

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

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.

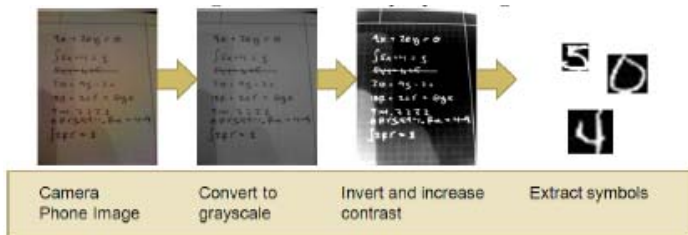


Fig 1 Image enhancement
Source (Hymes & Lewin, 2008)

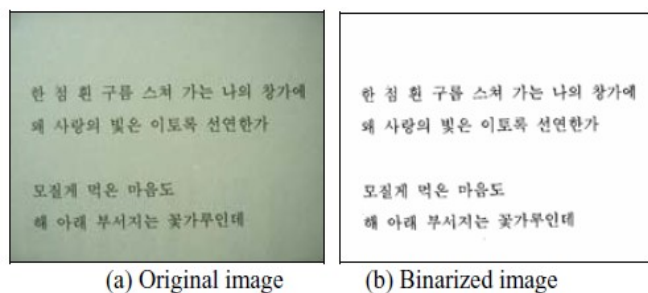


Fig 2 Binarization
Source: (Bae et al., 2005)

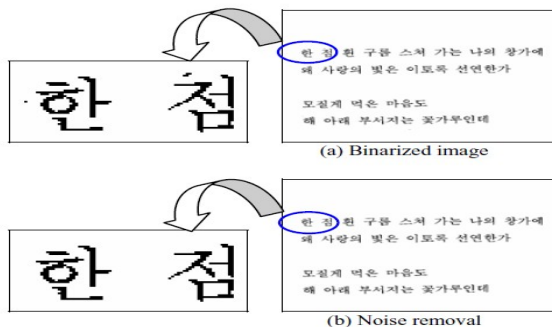


Fig 3 Noise reduction
Source: (Bae et al., 2005)

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.

Download English Version:

<https://daneshyari.com/en/article/4920295>

Download Persian Version:

<https://daneshyari.com/article/4920295>

[Daneshyari.com](https://daneshyari.com)