of comfort. They are not anxious social climbers, and they have no devotion to material things. All this seems to me to link them with life, and the origins of things. It's as if they were, in some sense, cosmic beings in violent and lovely contrast with us suburban creatures. All that is in my mind when I meet a young person. He may be conceited, ill-mannered, presumptuous or fatuous, but I do not turn for protection to dreary clichés about respect of elders as if mere age were a reason for respect. I accept that

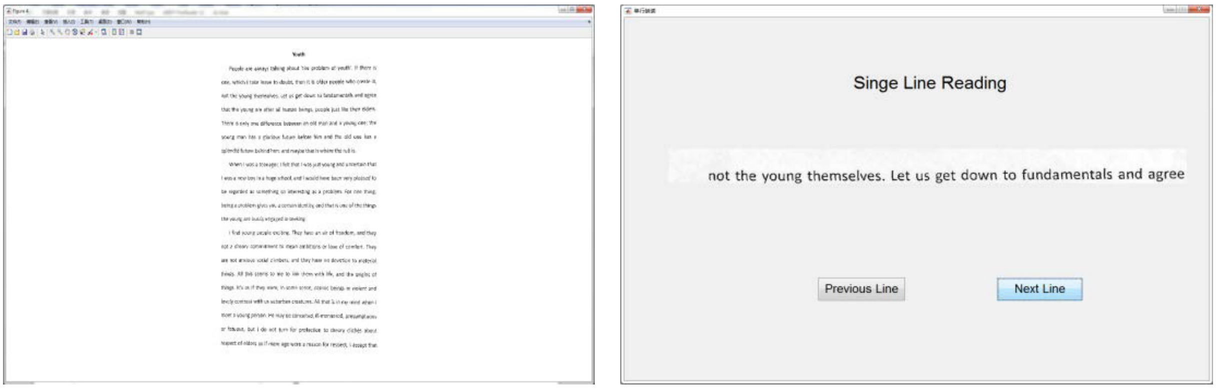
(a) (b)

FIGURE 6. Segmentation result of text line. (a) printed text image. (b) line projection result

5. **Evaluation.** In this section, we will evaluate our system from two aspects: text recognition accuracy and user feedback. The former determines the effectiveness of our system, and the latter determines the practicability of our system.

5.1. **Text recognition accuracy.** The accuracy of text recognition will greatly affect the effectiveness of our system. However the accuracy will be influenced by various factors, such as document page layout, different languages, etc. Therefore, to test the robustness and effectiveness of our system, different page layouts and languages are used: (i) A4 documents, English, 12 point font, single/double spacing, (ii) A4 documents, English, 14 point font, single/double spacing, (iii) English/Chinese documents. The statistic results are showed in Tab. 3.

From the table, we can see that the system achieves highest recognition rate when the page layout of the English document is double-spaced in 14 point font. If the font is 12 point or the row spacing is single, the recognition rate is decreased. Besides, compared with Chinese recognition, the system achieved much higher correct rate for English text



(a) (b)

FIGURE 7. Our software in the midst of reading. (a) whole text reading. (b) single line reading

recognition. Thant's because the Tesseract engine that used in our system has superior performance on English, but performs poor on Chinese document. Therefore, we can deduce that the system requires the English text should be double-spaced in at least 14 point font.

TABLE 3. Text recognition accuracy for different page layouts and languages

Page layout (English document)	**Recognition rate**
12 point font, single spacing	78%
12 point font, double spacing	85%
14 point font, single spacing	90%
14 point font, double spacing	98%
Language	**Recognition rate**
English (14 point font, double spacing)	98%
Chinese (14 point font, double spacing)	74%

5.2. **User feedback.** User feedback is also very important for evaluating our system. However, it can't be ignored that user feedback is a subjective experience, thus it may be biased and will influence our evaluation. To avoid this bias, a questionnaire (see Tab. 4) is used to objectively measure user feedback situations. The evaluation was performed by 18 subjects. The subjects were chosen from among voluntary undergraduate healthy students and visually impaired patients with low-vision or no-vision. Tab. 4 shows the evaluation results. 5-point likert scale is adopted to assess the user feedback. For the likert scale, 1 stands for strongly disagree, and 5 stands for strongly agree.

