Jurnal Teknologi

Full paper

Mobile Text Reader for People with Low Vision

Teng Ren Sin[a], Eileen Su Lee Ming[a*], Yeong Che Fai[b], Ong Jian Fu[a], Sim Yang Shane[a]

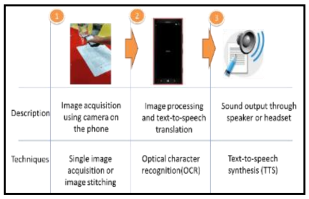
[a]*Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*
[b]*Center of Artificial Intelligence and Robotics, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

*Corresponding author: eileensu@utm.my

Article history

Received : 25 March 2015
Received in revised form :
11 April 2015
Accepted : 13 April 2015

Graphical abstract



Description	Image acquisition using camera on the phone	Image processing and text-to-speech translation	Sound output through speaker or headset
Techniques	Single image acquisition or image stitching	Optical character recognition(OCR)	Text-to-speech synthesis (TTS)

Abstract

People with low vision have visual acuity less than 6/18 and at least 3/60 in the better eye, with correction. The limited vision requires them to enhance their reading ability using magnifying glass or electronic screen magnifier. However, people with severe low vision have difficulty and suffer fatigue from using such assistive tool. This paper presents the development of a mobile text reader dedicated for people with low vision. The mobile text reader is developed as a mobile application that allows user to capture an image of texts and then translate the texts into audio format. One main contribution of this work compared to typical optical character recognition (OCR) engines or text-to-speech engines is the addition of image stitching feature. The image stitching feature can produce one single image from multiple poorly aligned images, and is integrated into the process of image acquisition. Either single or composite image is subsequently uploaded to a cloud-based OCR engine for robust character recognition. Eventually, a text-to-speech (TTS) synthesizer reproduces the word recognized in a natural-sounding speech. The whole series of computation is implemented as a mobile application to be run from a smartphone, allowing the visual impaired to access text information independently.

Keywords: Text reader; image stitching; optical character recognition; mobile application; text-to-speech; vision impairment

Abstrak

Golongan yang mempunyai penglihatan kabur mempunyai daya penglihatan kurang daripada 6/18 dan sekurang-kurangnya 3/60 dalam mata yang lebih baik, dengan pembetulan. Visi terhad memaksa mereka untuk meningkatkan kemampuan membaca mereka dengan menggunakan kanta pembesar atau pembesar skrin elektronik. Walau bagaimanapun, golongan berpenglihatan kabur mempunyai kesukaran dan mengalami keletihan dari menggunakan alat bantuan ini. Karya ini membentangkan pembangunan pembaca teks mudah alih khusus untuk orang berpenglihatan kabur. Pembaca teks mudah alih dibangunkan sebagai aplikasi mudah alih yang membolehkan pengguna merakam imej teks dan kemudian menterjemahkan teks ke dalam format audio. Salah satu sumbangan utama karya ini berbanding dengan enjin pengenalan aksara optik (OCR) lain atau enjin teks-ke-ucapan lain adalah penambahan ciri jahitan imej. Jahitan imej boleh menghasilkan satu imej tunggal dari beberapa imej yang tidak tersusun dan disepadukan dalam proses pemerolehan imej. Gambar tunggal atau komposit kemudiannya dimuat naik ke enjin OCR berasaskan awan untuk pengecaman aksara yang teguh. Akhirnya, pensintesis teks-ke-suara (TTS) menghasilkan semula perkataan yang dikenalpasti sebagai ucapan yang berbunyi semulajadi. Keseluruhan siri pengkomputeran ini dilaksanakan sebagai aplikasi mudah alih yang akan digunakan dari telefon pintar, membolehkan golongan terjejas penglihatan untuk mendapatkan maklumat teks secara berdikari.

