

# TEXT SIGNAGE RECOGNITION IN ANDROID MOBILE DEVICES

Oi-Mean Foong, Suziah Sulaiman and Kiing Kiu Ling

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia

Received 2013-10-15; Revised 2013-10-31; Accepted 2013-11-25

## ABSTRACT

This study presents a Text Signage Recognition (TSR) model in Android mobile devices for Visually Impaired People (VIP). Independence navigation is always a challenge to VIP for indoor navigation in unfamiliar surroundings. Assistive Technology such as Android smart devices has great potential to assist VIPs in indoor navigation using built-in speech synthesizer. In contrast to previous TSR research which was deployed in standalone personal computer system using Otsu's algorithm, we have developed an affordable Text Signage Recognition in Android Mobile Devices using Tesseract OCR engine. The proposed TSR model used the input images from the International Conference on Document Analysis and Recognition (ICDAR) 2003 dataset for system training and testing. The TSR model was tested by four volunteers who were blind-folded. The system performance of the TSR model was assessed using different metrics (i.e., Precision, Recall, F-Score and Recognition Formulas) to determine its accuracy. Experimental results show that the proposed TSR model has achieved recognition rate satisfactorily.

**Keywords:** Android Devices, Text Signage, Independence Navigation, Tesseract OCR

## 1. INTRODUCTION

According to World Health Organization (WHO), around 285 million people around the world are suffering from Visual Impairment and Blindness (WHO, 2013). In general, visual impairment is defined as the reduction of the eyesight that cannot be corrected by standard glasses or contact lens and it reduces a person's ability to perform some daily tasks. It is always a great challenge for VIP to walk along streets in town without the company of a guide dog or caretaker. With the advancement of assistive technology, the VIPs from the Western countries could perform their daily tasks like grocery shopping or even take a bus to go to various places independently. Some alternative ways had been proposed (Coughlan and Shen, 2013; Zhang *et al.*, 2013; Katz *et al.*, 2012; Sarfraz and Rizvi, 2007; Al-Saleh *et al.*, 2011) so that VIP can get around independently with the help of technologies. However, most of these systems are based on carrying heavy equipment or a computer

system and there are limited choices of text signage recognition applications that run on smart phones. Based on statistics from Infographic, approximately 46.9% of smart phone users from a total of 1.08 billion users are running their mobile applications in Android platform (<http://www.go-gulf.com/blog/smartphone>). Indeed, it has been a paradigm shift to design, develop and implement mobile text recognition applications for indoor signage objects.

The VIPs are not able to see the texts on the indoor sign boards clearly. However, there are many electronic gadgets available on the market but they are either too expensive or difficult for VIPs to handle those devices. We were inspired to initiate this research after a site visit to the National Council of the Blind Malaysia (NCBM) in Kuala Lumpur. From the interview sessions, we have learned that the VIP desperately needs affordable electronic gadgets to assist in navigating around various places independently without the help of caretakers (Foong and Safwanah, 2011).

**Corresponding Author:** Oi-Mean Foong, Department of Computer and Information Sciences  
Universiti Teknologi PETRONAS, Malaysia

This is especially pertinent for VIP who works as professionals such as lawyer, accountant or computer network manager (Tocheh, 2006).

Unlike previous TSR research which was developed in standalone system using MATLAB Release 2012 (Foong and Safwanah, 2011; Hairuman and Foong, 2011), an attempt was made to develop a Text Signage Recognition (TSR) Model in Android mobile devices for better portability. Based on our previous research, Hairuman and Foong (2011) applied canny edge detection algorithm, Hough Transformation and Shearing Transformation to detect and correct skewed signage images. In this study, the main objective is to propose a light weight text signage recognition model for VIPs in Android mobile devices using Tesseract OCR engine.

However, there were many challenges in text signage recognition such as low quality input images, limited bandwidth, low memory capacity, poor resolution and inadequate lighting. Adapting OCR on mobile devices poses another challenge, i.e., mobile devices have much lower processing capabilities than standalone workstation or personal computer. This portrayed an obstacle in particular for the recognition of signage images in mobile devices. The problem was intensified by the use of mobile applications that required real-time or on-the-fly processing. Fortunately, mobile devices offer its advantages beyond portability and ubiquity. Since smart phones are getting more affordable and widely use, they are no longer just a communication device. Smart phones or Android mobile devices are used to capture text images from signage objects and translate them to voice for VIPs or foreign tourists whose native languages are not in English.

