Detecting Text Based Image With Optical Character Recognition for English Translation and Speech using Android

Sathiapriya Ramiah
School of Computing & Technology
Asia Pacific University of Technology & Innovation, Technology Park Malaysia
57000 Bukit Jalil,
Kuala Lumpur, Malaysia
sathiapriya@apu.edu.my

Tan Yu Liong
Asia Pacific University of Technology & Innovation, Technology Park Malaysia
57000 Bukit Jalil,
Kuala Lumpur, Malaysia
yuliong1015@gmail.com

Manoj Jayabalan
School of Computing & Technology
Asia Pacific University of Technology & Innovation, Technology Park Malaysia
57000 Bukit Jalil,
Kuala Lumpur, Malaysia
manoj@apu.edu.my

Abstract*— Smartphones have been known as most commonly used electronic devices in daily life today. As hardware embedded in smartphones can perform much more task than traditional phones, the smartphones are no longer just a communication device but also considered as a powerful computing device which able to capture images, record videos, surf the internet and etc. With advancement of technology, it is possible to apply some techniques to perform text detection and translation. Therefore, an application that allows smartphones to capture an image and extract the text from it to translate into English and speech it out is no longer a dream. In this study, an Android application is developed by integrating Tesseract OCR engine, Bing translator and phones' built-in speech out technology. Final deliverable is tested by various type of target end user from a different language background and concluded that the application benefits many users. By using this app, travelers who visit a foreign country able to understand messages portrayed in different language. Visually impaired users are also able to access important message from a printed text through speech out feature.

Keywords—Android, OCR, text translator, text to speech

I. INTRODUCTION

Real world contains too many significant message and useful information but unfortunately most of them are written in different official language depends on the host country. Sometimes a signboard or any other notice could carry an important message or even danger. If the message is unreachable to mankind with different language background, it might cause important information to be missed out [1]. Besides that, it is inconvenient for a travelers to carry on their tasks in a foreign country if they don't understand the language used in that country. They need to carry a pocket dictionary or use online translation service in order to understand the message. However, a pocket dictionary might not be helpful if the users want to translate a language that does not group by alphabets [2]. It is also meant the same way in another study that users are unable to write the text of what they see. This issue might cause a communication breakdown for mankind from a different language background as they are unable to understand the language even though the pocket dictionary and online translation service provided [3].

In 2014, the World Health Organization (WHO) estimated that there are 285 million people to be visually impaired, total of 246 million with visual impairment and 39 million are blind. It is also reported that 90% of the world's visually impaired people are from low income group [4]. One of the major problems faced by visually impaired people is they are not capable in accessing printed text. Although there are numbers of assistive technology meant for visually impaired, most of these special devices are not convenient because it require custom modifications and some are too expensive. Almost 70% of visually impaired people are unemployed and most of them are unable to use assistive technology due to its cost [5]. This causes visually impaired users missed the opportunity to access important text that is present in the world to carry out day to day task efficiently.

In order to overcome these issues, this paper proposes to develop an Android application which capture text based image which carries important messages from real world and translate them into English and finally pronounce it.

II. PRIMARY STUDY ON TECHNOLOGY

A. Optical Character Recognition(OCR)

OCR is a technology that allows users to convert text or documents in images captured by an input device into an editable, searchable and reusable data type for further image processing. This technology enables a machine to recognize the characters automatically through an optical mechanism just like a human being use eyes to see an object in the world. At the early stage of introducing OCR, this technology encountered several problems such as limitations in terms of the quantity and complexity of the hardware and the algorithm [6]. However, OCR has been widely used in many areas including cheque processing, digital libraries, recognizing text from natural scenes, understanding hand-written office forms and etc. As years go on, OCR has evolved and became more and more mature with the advancement of technologies and contributions of well-known companies such as IBM, HP, Microsoft, Google and etc through ongoing researches.