Kata kunci: Pembaca teks; jahitan imej; pengenalan aksara optik; aplikasi mudah alih; teks-ke-suara; penglihatan terjejas

■1.0 INTRODUCTION

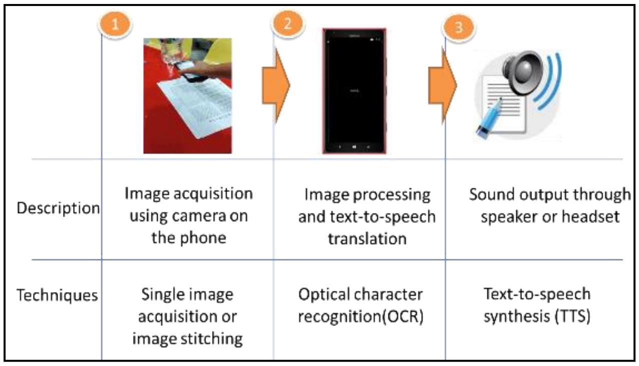
The range of visual impairment covers both total blindness and low vision. As defined by World Health Organization, low vision is visual acuity less than 6/18 and equal to or better than 3/60 in the better eye with the best correction like treatment or standard refractive correction [1]. Low vision is majorly caused by refractive error and secondly by cataract [2], an ocular pathology with aging as one of the reason. Hence, it is not wrong to state that the risk of getting low vision increases with age [3]. People with low vision are still able to sense some visual input. The limited vision allows them to enhance their reading ability by using magnifying glass or electronic screen magnifier. However, some people with severe low vision might still face difficulty and fatigue from using such assistive tool. Unlike total blindness, people with low vision typically do not go through Braille education and hence they cannot read using the sense of touch. One possible solution for them is a text reader, which produces speech that represents the text element.

Out of 285 million people who have vision impairment around the world, 246 million people have low vision, which makes up 86.3% of the visual impaired population [4]. In Malaysia, out of 518,000 Malaysian, 464,000 people have low vision, which makes up 89.6% of the visual impaired population in Malaysia [2]. A text reader can effectively help this huge population to reinstate their independence in readings [5]. Some text readers are either too bulky, too costly, inaccessible or non-portable. One example is the text reader as proposed by Tjajha et. al [6], which was developed on a non-portable laptop. Another version is a screen reader which focused only on reading the text element on the screen [7]. There are text readers which are handheld pointers which require the user to point at the word to read it out [8]. Our study targets a different outcome, which is to help the low vision user to access text information from hardcopy document in a low-cost and intuitive manner, as outlined in a design guidelines document [9]. To be efficient, our device aims to fulfill some binding specifications like simple interface, ergonomics, posing tolerances, portable, and finally and at a low cost. Hence, the simplest idea would be to capture a photo with the whole text and then, output the text element as audio. The whole process might only include one or two buttons to ensure intuitive usage by the user

Smartphone is becoming more and more popular and trendy as an assistive device for the visual impaired [10]. With the breakthrough of advanced hardware and mobile operating systems, mobile apps development is a simple and direct idea to be linked to our objective. As such, our study uses the smartphone, Windows Phone Nokia Lumia 920 which comprised of some hardware crucial for our device, such as big screen (768 × 1280 pixels), high memory (32 GB storage, 1 GB RAM), high resolution camera (8 MP, 3264 × 2448 pixels), and a robust CPU (Qualcomm MSM8960 Snapdragon, Dual-core 1.5 GHz Krait). The big screen is essential to address the accessibility of visual impaired user. The high memory and robust CPU ensure a smoother and faster processing. The high resolution camera assures high accuracy of image acquisition which eventually enhances the speech output. In terms of the development tool, we use Visual Studio 2012 with Windows Phone 8 SDK which minimally requires a laptop or computer with Windows 8 Pro 64-bit, 4GB RAM, and Hyper-V for smooth operation of emulator. The application, namely SightShare Reader, is developed using C# language and XAML language. The above-mentioned technology is believed to be sufficient to overcome the limitation of such application in the past [11].