Hence, we focus on offline mode in the development of text signage recognition for the VIP in Malaysia. The text images were translated by Tesseract OCR engine. Random public signage was snapped. Texts were extracted from the captured image that will be translated to text characters. Eventually, the text will be recognized and read out the text. The authors had implemented the TSR system in Samsung Galaxy S2 device using Android version 2.3.

The remainder of the paper is organized as follows. In Section 2 is the literature review. Section 3 describes the proposed Text Signage Recognition model. Section 4 includes the experimental results. Section 5 discusses the experimental findings. Section 6 describes some of the challenges and limitations of text Signage recognition. Section 7 states the conclusion and future work.

## 2. RELATED WORK

Recent researches in OCR using Neural Network or pattern recognition in which texts in an image are translated into a format that computers can manipulate using mathematical formula. There are numerous devices and system applications that are beneficial to the VIPs. One of the most common devices is the domestic helper robot that can recognize scene texts in real-time (Ruiz *et al.*, 2013). Case *et al.* (2011) had built a robot to read the texts from office signage and performed a semantic mapping between the office location and the name of the office occupant. Those devices are helpful to the VIPs but there exists several drawbacks. First of all, the devices are cumbersome to carry around. Secondly, the devices lack ergonomics features for the VIP to operate on the devices.

The rapid development in mobile technology and the increasing number of smart phones users have captivated the attention of mobile application developers. Quite a number of OCR related applications that run on smart phones had been developed. There are numerous text detection/recognition and translation applications (Hsueh, 2011; Lue *et al.*, 2010) from the market but they are without speech synthesizer. Those applications can be used at anytime and anywhere because they can be installed on the smart phones and carried around. However, it is mandatory to have internet connectivity for smart phone users to run their mobile applications in back-end server. Undeniably, it will be costly to do roaming if free WIFI is unavailable.

A considerable amount of related work that incorporates speech synthesizing ability with text recognition could be found in Tian *et al.* (2010; 2013); Yi and Tian (2011; 2012; 2013); Wang *et al.* (2013) and Arditi and Tian (2013). The central theme of these works revolves around independence navigation and way finding by exploiting the capabilities of camera-based computer vision. Tian *et al.* (2010; 2013) and Wang *et al.* (2013) described in detailed manner their way finding methods to access unfamiliar indoor environments such as different rooms and other building amenities. Similarly, Yi and Tian (2011; 2012) detailed out their proposed framework that could extract text regions from scene images with complex backgrounds and multiple text appearances. Taking a different perspective that is more to the end-users' viewpoint, Arditi and Tian (2013) reported on the user interface preferences in the design of a camera-based navigation and way finding. The research's finding confirmed on the need for devices that can guide the VIPs to architectural landmarks in their environment. Likewise, the finding indicated preferences towards device

interfaces that could give the VIPs control over the new surroundings learnt.

Recently Zhang *et al.* (2013) conducted a study on state-of-the-art algorithms on scene text classification and detection and claimed that his proposed algorithm significantly outperformed the existing algorithms for character identification. Chang (2011) came up with an intelligent text detection and Extraction technique using Canny Edge Detection algorithm in addition to using fast Connected Component algorithm to filter noise to obtain features and candidate text. Gonzalez and Bargasa (2013) applied geometric and gradient properties to design a text reading algorithm for natural images. Carlos *et al.* (2013) presented a method for perspective recovery of natural scene texts where the geometry of the characters was estimated to rectify homography for every line of text. Koo and Tian (2013) and Shi *et al.* (2013) extracted connected components in images using Maximally Stable Extremal Region (MSER) algorithm to perform scene text detection and recognition. In addition, Liu and Smith (2013) invented a simple algorithm that used the density of special symbols to detect equation regions in printed document. They claimed that the performance of the Tesseract OCR engine had improved after the detector was enabled. Thus, we decided to deploy Tesseract OCR engine as a backend engine for this research because we discover that Tesseract OCR engine is open source (<http://code.google.com/p/tesseract-ocr/>) and it is ranked the top three of the optical character recognition system at the Fifth Annual Test of OCR Accuracy (Rice *et al.*, 1996).