An OCR system is a combination of several subsystems, and each of the subsystem itself is dedicated to solve certain

978-1-4673-9572-4/15/$31.00 ©2015 IEEE

problems and perform different roles in image processing [7]. Although there are numerous algorithms available out there, most still follow the core steps which are discussed below:

1. *Image Capturing and Preprocessing*

Firstly, images are captured in this phase by a camera or some other device. Extraction of scanned images with a white background and black character foregrounds are easy to be detected but the camera captured images may contain noise because of the environmental reasons and low brightness of the images [8]. Therefore, there are some techniques like image enhancement, binarization and noise reduction to be done in the preprocessing phase to increase the performance and accuracy of a character recognition system.

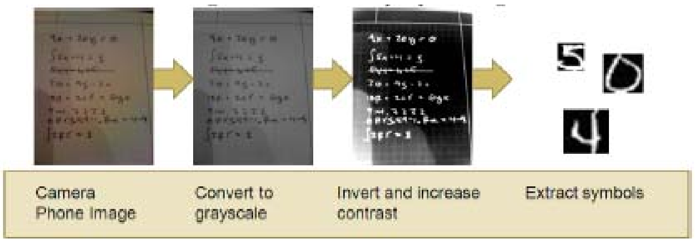
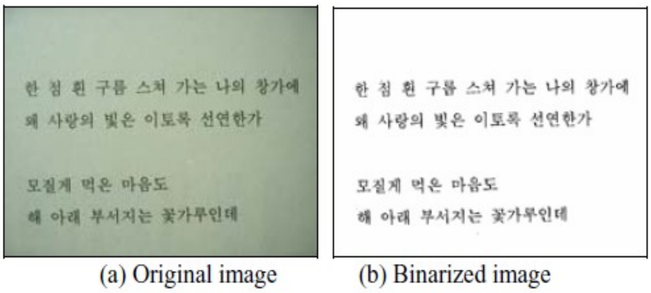
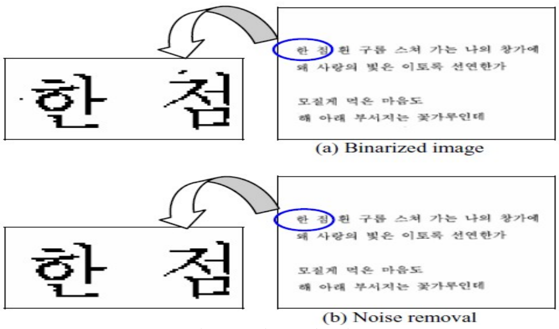


Fig 1 Image enhancement
Source (Hymes & Lewin, 2008)


한 점 흰 구름 스쳐 가는 나의 창가에
왜 사랑의 빛은 이토록 선연한가
모질게 먹은 마음도
해 아래 부서지는 꽃가루인데

(a) Original image (b) Binarized image

Fig 2 Binarization
Source: (Bae et al., 2005)



(a) Binarized image

(b) Noise removal

Fig 3 Noise reduction
Source: (Bae et al., 2005)

2. *Text segmentation*

Extraction of correct character boundaries is very important for recognizing a character [8]. The segmentation of a binary image in a regular sequence can be categorized into lines, words and characters. There are many well known segmentation methods available which are projection, region growing and tracing contour etc.

3. *Text recognition*

In the recognition phase, after the character is segmented, they are normalized by removing noise. Lastly, OCR extracts the character and recognizes it.

Although OCR is useful, it is not perfect without any issues. Researchers have discovered problems such as light condition, text skew, perception distortion, misalignment of text, blur and difficulty in recognizing handwritten document [3]. So these are the challenges for the technologist to conduct further research on enhancing this technology.

With the evolving and remarkable technology in mobile device, mobile phones are now capable to capture high resolution images with at least 1280 X 960 pixels, which are more compatible and have higher chances to be detected with OCR. Implementing OCR application on a mobile device could be realistic as there are many ongoing researches in this field [9]. However, mobile phones have their drawbacks too. One of the drawbacks is limited power to run complex software like OCR engines due to limited hardware and memory resources unlike desktops. In addition, real time response is another critical issue need to be considered.