■2.0 METHODOLOGY

Figure 1 illustrates the scheme of the proposed system. At the photo capturing page, there is a button which indicates the accomplishment of the image acquisition task. If only one picture is taken, the event of pressing the button will directly activate the optical character recognition (OCR) processing on the picture. However, if more than one picture is taken, the event of pressing the button will activate the process of image stitching. Similar to the single photo, the composite photo is uploaded to the cloud-based OCR engine for text extraction. Once the extracted text is downloaded as string element, speech is started immediately.



	1	2	3
Description	Image acquisition using camera on the phone	Image processing and text-to-speech translation	Sound output through speaker or headset
Techniques	Single image acquisition or image stitching	Optical character recognition(OCR)	Text-to-speech synthesis (TTS)

Figure 1 System configuration of the proposed device which consists of the three fundamental blocks: image acquisition, OCR and TTS

2.1 Image Stitching

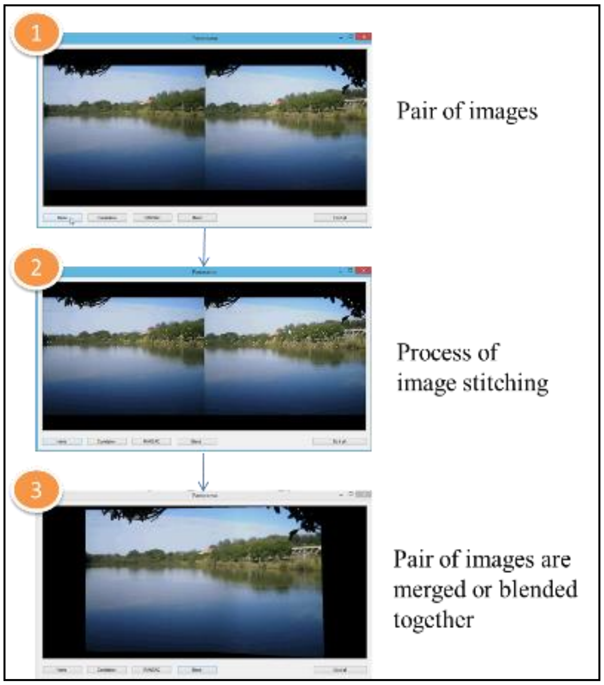

1
Pair of images
2
Process of image stitching
3
Pair of images are merged or blended together

Figure 2 Capability of image stitching

Panoramic image stitching is the process to produce a high resolution mosaic image from a series of smaller, partially overlapping images [12]. Figure 2 illustrates the capability of image stitching. This function is prevalently used to take pictures of natural panorama. Image stitching technique can be used to compose a map with sequence of camera frames taken from a unmanned aerial vehicle (UAV) [13]. Meanwhile, in this paper, the technique is used for document stitching. When using a phone camera to capture a large document, e.g. an A1-size paper filled with small-font texts, the texts in that captured image might be too blurry if the whole A1-size document is forcefully fitted into one single image. To perform any character recognition using that camera image will likely result in failure due to the blurry texts. Image stitching feature allows user to capture the A1 document as multiple segments, with each segment having sharp focus of the texts. The individual segments can then be stitched together to form a single A1 document, resulting in crisp and clear texts for successful optical character recognition.

In this study, Accord.NET Framework, an extension framework for Aforge.NET, also a popular framework for image processing, computer vision and artificial intelligence, is used to perform automatic image stitching. The algorithm will join one pair of images at one time. The general idea consists of identifying common points between the two overlapped images and then studying one of the images on top of the other in an effort to match those points. Figure 3 shows the flowchart of stitching a pair of images until mosaic image is produced.

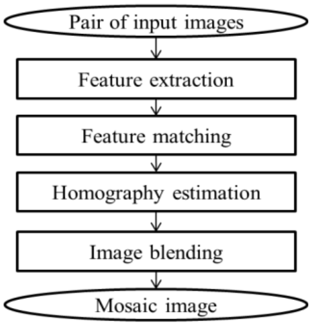


Figure 3 Flowchart of automatic image stitching using Accord.NET framework

1) *Feature Extraction*

The first step of the image stitching process was to align the images properly. Image alignment algorithm made use of interest points or features to identify corresponding relationships between partially overlapped images[14]. In this study, corners were chosen to be the features to track and Harris corner detection algorithm was used to detect these corners in an image. Corners were defined as intersection of two edges and were good features because they were distinctive [15]. Several corner detection algorithms were reviewed and Harris corner detector was chosen because of its above average performance and simplicity to implement [16, 17]. The Harris corner detector formula is as stated in Equation 1 and corners are denoted by window with large variation of intensities [18]:

$$E(u,v) = \sum_{x,y} w(x,y)[I(x + u, y + v) - I(x,y)]^{2} \qquad (1)$$

where

w(x,y) is window at position (x,y)
I(x,y) is the intensity at position (x,y)
I(x + u, y + v) is the intensity of the moved window at (x + u, y + v)

2) *Feature Matching*

After the interest points (corners) were identified, they had to be correlated. By using a maximum correlation rule, a window of pixels around every point in the first image was analyzed and correlated to a window of pixels around every other point in the second image. Points with maximum bidirectional correlation were considered as corresponding pairs.

3) *Homography Estimation*

After identifying two sets of correlated points, an image transformation model which can translate points from one set to another must be defined. The model had to study one of the pair of images on top of the other while matching most of the correlated points. Hence, a studyive transformation, or homography matrix, was created using random sample consensus (RANSAC) algorithm [19]. The homography matrix was represented as a 3x3 matrix and the last value in the matrix was interpreted as a scale parameter and was fixed at 1. At this stage, incorrect correlations which may cause troubles in blending phase were removed [13].

$$\begin{bmatrix} h_{11} & h_{21} & h_{31} \\ h_{12} & h_{22} & h_{32} \\ h_{13} & h_{23} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} \qquad (2)$$

where

H is the 3×3 homography matrix
x and y denotes the position of pixel in first image
$x'$and y' denotes the position of matched pixel in second image

4) *Image Blending*

Image blending was the process to remove seam or edges and make a compact image [20]. Translation and rotation of image were performed in this phase to blend the pair of images together according to the homography matrix [13].

The above-mentioned series of image stitching operation was considered as one iteration. The iteration was repeated for every photo captured by the user.

2.2 Optical Character Recognition

Optical character recognition (OCR) was the machine replication of human reading [21]. The OCR algorithm can extract text elements (feature) from an image using image processing, computer vision, and artificial intelligence. Based on performance evaluation of existing OCR engines [22], ABBYY Cloud OCR SDK was selected as the tool for this stage. The commercial engine can recognize up to 194 languages including English, Malay, Chinese, Japanese, and Korean and so on. Except from performing character recognition, the engine also performed pre-processing features like deskewing, automatic page orientation detection, perspective correction, texture removal and resolution correction, which drastically enhanced the accuracy of the result. As it works on cloud, the Internet connection was needed to upload the targeted image and to download the recognized text in string type.

2.3 Text-to-speech Synthesis

After downloading the string typed text element from ABBYY cloud-based OCR engine, the text was synthesized to natural-sounding speech using a robust Windows' API called Windows.Phone.Speech.Synthesis. The API can support multiple languages including English, Chinese, Japanese and Korean. However, language packs for languages, excluding English, need to be preinstalled in the smartphone.

■3.0 RESULTS AND DISCUSSION

Figure 4 illustrates the overall pages and navigation of SightShare Reader application. The first page after activating the application is the main page. Three options are provided here: adjust the setting, browse picture from phone memory, and capture photo from hardcopy documents. The setting page allows the user to choose language for the text-to-speech API as well as certain specifications for the voice output and camera. At the camera page, the user can capture up to maximum 5 photos which, with the pressing of the button “Finish taking pictures”, will be automatically stitched into one image and sent to the processing page. If only one photo is captured, image stitching will not be performed. The single image will be sent to the processing page automatically. The same thing happens for browsing pictures from the picture gallery. However, only one picture can be chosen by the user in a time. Image stitching does not apply here. At the final processing page, the final image is uploaded to cloud-based OCR engine, text element downloaded and directly spoken out. Hence, the page will also display the downloaded text element.