### 3. PROPOSED TEXT SIGNAGERECOGNITION (TSR) MODEL

As mentioned earlier, the Tesseract OCR engine has been shortlisted for this research because it is reliable, open source and free of charge. The recognition process of the Tesseract OCR engine mainly consists of features detection, segmentation, features extraction and lastly character recognition steps (Smith, 2007).

To shorten the development time, the authors had adopted also the Open Source Computer Vision (OpenCV)'s libraries for image processing because it contains a list of real time computer vision library functions which is free for academic use (<http://opencv.willowgarage.com/wiki/Welcome/>).

**Figure 1** shows the proposed Text Signage Recognition (TSR) Model. In a nutshell, the TSR model recognizes text in signage image based on the following steps:

- Image Acquisition

- Image Pre-Processing
- Character Recognition using Tesseract OCR Engine
- Speech Synthesizing Process

#### 3.1. Image Acquisition

The text image was captured using Samsung Galaxy S2 device and stored as standard jpeg format or the image was selected from secured digital card in Android smart phone. The image may contain some unwanted noises such as blur images, complex background objects, illuminating objects. Thus, pre-processing steps are required to smoothen or reduce these noises for better recognition rate.

#### 3.2. Image Pre-processing

The image pre-processing eliminates unwanted noise in signage objects. It comprises of a series of operations like convert to grayscale, morphology transformation, thresholding, find the edges of the characters, copy those characters in new white background and finally convert the gray scale image to binary image. The entire steps of image preprocessing are illustrated in **Fig. 2**.

The original colored image in **Fig. 3** is converted into gray scale image by using `cvtColor()` function with "COLOR\_BGR2GRAY" color space conversion code. (<http://opencv.willowgarage.com/wiki/Welcome/>). The gray scale image carries intensity information.

The formula used for color space conversion to gray scale is as shown in Equation (1) (Juan-Li *et al.*, 2010):

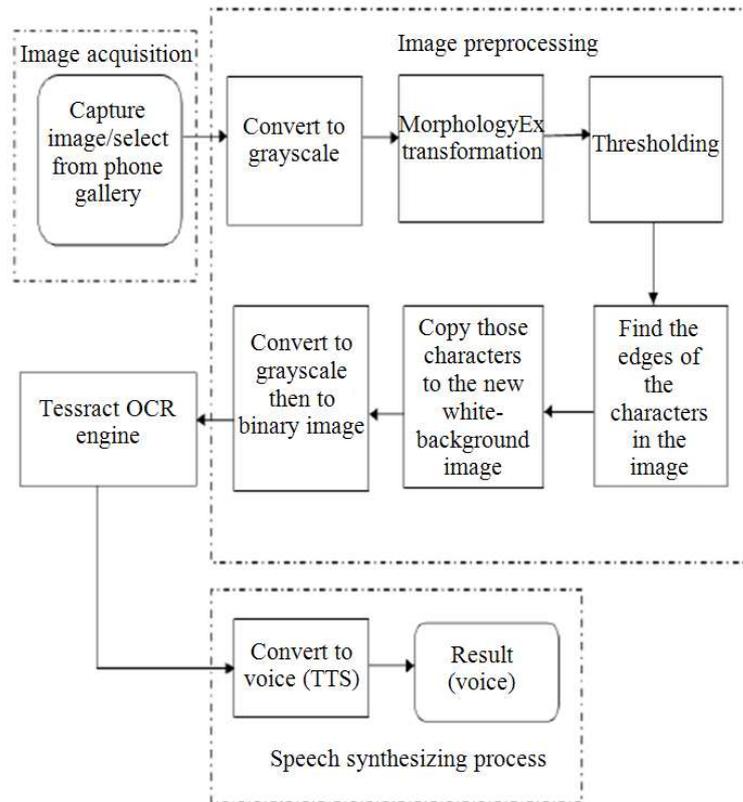
$$Y = 0.299R + 0.587G + 0.114B \quad (1)$$

where, R, G and B are the normalized input Red, Green and Blue pixel values.