B. *Translator*

Most of the language translator engines are available on the web based application compared to mobile based [11]. However there are translation engines provided in mobile environment with limited number of supported languages.

As research conducted, there are 3 different algorithms of how a translator engine can be performed. They are based on crowdsourcing [3], where the OCR resulted character will be sent to a group of human workers to carry out translation task, an online translation service [12], where character extracted will be delivered to Google Translator to do the actual text to text translation, and lastly real time translator, which integrates a translator engine and dictionary on mobile phone to do the translation.

Crowdsourcing translation requires too many human resources. The translations made sometimes are not reliable and slow because of translating word by word sequentially. Another issue might be faced that if the translation is done word by word, the meaning of the original sentences might be affected and may lead to a different meaning [11].

On the other hand, online translation service provides more reliable translation to requester and it only requires a small amount of data package to be embedded at the backend of the smartphone. Internet connection is required to update the translator on need basis to ensure more accurate translation. It is useful to access the right information instead of getting wrong information.

Real time translator is considerable effective too. However, it requires huge data to be embedded into the backend of smartphone and languages need to be updated to the latest version every time it is used.

III. Existing System

Existing systems that are similar to the proposed system are commercial applications such as Google Goggles, ABBYY Fototranslate and Word Lens.

Word Lens is an augmented reality translation application that helps users to access their daily information from the real world without connecting to internet. It works in iOS and provides translation for European languages. Word Lens is currently not capable of detecting unscripted language such as Chinese, Korean and Japanese which does not group by alphabet and it does not provide a text to speech function.

Google Goggles is another interesting application introduced by Google that are available for Android and iPhone. Google has large databases that allow users to stay connected to it. This application provides users the ability to capture the images of different kinds of objects such as text, landmarks, books, artwork and logos and match it with the keyword in database. Then it will return the related information as a final result. The performance and accuracy reduces without internet connection and it does not provide a text to speech function.

ABBYY FotoTranslate is a commercial application to instantly do translation on Nokia smartphones. The application can work offline without internet connection and support up to 8 languages and 30 directions for translation. It is currently available in Nokia and it does not provide a text to speech function either.

Besides it, several studies are also conducted such as summarized on Table 1 below.

TABLE I OCR Projects On Mobile Devices

Authors, Publication Year	Mobile OCR Engine	OS	Translator	Text-To-Speech
L. Zhifang, L. Bin and G. Xiaopeng, 2010 [3]	Not Stated	Not Stated	Crowd-sourcing	-
H. Nakajima, Y. Matsuo, M. Nagata and K. Saito, 2005 [2]	Not Stated	Not Stated	Yes, Not Stated	-
M. Laine and O. Nevalainen, 2006 [7]	Not Stated	Symbian	-	-
K. Bae, K. Kim, Y. Chung and W. Yu , 2005 [8]	Not Stated	Not Stated	-	-
A. Canedo-Rodriguez, S. Kim, J. Kim and Y. Blanco-Fernandez , 2009 [1]	Not Stated	Not Stated	Yes, Not Stated	-
V. Fragoso, S. Gauglitz, S. Zamora, J. Kleban and M. Turk , 2011 [12]	Tesseract	Maemo5 (Nokia N900)	Google Translator	-
Koga M, R. Mine, T. Kameyama, T. Takahashi, M. Yamazaki and T. Yamaguchi , 2005 [13]	Not Stated	Not Stated	Yes, Not Stated	-
L. Chang and S. ZhiYing , 2009 [10]	Not Stated	Not Stated	-	-
A. Shaik, G. Hossain and M. Yeasin, 2010 [5]	Tesseract	Android	-	Yes