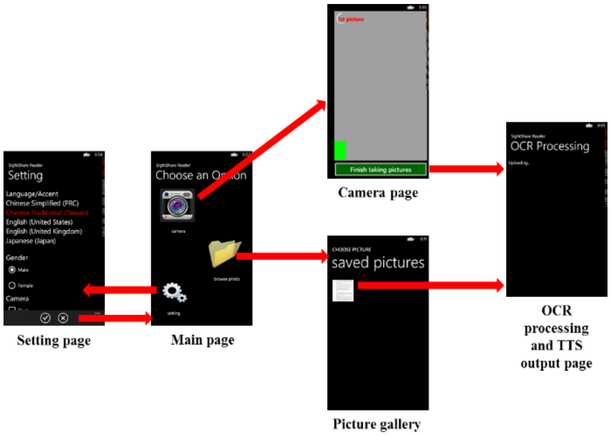


Figure 4 Pages and navigation of SightShare Reader application

3.1 Image Stitching

A simple experiment is conducted to identify the relationship between sizes of input images with the time needed for one complete iteration of image stitching. The result is shown in Table 1. One of the successful result of this experiment is shown in Figure. 5. The overlapping region covers about 30% of the composite image.

The second experiment is conducted to investigate the relationship between the numbers of input images with the required time to completely stitch all the input images.

Table 1 Time taken to stitch two images for different size of input images

Size of input images (pixel)	Time taken to stitch two images (second)
640 × 480	21
844 × 475	37
1024 × 576	66

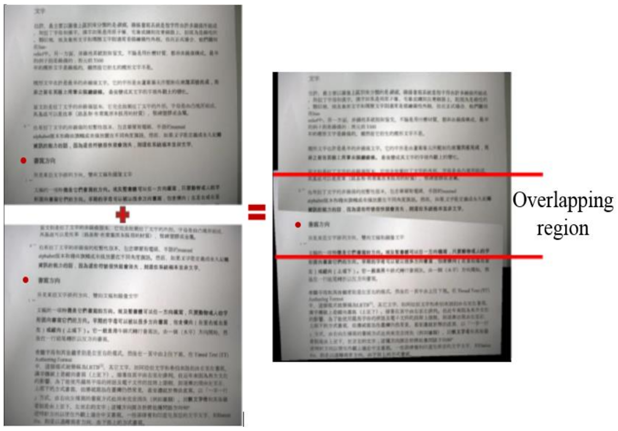


Figure 5 The successful result of image stitching on text-based images of 1024×576 pixels in the first experiment

However, unlike first experiment, the sample image in second experiment, as shown in Figure 6, is non-text image of 844×475 pixels with big objects of different colours and more importantly, a wider overlapping region which covers about 50% of the resulted image. Most of the input images in second experiment has wide overlapping region. The time taken is recorded in Table 2.

Table 2 Time taken for two or more iteration of image stitching

Number of input images	Time taken for image stitching (second)
2	12
3	20
4	23
5	31

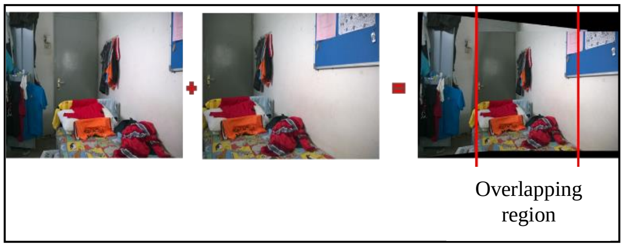


Figure 6 One of the successful result for image stitching of 2 input images of 844×475 pixels in the second experiment

From both experiment, some useful inference can be made:

1) The wider the overlapping region, the faster the processing time. This is prominent in second experiment as the time taken for image stitching is not linearly proportionate to the number of input images.