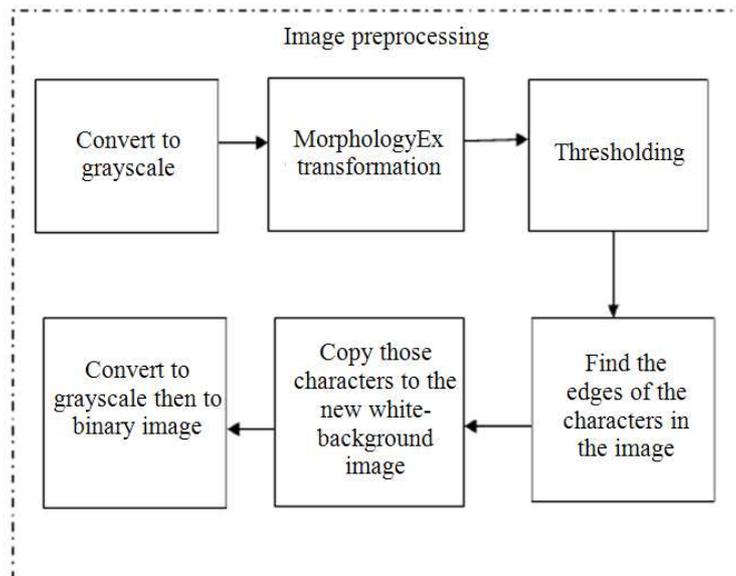
Morphological Transformations will be carried out next on the image as shown in **Fig. 4**. `Morph_Open` (<http://opencv.willowgarage.com/wiki/Welcome/>) is the morphological operation that has been used in this research project. The image was first eroded and then dilated as shown in the following codes:

```
dst = dilate(erode (src,element), element)
```

Thresholding operation that was performed to convert from gray scale to binary image based upon Otsu's algorithm (Otsu, 1979). The digital gray scale image was partitioned into multiple segments or sets of pixels. The selected pixels must fall within the limit of the threshold value. Otherwise the pixels values that are either above or below the threshold value will be rejected.



**Fig. 1.** Proposed text signage recognition model



**Fig. 2.** Image pre-processing steps



Fig. 3. Original colored image



Fig. 5. Edges of the characters



Fig. 4. Gray scale image



Fig. 6. Masked image

By searching for the contours of characters in the binary image, the edges of the characters in the binary image were found as shown in Fig. 5. The function `findContours()` connects all the continuous points with the same intensity together from the binary image (<http://opencv.willowgarage.com/wiki/Welcome/>).

A mask was created by applying the `drawContours()` function from the OpenCV library (<http://opencv.willowgarage.com/wiki/Welcome/>). The mask would eliminate and darken all background objects in the image and allow only the characters on the signage image to be visible as depicted in Fig. 6.

The masked image was copied and pasted onto another image with white background as illustrated in Fig. 7.

The image is finally converted to binary image before proceeding to the character recognition process.

### 3.3. Character Recognition

The processed binary image will be further processed and converted to text during the character recognition process (Smith, 2007) without any internet connection. The characters outlined in the binary image were analyzed into Blobs using Connected Component Analysis. The Blobs were organized into text lines and then further divided into words.



Fig. 7. New image with white background

### 3.4. Speech Synthesizing Process

The words or texts were then fed into the speech synthesizer for the Text-To-Speech (TTS) translation. The TTS read out the words if recognized correctly.

Otherwise, it either read out all the characters in the word one after another or no output was produced.

#### 4. RESULTS

For the TSR prototype, the authors adopted a new paradigm shift to design its mobile user interface as simple as possible. Every option is performed in a few clicks. **Figure 8** shows the mobile user interface design that consists of two options, i.e., either to select image from phone gallery or from its camera in the homepage.

##### 4.1. Experiment Setup and Implementation

A total of 60 different samples of signage images were selected from the ICDAR 2003 dataset (<http://algoval.essex.ac.uk/icdar/RobustWord.html>). Of all the random samples selected, 60% of the selected images were used for training while the remaining samples were used for testing. The aim of the experiment was to determine the recognition rate of the TSR model. We used two open sources libraries i.e., OpenCV (Bradski and Kaehler, 2008) and Tesseract OCR engine during the development of the TSR mobile application. We installed the proposed TSR model on Samsung Galaxy SII device for system training and testing.