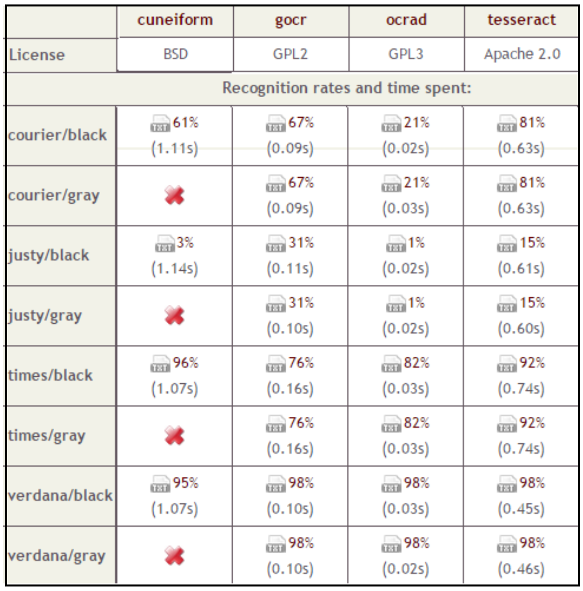
Studies shown on Table 1 are related to OCR projects on mobile or PDA environment. Most of them are not published to the market. Based on table above, most of the studies are mainly focused on the OCR technology which recognizes the characters only. There is no research includes text to speech feature except the last shown in the table. Besides that, no research conducted to incorporate OCR engine, translator and text to speech features. Therefore the proposed application still has its unique and trends compare to above listed researches.

IV. Chosen Technologies

A. OCR

There are numerous open sources as well as commercial OCR engines available in the market today with their own strengths and weaknesses. Many open source communities offer engines such as GOCR, Cuneiform, OCRAD, Tesseract and OCROPUS. There are commercially available OCR engines such as ABBYY Fine Reader, OmniPage, and Microsoft Office Document Imaging.

Tesseract is an open source engine developed by HP labs between years 1985 to 1995 and then handed over to Google Inc. in 2006. Tesseract combined with the Leptonica Image Processing library which can read a wide variety of image formats and convert them to text in over 60 languages. It works well on all computer operating system as well as Android and iPhone mobile platform. Due to popularity of Tesseract being open source engine, there are a lot of academic experiments and OCR software developments conducted successfully. Based on study conducted between OCRAD, GOCR and Tesseract, found out that the Tesseract outperform other open source engines. Despite of unclean data, Tesseract proved the best free and open source OCR engine in term of accuracy and processing time as shown in Fig 4 [14].



	cuneiform	gocr	ocrad	tesseract
License	BSD	GPL2	GPL3	Apache 2.0
Recognition rates and time spent:				
courier/black	61% (1.11s)	67% (0.09s)	21% (0.02s)	81% (0.63s)
courier/gray	✗	67% (0.09s)	21% (0.03s)	81% (0.63s)
justy/black	3% (1.14s)	31% (0.11s)	1% (0.02s)	15% (0.61s)
justy/gray	✗	31% (0.10s)	1% (0.02s)	15% (0.60s)
times/black	96% (1.07s)	76% (0.16s)	82% (0.03s)	92% (0.74s)
times/gray	✗	76% (0.16s)	82% (0.03s)	92% (0.74s)
verdana/black	95% (1.07s)	98% (0.10s)	98% (0.03s)	98% (0.45s)
verdana/gray	✗	98% (0.10s)	98% (0.02s)	98% (0.46s)

Fig 4 Comparison of open source OCR engine
Source: (Andreas Gohr, 2010)

B. *Translator*

For text translation, online translation service is chosen. There are many translators available for example NiuTrans, Systran, Google Translate, OpenLogos, Bing Translator, GramTrans, Babylon and etc. Out of many, three translators which are more suitable for mobile development compared which are Google Translate, Babylon and Bing Translator.