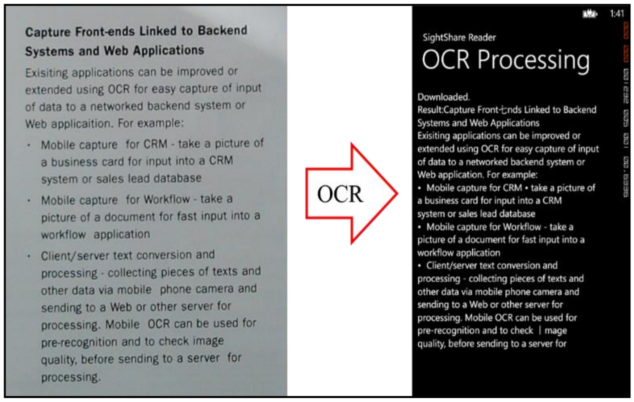
2) The bigger the size of input images, the slower the processing time, but there is less information loss compared to image with smaller size.

3) Failure of image stitching mostly occurs when the camera is not hovered in a direction which is parallel to the surface of document. In other words, either the angle of camera or the distance of camera from the document is not uniform throughout the process of image acquisition. Failure also occurs when more than 5 pictures are taken as input images.

4) The prolonged time of processing in the first experiment might be caused by the non-uniform lighting as there is non-uniform shadow on the input images. However, this statement is just a speculation. More tests are needed.

3.2 Optical Character Recognition

The results of OCR engine in two different languages are shown in Figure 7 and Figure 8. High accuracy of the OCR output is confirmed from the results.



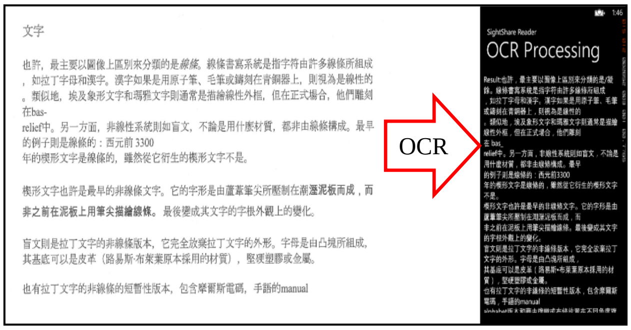
Capture Front-ends Linked to Backend Systems and Web Applications

Exisiting applications can be improved or extended using OCR for easy capture of input of data to a networked backend system or Web applicaition. For example:

- Mobile capture for CRM - take a picture of a business card for input into a CRM system or sales lead database
- Mobile capture for Workflow - take a picture of a document for fast input into a workflow application
- Client/server text conversion and processing - collecting pieces of texts and other data via mobile phone camera and sending to a Web or other server for processing. Mobile OCR can be used for pre-recognition and to check image quality, before sending to a server for processing.

Downloaded.
Result:Capture Front七nds Linked to Backend Systems and Web Applications
Exisiting applications can be improved or extended using OCR for easy capture of input of data to a networked backend system or Web application. For example:
• Mobile capture for CRM • take a picture of a business card for input into a CRM system or sales lead database
• Mobile capture for Workflow - take a picture of a document for fast input into a workflow application
• Client/server text conversion and processing - collecting pieces of texts and other data via mobile phone camera and sending to a Web or other server for processing. Mobile OCR can be used for pre-recognition and to check I mage quality, before sending to a server for

Figure 7 Result of OCR in English



文字

也許，最主要以圖像上區別來分類的是線條。線條書寫系統是指字符由許多線條所組成，如拉丁字母和漢字。漢字如果是用原子筆、毛筆或鑄刻在青銅器上，則視為是線性的。類似地，埃及象形文字和瑪雅文字則通常是描繪線性外框，但在正式場合，他們雕刻在bas-relief中。另一方面，非線性系統則如盲文，不論是用什麼材質，都非由線條構成。最早的例子則是線條的：西元前 3300 年的楔形文字是線條的，雖然從它衍生的楔形文字不是。