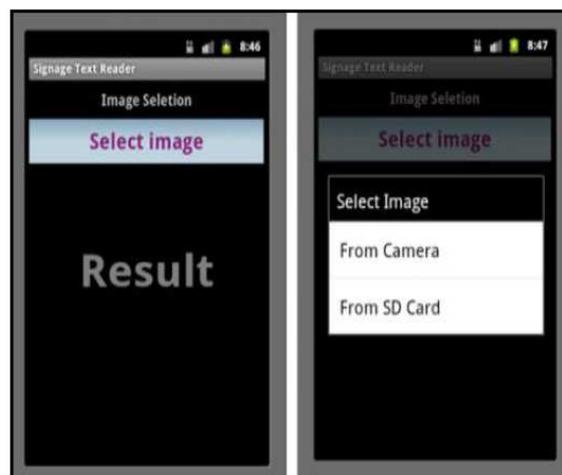
##### 4.2. Training and Testing

Prior to system performance testing, we had four volunteers/users (3 males and 1 female) between 20-22 years old to train the TSR mobile application. The users manually input those images into the proposed model for training purpose. Samples of training and testing data were displayed in **Fig. 9** and **10** respectively. The results of training and testing were tabulated in **Table 1**.

The Recognition Rate (RR) of the proposed TSR model is calculated using Equation (2):

$$RR = \frac{\text{No.of Recognized Signage}}{\text{Total No.of Signage}} \times 100\% \quad (2)$$

Based on our observation, the recognition rate for training is 16.7% higher than that for testing dataset as illustrated in **Table 1**. The huge discrepancy between training and testing dataset was due to noises that were inherent in the selected testing images. The testing samples downloaded from the ICDAR2003 dataset contained complex background objects which constituted the noise.



**Fig. 8.** TSR in android mobile device



**Fig. 9.** Sample ICDAR 2003 Training Images



**Fig. 10.** Sample ICDAR 2003 testing images

##### 4.3. User Evaluation on Proposed TSR Model

The TSR model was tested by the four volunteers/users to estimate its recognition rate. All users were blind-folded. 15 signage images were randomly selected in which some characters were slanted or skewed. Each signage image was captured using Samsung Galaxy SII device. If the signage was recognized correctly (T = True), a tick (✓) was placed in the entry. Otherwise the signage was not recognized (F = False). **Table 2** shows the user evaluation results for the ICDAR 2003 Robust Word Recognition dataset.

**Table 1.** Training And testing results on ICDAR 2003

	No. samples	No. recognized samples	Recognition Rate (RR)
Training	36	33	91.7 %
Testing	24	18	75.0 %
Total	60	51	

**Table 2.** Evaluation of blind-folded users

Sample Signage Images	Blind-Folded Users							
	1		2		3		4	
	T	F	T	F	T	F	T	F
STAFF	√		√		√		√	
Alarm		√	√		√			√
AGENTS	√			√	√		√	
PEPPER	√		√			√	√	
2002	√		√		√		√	
GAS	√		√		√		√	
TRAFFIC	√		√		√		√	
South	√		√		√		√	
Videos	√		√		√			√
1946		√	√			√	√	
EST.	√			√		√		√
E14	√		√		√		√	
FLATS	√		√		√		√	
SERVICE	√		√		√		√	
YOU	√		√		√		√	

From the experimental results illustrated in **Table 2**, it appeared that the TSR model could not detect and recognize the text signage for “Alarm” by user 1 and user 4, “AGENTS” by user 2, “PEPPER” by user 3, “Videos” by user no. 4, “1946” by user 1 and user 3, “EST.” by user 2 and “YOU” by user 2 and user 3. However, the TSR could recognize signage for “STAFF”, “2002”, “GAS”, “TRAFFIC”, “South”, “E14”, “FLATS” and “SERVICE”. In general, most slanted or skewed text images could be detected and recognized by the Tesseract OCR engine.