Google translate and Bing translate are no longer new translation service to users who always use translators. They are well known translators which supports many languages all over the world. Both of them are free to use translator. On the other hand, Babylon has big difference with those mentioned above. Babylon is a commercial product that has many versions and it works well with many operating systems.

Experiment conducted on Google Translate, Babylon and Bing Translate to identify the best translator. Findings showed that Bing Translator performs better in producing more accurate result [15]. Furthermore, Bing translator is not banned in any country unlike Google services.

C. *Text To Speech*

For text to speech, phone built-in feature would perform the speech out service. Android libraries such as android.text and android.speech will be used mainly for this purpose.

V. SYSTEM DESIGN

A. *Architecture Design*

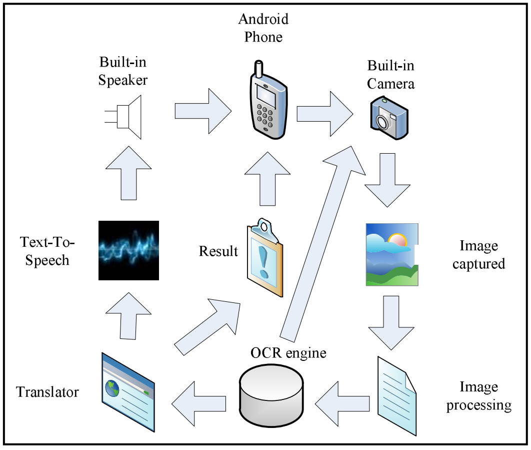


Fig 5 Architecture of the proposed application

Fig 5 presents the overall architecture of the proposed application. Once the user captures an image of text with the built-in camera of an Android phone, the image is sent for processing. The image will be processed to reduce the noise and delivered to OCR engine. The selected OCR engine will extract and recognize the text. However, if the result is not detectable based on the OCR engine, the phone will navigate user back to capture mode. Once the text extraction completed, now they will be sent to translator. In this phase, the result of OCR will be translated into English by translator and finally voice out by the phone built-in speaker.

B. *Flowchart*

Fig 6 below shows the flowchart for the proposed application.

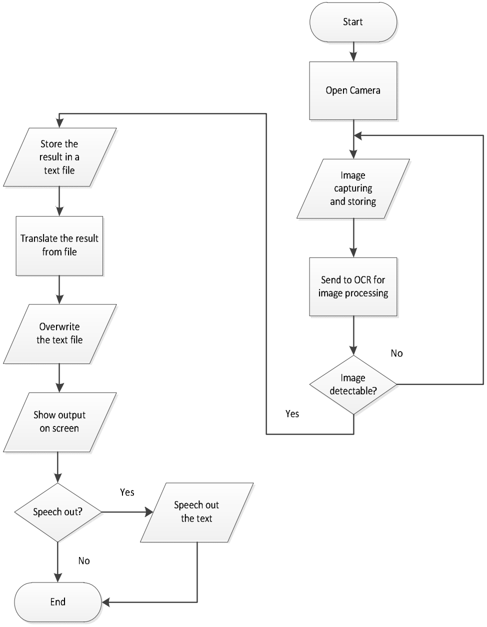

Start
Open Camera
Image capturing and storing
Send to OCR for image processing
Image detectable?
No
Yes
Store the result in a text file
Translate the result from file
Overwrite the text file
Show output on screen
Speech out?
Yes
Speech out the text
No
End

Fig 6 Flowchart of the application

C. *Package Diagram*

Fig 7 below shows the package diagram on the interactions of libraries that used for implementation.