楔形文字也許是最早的非線條文字。它的字形是由蘆葦筆尖所壓制在潮溼泥板而成，而非之前在泥板上用筆尖描繪線條。最後變成其文字的字根外觀上的變化。

盲文則是拉丁文字的非線條版本，它完全放棄拉丁文字的外形。字母是由凸塊所組成，其基底可以是皮革（路易斯·布萊葉原本採用的材質），堅硬塑膠或金屬。

也有拉丁文字的非線條的短暫性版本，包含摩爾斯電碼，手語的manual

Result:也許，最主要以圖像上區別來分類的是/線條。線條書寫系統是指字符由許多線條所組成
，如拉丁字母和漢字。漢字如果是用原子筆、毛筆或鑄刻在青銅器上，則視為是線性的
。類似地，埃及象形文字和瑪雅文字則通常是描繪線性外框，但在正式場合，他們雕刻
在 bas_
relief中。另一方面，非線性系統則如盲文，不論是用什麼材質，都非由線條構成。最早
的例子則是線條的：西元前3300
年的楔形文字是線條的，雖然從它衍生的楔形文字不是。
楔形文字也許是最早的非線條文字。它的字形是由蘆葦筆尖所壓制在潮溼泥板而成，而
非之前在泥板上用筆尖描繪線條。最後變成其文字的字根外觀上的變化。
盲文則是拉丁文字的非線條版本，它完全放棄拉丁文字的外形。字母是由凸塊所組成，
其基底可以是皮革（路易斯•布萊葉原本採用的材質），堅硬塑膠或金屬。
也有拉丁文字的非線條的短暫性版本，包含摩爾斯電碼，手語的manual

Figure 8 Result of OCR in Chinese traditional

3.3 Text-to-speech Synthesis

The setting of the language affects a lot on the result. If the language is set as English, the API will skip the non-English part in the text. However, if the language is set as Chinese, it can still synthesize speech for the English text element, but in the accent of Chinese. A prominent result is shown when English text is spoken in Japanese accent. All in all, the naturalness of the speech, the most important specification in text-to-speech [23], is guaranteed.

For the phase of image stitching, more experiment need to be conducted to identify the best setting for the algorithm. Specifications which require consideration are percentage of overlapping region, non-uniform lighting, sizes of input images, camera's position and angle during image acquisition, time of processing, the requirement to implement pre-processing and what method to use if pre-processing is required.

For the stage of optical character recognition, the SDK shows the most stable and robust result compared to other process. It is insensitive to orientation, shape and size of uploaded image. It has a powerful pre-processing done on the cloud. Although time is wasted during the uploading and downloading process, the processing performed on the cloud does not use a lot of time. A fast data plan or Wi-Fi can speed up this process.

For the part of text-to-speech, the biggest limitation is the setting of the language. It is hoped that it can automatically detect the language of the text element. Secondly, a pause and play button has to be added to fulfill the design guideline of this feature [24].

For the overall application, more consideration has to be taken on the design of user experience to ensure an efficient usage by the visually impaired user. The application has to be tested on the user with low vision as last confirmation for its functionality.

■4.0 CONCLUSION

A smartphone application is developed to help the people with vision impairment to read the text element on hardcopy document, book, newspaper, or other media. The language support of the application can efficiently help the user to read in different languages. Either the high resolution camera or the additional feature like image stitching can effectively help the user during image acquisition. The robust OCR engine assures a reading experience with high accuracy. The function of text-to-speech can alleviate the fatigue of reading using magnifier. The accessibility and user experience design are focused as these are the most important specifications which can motivate the user to use the reading aid. Testing will be done to confirm the accessibility with visually impaired user.

Acknowledgement

This study is supported by the Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia Research Grant Q.J130000.2523.08H77.

References

[1] World Health Organization. 2007. Vision 2020 Global Initiative for the Elimination of Avoidable Blindness Action Plan 2006-2011, France.

[2] M. Zainal, S. M. Ismail, A. R. Ropilah, H. Elias, G. Arumugam, D. Alias, J. Fathilah, T.O Lim, L. M. Ding, P.P. Goh. 2002. Prevalence of Blindness and Low Vision in Malaysian Population: Results from the National Eye Survey. *British Journal of Ophthalmology*. 86(9): 951–956.