#### 4.4. Comparison of Performance Evaluation Using Precision, Recall and F-Score

In this subsection, we compared the proposed TSR model’s system performance results with other alternative models in terms of Precision, Recall and F-

Score metrics. The F-Score indicates the Harmonic Mean of Precision and Recall metrics. The Precision, Recall and formulas are given in Equation (3 to 5) respectively:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

where, TP represents True Positive for text that can be recognized and pronounced correctly, FP represents False Positive for text that can be recognized but does not pronounced correctly and FN represents False Negative for text that cannot be recognized at all.

## 5. DISCUSSION

We observed that the TSR model had encountered some hiccups in text recognition process. Firstly, it was noticed that those signage images with white colored text failed to be recognized correctly. This was because when texts were extracted, masked, copied followed by pasted onto a new white background image, the white texts were blended into the new image that had white background. Secondly, it recognized the character ‘O’ incorrectly as ‘Q’ and vice versa. Thirdly, signage that had illuminating texture cannot be recognized at all. Fourthly, any incomplete and cropped character in the captured image would be recognized wrongly by the TSR model as well. On the other hands, the proposed TSR model could recognize well on text signage that had simple/plain background. In addition, sometimes the speech synthesizer failed to speak the word correctly. For example, it pronounced the word “YOU” as “Y”, “O” and “U” by reading out the characters one by one. The signage recognition would deteriorate further if the image captured were either too dim or too bright since the users were blind-folded.

As there was no prior comparative study conducted on the ICDAR 2003 dataset using the Recognition Rate formula in Equation (2), further testing on more ICDAR 2003 images were conducted using alternative metrics (i.e., Precision, Recall, F-Score). The experimental results of the proposed TSR model were demonstrated in **Table 3**.

**Table 3.** Benchmark results on ICDAR 2003 dataset

Method	Precision	Recall	F-Score
Yao's method	0.62	0.65	0.63
Proposed TSR	0.60	0.63	0.61
Ashida's method	0.55	0.46	0.50

From **Table 3**, Yao *et al.* (2013) demonstrated that their algorithm had achieved average F-Score of 0.63 by minimizing energy cost function for each region in its surrounding text. The proposed TSR model which used the Tesseract OCR engine had performed better than Ashida's method (Lucas *et al.*, 2003) that gave an average F-Score of 0.50. But our TSR model produced average F-Score of 0.61 which was below Yao's method (Yao *et al.*, 2013).

## 6. CHALLENGES AND LIMITATIONS

Text detection and recognition have posed many challenges to research community. Researchers have attempted to bridge the research gap by improving the text detection and recognition techniques. Yet, there exist plenty of rooms for improvement. The challenges still persist in scene text recognition such as cluttered or complex background, small font size, inappropriate font style, slanted/skewed images, embossed images and poor lighting are the main contributing factors of poor detection and recognition rates. However, the greatest challenge for the visually impaired or blind people in capturing scene text is that the captured image tends to be out of focus, missing or partial text appeared in the captured image, or even image that contains illuminating color. Undoubtedly, lack of contrasting colors between foreground and background objects would also be a stumbling block to better text detection and recognition.

In the contrary, with the advancement in mobile technology, the prices of mobile devices have become cheaper and affordable. But, all mobile devices still have little memory and limited processing capability to process indoor text images. The offline operating mode is helpful in avoiding heavy roaming costs while using the TSR mobile application abroad.

## 7. CONCLUSION

The proposed TSR model which was deployed with Tesseract OCR engine could successfully recognize the text images from the International Conference on Document Analysis and Recognition (ICDAR) 2003 dataset with satisfactory recognition rate. We attempt to build a Text Signage Recognition model in Android mobile devices for visually impaired people in Malaysia.

Our approach is simple and easy to implement. The proposed model does not require the internet connection during its operating mode because the Tesseract OCR engine was embedded in the Android mobile devices. However, the major drawback for implementing text signage recognition is that its recognition performance suffers when compared to state-of-the-art algorithms for text extraction and recognition. Yet, the proposed TSR Model is capable of translating texts to speech through the built-in speech synthesizer.