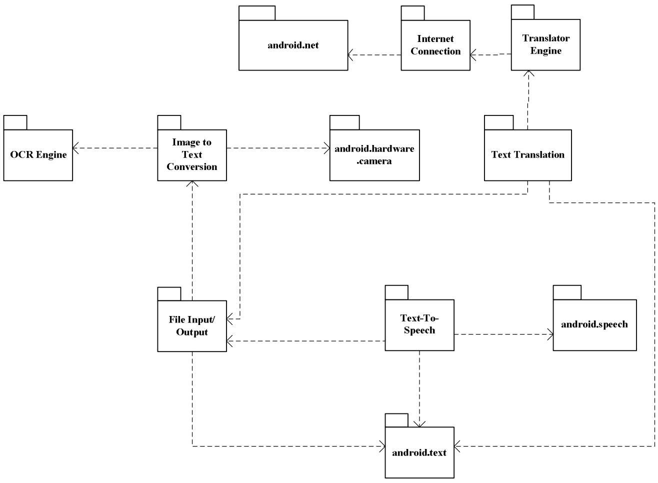


Fig 7 Package diagram of the application

VI. IMPLEMENTATION & TESTING

From the research conducted, Tesseract OCR engine and Leptonica Image Processing Libraries were selected to extract the text from image captured by phone's camera. Furthermore, enhanced feature for this research is the translation of words. The OCR result will be passed to Microsoft Translator API to provide multi language support for source language and target language will be English. Final deliverable is deployed and tested on an Android device.

Application interface of final deliverable shown on following figures:

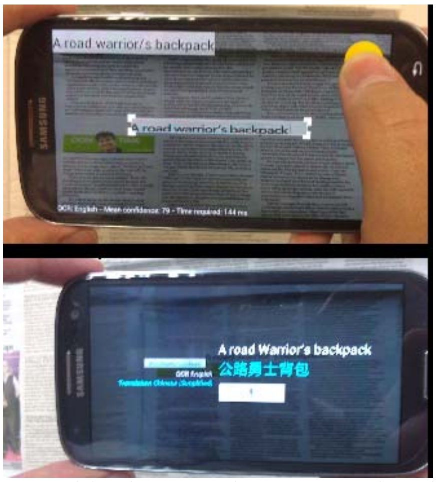

A road warrior's backpack
A road warrior's backpack
OCR: English - Mean confidence: 79 - Time required: 144 ms
A road Warrior's backpack
公路勇士背包
OCR English
Translation Chinese (Simplified)

Fig 8 Recognizing the sentence in newspaper and translation

A sentence from an English newspaper is captured. The preview displayed on the screen and it is translated from English to Chinese (Simplified) as shown in Fig 8.

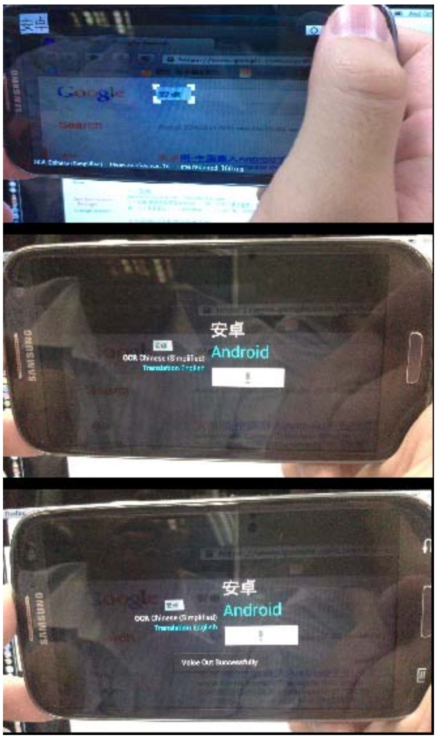


Fig 9 Recognizing Chinese word, translation and voice out

Chinese word is captured and it is translated from Chinese (Simplified) to English. Finally, the application displays "Voice out successfully" notification on the screen if the voice out language is recognized as English and once the user pressed the Voice Out button as shown in Fig 9.

The application was tested with 50 target end users to ensure the functionality, accuracy, performance and quality. 40 respondents were travelers who are from different language background and the remaining 10 were partially impaired people. At the end of the testing, 95% of respondents are happy with the functionalities and accuracy of the application. 5% participants suggested that the application should diversify for more languages. Similarly, 10% respondents suggest integrating more voice out languages. Overall, the current features and quality of the application satisfy the users' needs.