As can be seen in Tab. 4, the healthy users generally liked the current implementation, and the low-vision users found it generally easy and enjoyable to access text with the BlindReader. While the overall experience with the BlindReader was rated as mediocre by the blind users. They commented that this was mainly because the menu is hard to operate and the synthesized voice sounds unpleasant. Nevertheless, all the participants reported that the BlindReader could help them to fulfill everyday study and collect more information for themselves.

Table 4. Results from the questionnaire on 5-point Likert scale

group	**Healthy users**	**Low-vision users**	**Blind Users**
General			
The overall experience was enjoyable	4	4	3
Accessing Text with BlindReader was easy	5	4	3
Reading with the BlindReader was enjoyable	4	5	4
Reading with the BlindReader was easy	5	4	3
Comparison to other text reading aids			
Accessing text with the BlindReader felt easier	5	4	3
Reading with the BlindReader felt easier	4	4	3
Independence			
Feel greater desire to become able to read independently	4	4	4
Feel the desire to use the BlindReader to access text	4	3	2

6. **Conclusion.** In this paper, we introduced BlindReader, a visual assistant system. This system can assist visually impaired people to read printed text. Some computer vision algorithms are introduced to adjust the intensity and orientation of captured document images in preprocessing stage. Using the system, the low-vision users can read the text more clearly just like the way people use the magnifying glass and the blind users can hear the recognized words synthesized to audio speech for the whole text or the single text line. The main novelty of our work are: i) an assistant system is conceptualized and implemented to support text reading for the visually impaired people; ii) each line of the text image is segmented through layout analysis and then these text line images are recognized by using the OCR engine; and iii) we report findings from two evaluations to determine the effectiveness and usability of the BlindReader. Evaluation results show that an improvement in visually impaired people's study interest can be achieved by using our system.

However, our system also has some limitations: i) the system requires the text should be double-spaced in at least 14 point font, which is hard to satisfy in many real applications; and ii) the text recognition may have low accuracy when deal with other languages rather than English. In the future, we will try to investigate the layout analysis and OCR techniques that can handle complicated text styles and several languages, and we hope the BlindReader could help researchers to design more and more reading aid systems to bring more convenience to visual disabled people. We also look to go beyond usage for person with a visual impairment, and the BlindReader may be useful to support early language learning for preschool children and translating different foreign languages for them.

Acknowledgements. This work was supported in part by the National Undergraduate Programs for Innovation, National Natural Science Foundation of China (No. 61502537), Hunan Provincial Natural Science Foundation of China (2018JJ3681), Postdoctoral Science Foundation of Central South University (No. 126648), Changsha science and technology projects (kq1606004), and the Key Laboratory of Hunan Province for New Retail Virtual Reality Technology (2017TP1026).

REFERENCES

[1] A W. Black and K. Q. Lenzo, Flite: a small fast run-time synthesis engine. In *Proceedings of Isca Tutorial & Research Workshop on Speech Synthesis Perthshire*, vol. 99, no. 4, pp. 20–24, 2001.

[2] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White and T. V. Yeh, Nearly real-time answers to visual questions. In *Proceedings of International Cross Disciplinary Conference on Web Accessibility*, pp. 333–342, 2010.

[3] J. P. Peters, C. Thillou, S. Ferreira, Embedded reading device for blind people: a user-centered design. In *Proceedings of IEEE Applied Imagery Pattern Recognition Workshop*, pp.217–222, 2004.

[4] R. Shilkrot, J. Huber, M. E. Wong, P. Maes, S. Nanayakkara, FingerReader: A wearable device to explore printed text on the go. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp.1–10, 2005.

[5] H. Wu, B. J. Zou, Y. Q. Zhao, J. J. Guo, Scene text detection using adaptive color reduction, adjacent character model and hybrid verification strategy. *Visual Computer*, vol.33, no.1, pp.113–126, 2015.

[6] R. Smith, An overview of the tesseract OCR engine. In *Proceedings of International Conference on Document Analysis & Recognition.* no.2, pp.629–633, 2007.

[7] G. F. Meng, S. M. Xiang, N. N. Zheng, C. H. Pan, Nonparametric illumination correction for scanned document images via convex hulls. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, no.7, pp.1730–1743, 2013.

[8] G. F. Meng, C. H. Pan, N. N. Zheng, C. Sun, Skew estimation of document images using Bagging. *IEEE Transactions on Image Processing*, vol.19, no.7, pp.1837–1846, 2010.