At this juncture, the TSR model could recognize and pronounce English words only. In future, we intend to enhance the TSR prototype by translating any English texts to different languages such as Malay or Mandarin languages.

## 8. ACKNOWLEDGEMENT

Special thanks and gratitude to the National Council of the Blind Malaysia (NCBM) and the Research and Innovation Office (RIO) of Universiti Teknologi PETRONAS for providing support in this research.

## 9. REFERENCES

- Al-Saleh, A., A. El-Zaart and A.M. Al-Salman, 2011. Dot detection of Braille images using a mixture of beta distribution. *J. Comput. Sci.*, 7: 1749-1759. DOI: 10.3844/jcssp.2011.1749.1759
- Arditi, A. and Y. Tian, 2013. User interface preferences in the design of a camera-based navigation and wayfinding aid. *J. Visual Impairment Blindness*, 107: 118-129.
- Bradski, G. and A. Kaehler, 2008. *Learning OpenCV*. 1st Edn., O'Reilly Publications, Sebastopol, ISBN-10: 0596554044, pp: 580.
- Carlos, M.G., M. Majid, S. Jose and L.G.M. Jose 2013. Fast perspective recovery of text in natural scenes. *Image Vis. Comput.*, 31: 714-724. DOI: 10.1016/j.imavis.2013.07.002
- Case, C., B. Suresh, A. Coates and A.Y. Ng, 2011. Autonomous sign reading for semantic mapping. *Proceedings of the IEEE Conference on Robotics and Automation*, May 9-13, IEEE Xplore Press, Shanghai, pp: 3297-3303. DOI: 10.1109/ICRA.2011.5980523
- Chang, R.C., 2011. Intelligent text detection and extraction from natural scene images. *Proceeding of the 15th North-East Asia Symposium on NANO, Information Technology and Reliability*, Oct. 24-26, IEEE Xplore Press, Macao, pp: 23-28. DOI: 10.1109/NASNIT.2011.6111115