VII. Conclusion

In this paper, an Android application to capture text, translate and voice it out has been designed and developed. Final deliverable has been tested and results obtained are promising which demonstrate that this study successfully addressed the problems discussed in Section I. This study proves that the application is mainly convenient for travelers to reduce the language barrier during their visit to another country which uses different language to portray information. Not to forget that speech out feature is beneficial for visually impaired users to access printed text which may carry significant messages. This application developed as a prototype to obtain users' feedback in the first iteration and it can be further improved and enhanced. In future, it can be upgraded with better OCR engine, translator services or even by multi supported text to speech engine. By doing so, it can significantly improve the performance of the system to ensure a better quality application.

Acknowledgement

The authors wish to thank everyone who provided insight and expertise that greatly assisted this study. Appreciation to university management for the support and resources provided throughout the research.

References

[1] A. Canedo-Rodriguez, S. Kim, J. Kim and Y. Blanco-Fernandez, 'English to Spanish translation of signboard images from mobile phone camera', *IEEE Southeastcon 2009*, 2009.

[2] H. Nakajima, Y. Matsuo, M. Nagata and K. Saito, 'Portable Translator Capable of Recognizing Characters on Signboard and Menu Captured by Built-in Camera', in *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, 2005.

[3] L. Zhifang, L. Bin and G. Xiaopeng, 'Test automation on mobile device', *Proceedings of the 5th Workshop on Automation of Software Test - AST '10*, 2010.

[4] Who.int, 'WHO | Visual impairment and blindness', 2015. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs282/en/. [Accessed: 28- Sep- 2015].

[5] A. Shaik, G. Hossain and M. Yeasin, 'Design, development and performance evaluation of reconfigured mobile Android phone for people who are blind or visually impaired', *Proceedings of the 28th ACM International Conference on Design of Communication - SIGDOC '10*, 2010.

[6] S. Mori, C. Suen and K. Yamamoto, 'Historical review of OCR research and development',*Proceedings of the IEEE*, vol. 80, no. 7, pp. 1029-1058, 1992.

[7] M. Laine and O. Nevalainen, 'A Standalone OCR System for Mobile Cameraphones', *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2006.

[8] K. Bae, K. Kim, Y. Chung and W. Yu, 'Character Recognition System for Cellular Phone with Camera', *29th Annual International Computer Software and Applications Conference (COMPSAC'05)*, 2005.

[9] Hymes, K. & Lewin, J. 'OCR for Mobile Phones', 2008.

[10] L. Chang and S. ZhiYing, 'Robust pre-processing techniques for OCR applications on mobile devices', *Proceedings of the 6th International Conference on Mobile Technology, Application & Systems - Mobility '09*, 2009.

[11] S. Fong, A. Elfaki, M. bin Md Johar and K. Aik, 'Mobile language translator', *2011 Malaysian Conference in Software Engineering*, 2011.

[12] V. Fragoso, S. Gauglitz, S. Zamora, J. Kleban and M. Turk, 'TranslatAR: A mobile augmented reality translator', *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, 2011.

[13] Koga M, R. Mine, T. Kameyama, T. Takahashi, M. Yamazaki and T. Yamaguchi, 'Camera-based Kanji OCR for mobile-phones: practical issues', *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005.

[14] A. Gohr, 'Linux OCR Software Comparison', *Splitbrain.org*, 2015. [Online]. Available: http://www.splitbrain.org/blog/2010-06/15-linux_ocr_software_comparison. [Accessed: 27- Sep- 2015].

[15] G. Erichsen, ' Which Online Translator Is Best? ', *About.com Education*, 2015. [Online]. Available: http://spanish.about.com/od/onlinetranslation/a/online-translation.htm. [Accessed: 29- Sep- 2015].