[3] K. Loh, J. Ogle. 2004. Age Related Visual Impairment in the Elderly. *The Medical Journal of Malaysia*. 59(4): 562.

[4] Visual Impairment And Blindness Fact Sheet N°282, 24 Dec 2013 Available from: http://www.who.int/mediacentre/factsheets/fs282/en/.

[5] T. H. Margrain. 2000. Helping Blind and Partially Sighted People to Read: the Effectiveness of Low Vision Aids. *British Journal of Ophthalmology*. 84(8): 919–921.

[6] T. V. Tjahja, A. S. Nugroho, J. Purnama, N. A. Azis, R. Maulidiyatul Hikmah, O. Riandi, B. Prasetyo. 2011. Recursive Text Segmentation For Indonesian Automated Document Reader for people with Visual Impairment. International Conference on Electrical Engineering and Informatics (ICEEI). 1–6.

[7] M. Dorigo, B. Harriehausen-Mühlbauer, I. Stengel, P. S. Dowland. 2011. Survey: Improving Document Accessibility from the Blind and Visually Impaired User's Point of View. Lecture Notes in Computer Science. 6768(2): 129–135

[8] U. Minoni, M. Bianchi, and V. Trebeschi. 2001. A Handheld Real-Time Text Reader. IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA). 354–359.

[9] M. A. Hersh. 2010. The Design and Evaluation of Assistive Technology Products and Devices Part 1: Design. In: JH Stone, M Blouin, editors. *International Encyclopedia of Rehabilitation*.

[10] E. Peng, P. Peursum and L. Li, S. Venkatesh. 2010. A Smartphone-Based Obstacle Sensor for the Visually Impaired. *Ubiquitous Intelligence and Computing*. Springer. 590–604.

[11] J. Leimer. 1962. Design Factors in the Development of an Optical Character Recognition Machine. Information Theory. *IRE Transactions*. 8(2): 167–171.

[12] C.Y. Chen and R. Klette. 1999. Image Stitching—Comparisons and New Techniques. *Computer Analysis of Images and Patterns*. Springer.

[13] V. Cani. 2011. *Image Stitching for UAV Remote Sensing Application*. Master Thesis, Universitat Politècnica de Catalunya, Spain.

[14] R. Szeliski. 2006. Image Alignment and Stitching: A Tutorial. Foundations and Trends. *Computer Graphics and Vision*. 2(1): 1–104.

[15] J. Shi and C. Tomasi. 1994. Good Features to Track. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94). 593–600.

[16] J. Chen, L.H. Zou, J. Zhang, L.H. Dou. 2009. The comparison and Application of Corner Detection Algorithms. *Journal of Multimedia*. 4(6): 435–441.

[17] J. Liu, A. Jakas, A. Al-Obaidi, Y. Liu. 2009. A Comparative Study of Different Corner Detection Methods. IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA). 509–514.

[18] R. Chandratre and V. Chakkarwar. 2014. Image Stitching using Harris Feature Detection and Random Sampling. *International Journal of Computer Applications*. 89(15): 14–19.

[19] M. Brown and D. G. Lowe. 2007. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*. 74(1): 59–73.

[20] V. Rankov, R. J. Locke, R. J. Edens, P. R. Barber, B. Vojnovic. 2005. An Algorithm for Image Stitching and Blending. *Biomedical Optics*. International Society for Optics and Photonics. 190–199.

[21] V. K. Govindan and A. P. Shivaprasad. 1990. Character Recognition—A Review. *Pattern Recognition*. 23(7): 671–683.

[22] O. Krejcar. 2012. Smart Implementation of Text Recognition (OCR) for Smart Mobile Devices. *In INTELLI, The First International Conference on Intelligent Systems and Applications*. 19: 24.

[23] M. Tatham and E. Lewis. 1996. Improving Text-to-Speech Synthesis. Proceedings of Fourth International Conference on Spoken Language. 3: 1856–1859.

[24] F. Holm, S. Pearson. 1998. User Interface Controller for Text-to-Speech Synthesizer. *US Patent* US5850629 A.