- Coughlan, J.M. and H. Shen, 2013. Crosswatch: A system for providing guidance to visually impaired travelers at traffic intersection. *J. Assistive Technol.*, 7: 131-142. DOI: 10.1108/17549451311328808
- Foong, O.M. and N.S.B. M. Razali, 2011. Signage recognition framework for visually impaired people. Proceedings of the International Conference on Computer Communication and Management, (CCM' 11), IACSIT Press, Singapore, pp: 488-492.
- Gonzalez, A. and L.M. Bergasa, 2013. A text reading algorithm for natural images. *Image Vis. Comput.*, 31: 255-274. DOI: 10.1016/j.imavis.2013.01.003
- Hairuman, I.F.B. and O.M. Foong, 2011. OCR signage recognition with skew and slant correction for visually impaired people. Proceedings of 11th International Conference on Hybrid Intelligent Systems, Dec. 5-8, IEEE Xplore Press, pp: 306-310. DOI: 10.1109/HIS.2011.6122123
- Hsueh, M., 2011. Interactive Text Recognition and Translation on a Mobile Device. University of California, Berkeley.
- Juan-Li, H., D. Jia-Bing and Z. Shan-shan, 2010. A novel algorithm for color space conversion model from CMYK to LAB. *J. Multimedia*, 5: 159-166. DOI: 10.4304/jmm.5.2.159-166
- Katz, B.F.G., F. Dramas, G. Parsehian, O. Gutierrez and S. Kammoun *et al.*, 2012. NAVIG: Augmented reality guidance system for the visually impaired. *Virtual Reality*, 16: 253-269. DOI: 10.1007/s10055-012-0213-6
- Koo, H. and Y. Tian, 2013. Scene text detection via connected component clustering and nontext filtering. *IEEE Trans. Image Process.*, 22: 2296-2305. DOI: 10.1109/TIP.2013.2249082
- Liu, Z. and R. Smith, 2013. A simple equation region detector for printed document images in tesseract. Proceedings of 12th International Conference on Document Analysis and Recognition, Aug. 25-28, IEEE Xplore Press, Washington, DC., pp: 245-249. DOI: 10.1109/ICDAR.2013.56
- Lucas, S.M., A. Panaretos, L. Sosa, A. Tang and S. Wong *et al.*, 2003. ICDAR 2003 robust reading competitions. Proceedings of the 7th International Conference on Document Analysis and Recognition, Aug. 3-6, IEEE Xplore Press, pp: 682-687. DOI: 10.1109/ICDAR.2003.1227749
- Lue, H.T., M.G. Wen, H.Y. Cheng, K.C. Fan and C.W. Lin *et al.*, 2010. A novel character segmentation method for text images captured by cameras. *ETRI J.*, 32: 729-739. DOI: 10.4218/etrij.10.1510.0086
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernet.*, 9: 62-66. DOI: 10.1109/TSMC.1979.4310076
- Rice, V.S., F.R. Jenkins and T.A. Nartker, 1996. The Fifth Annual Test of OCR Accuracy. 1st Edn., Information Science Research Institute, Las Vegas, pp: 44.
- Ruiz, J.A.A., P. Ploger and G.K. Kraetzschmar, 2013. Active Scene Text Recognition for a Domestic Service Robot. In: RoboCup 2012: Robot Soccer World Cup XVI, Chen, X., P. Stone and L.E. Sucar (Eds.), Springer, ISBN-10: 3642392512, pp: 249-260.
- Sarfraz, M. and S.M.A.J. Rizvi, 2007. Indoor navigation aid for the visually impaired. Proceedings of the Geometric Modeling and Imaging, (MAI' 2007), pp: 127-132.
- Shi, C., C. Wang, B. Xiao, Y. Zhang and S. Gao, 2013. Scene text detection using graph model built upon maximally stable extremal regions. *Patt. Recogn. Lett.*, 34: 107-116. DOI: 10.1016/j.patrec.2012.09.019
- Smith, R., 2007. An overview of the tesseract OCR engine. Proceedings of the 9th International Conference on Document Analysis and Recognition, Sept. 23-26, IEEE Xplore Press, Parana, pp: 629-633. DOI: 10.1109/ICDAR.2007.4376991
- Tian, Y., X. Yang, C. Yi and A. Arditì, 2013. Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. *Mach. Vis. Applic.*, 24: 521-535. PMID: 23630409
- Tian, Y.L., C. Yi and A. Aries, 2010. Improving computer vision-based indoor wayfinding for blind persons with context information. Proceedings of the 12th International Conference on Computers Helping People with Special Needs, Jul. 14-16, Springer Berlin Heidelberg, Vienna, Austria, pp: 255-262. DOI: 10.1007/978-3-642-14100-3\_38
- Tocheh, M.G., 2006. Siapa Bersama Mereka. National Council for the Blind (NCBM), Laser Press Sdn Bhd, Malaysia.
- Wang, S., X. Yang and Y. Tian, 2013. Detecting signage and doors for blind navigation and wayfinding.. *Netw. Model. Anal. Health Inform. Bioinform.*, 2: 81-93. PMID: 23914345
- WHO, 2013. Visual Impairment and Blindness. World Health Organization.
- Yao, L., C. Shen, W. Jia and A.V.D. Hengel, 2013. Leveraging surrounding context for scene text detection. Proceedings of the IEEE ICIP, (ICIP' 13), pp: 2264-2268.

- Yi, C. and Y. Tian, 2011. Assistive text reading from complex background for blind persons. Proceedings of the 4th International Conference on Camera-Based Document Analysis and Recognition, Sept. 22-22, Springer Berlin Heidelberg, Beijing, China, pp: 15-28. DOI: 10.1007/978-3-642-29364-1\_2
- Yi, C. and Y. Tian, 2012. Localizing text in scene images by boundary clustering, stroke segmentation and string fragment classification. IEEE Trans. Image Process., 21: 256-4268. DOI: 10.1109/TIP.2012.2199327
- Yi, C. and Y. Tian, 2013. Text extraction from scene images by character appearance and structure modeling. Comput. Vis. Image Understand., 117: 182-194. DOI: 10.1016/j.cviu.2012.11.002
- Zhang, H., K. Zhao, Y.Z. Song and J. Guo, 2013. Text extraction from natural scene image: A survey. Neurocomputing, 122: 310-323. DOI: 10.1016/j.neucom.2013.05.037