Optical Character Recognition (OCR) Performance in Server-based Mobile Environment

Teddy Mantoro[1,2], Abdul Muis Sobri[1], Wendi Usino[1]
[1]Postgraduate Program, FTI, University of Budi Luhur, Jl. Raya Ciledug, Jakarta 12260 Indonesia.
[2]Faculty Science and Technology, Universitas Siswa Bangsa International, Jakarta 12780 Indonesia.
teddy@ieee.org, abdulmuiss@yahoo.co.id, wendi.usino@bl.ac.id

***Abstract*—There are several Optical Character Recognition (OCR) mobile applications on the market running on mobile devices, both android and iOS (iPhone, iPad, iPod) platforms. The limitations of mobile device processor hinder the possible execution of computationally intensive applications that need less time of process. This paper proposes a framework of Optical Character Recognition (OCR) on mobile device using server-based processing. Comparison methods proposed by this paper by conducting a series of tests using standalone and server-based OCR on mobile devices, and compare the results of the accuracy and time required for the entire OCR processing. Server-based mobile OCR obtains 5% higher character recognition accuracy than the standalone OCR and its format recognition accuracy is 99.8%. The framework tries to overcome the limitation of mobile device capability process, so the devices can do the computationally intensive application more quickly.**

Keywords-OCR; Optical Character Recognition; mobile device; image processing; accuracy;

I. INTRODUCTION

Modern high-end mobile cameras are capable of image capturing up to 41 megapixels [9], which is usually quite adequate for performing optical character recognition. Having OCR software on a mobile device has obvious benefits such as: text editing of text printed-paper, translation, reproducing printed books from a saved soft copy. However, OCR standalone applications that are installed on a mobile device have limitations such as processing power and lifetime of battery.

For a single page OCR, using standalone OCR mobile application is adequate for performing character recognition. But using it to recognize multiple pages is very inconvenient, because it takes more time to capture an image, process it using OCR application, and compare the results with the installed language to correct mistakes during word detection. All these processes are limited to the mobile device's processor, memory and battery lifetime.

This paper proposes a technique to perform multiple pages of text recognition on a mobile device and take advantage of the cloud server's ability to process heavy work. The results will be compared with OCR results from some OCR applications on market, based on time of process and character detection accuracy. The examination use two models of mobile devices with different specification: HTC Android Phone and Apple iPad 4 to examine the OCR process, a personal computer as a server, and internet connection to communicate between mobile devices and the server. The results of the test shows that multiple pages OCR will take less time of process with the platform described and discussed on the next topic. Our experimental results described that server-based multiple page OCR on mobile devices will dramatically increase the total of character and text formatting accuracy of OCR.

The main contributions of this study are:

- Technique of multiple pages character recognition using standalone and server-based mobile device application.
- Technique of image correction to increase speed process and recognition accuracy

The rest of this paper is organized as follows. The next Section discusses OCR related work and image processing on mobile devices. Section 3 discusses the framework proposed in this paper. Standalone Mobile OCR on Section 4 discusses some applications on the market related mobile devices OCR and its features. Testing on Section 5 discusses about testing OCR and results using OCR standalone applications and server-based OCR. The method of OCR using mobile device and server and the advantages are explained here. Conclusions and future work of this topic discussed on Section 6.

II. RELATED WORK

Discussions about OCR either about the application running on a desktop computer or mobile computer was becoming a topic of interest. There are currently several commercially available OCR systems on the market today such as Abby FineReader and OmniPage, both applications running on desktop computer. OCR Kit, Mobile OCR and Scanthing are running on mobile device. In [1] presented OCR droid, a generic framework for developing OCR-based application on mobile phones. However, this discussion focused on using orientation sensor, embedded high-end camera and digital image processing technique to solve OCR issues related to camera-captured images. In [2] describes a simple OCR system that was implemented in Symbian C++

978-1-4799-2758-6/13 $31.00 © 2013 IEEE
DOI 10.1109/ACSAT.2013.89

to be run on a stock Nokia 6630 camera phone. However, this system does not describe how fast the system performs OCR on single and multiple pages. In [3] proposes a Text Extraction algorithm for the context of language translation of scene text images with mobile phones, which is fast and accurate at the same time. The author claims that the algorithm uses very efficient computations to calculate the Principal Color Components of a previously quantized image. The author also compares the algorithm with other algorithms using commercial OCR, achieving accuracy rates more than 12% higher, and performing two times faster, and the methodology is more robust against common degradations, such as uneven illumination, or blurring. However, no discussions have described the ability of OCR processing on multiple pages.

III. Optical Character Recognition (OCR) Framework for Mobile Device

The framework that is proposed on this discussion is to take an image of multiple printed-papers using a mobile device's camera. The image taken from 10 sheets of printed-paper is removed from a book. The color of text is black, and the paper color is white. The images taken with the best method as described in Section 5. The image is captured sheet per sheet. After the first image is captured, the image is then directly sent to a server. The Server processes the image using OCR application directly and sends the text file back to the mobile device.

The application used to perform real OCR on the server is Abby Fine Reader Express Edition build 8.0.0.4123. At the time of testing, the application was running on the computer Macbook pro 13", 64-bit, 8 MB memory, processor 2.3 GHz Intel Core i5, OSX Version 10.9. On this version, the application is able to convert image files into text files, spreadsheet files, HTML files and PDF files.

Sending process from a mobile device to an OCR server is a background process, occurred after the picture is taken, so the process can carry on using the device to capture the next page, as shown in the architecture in Figure 1.

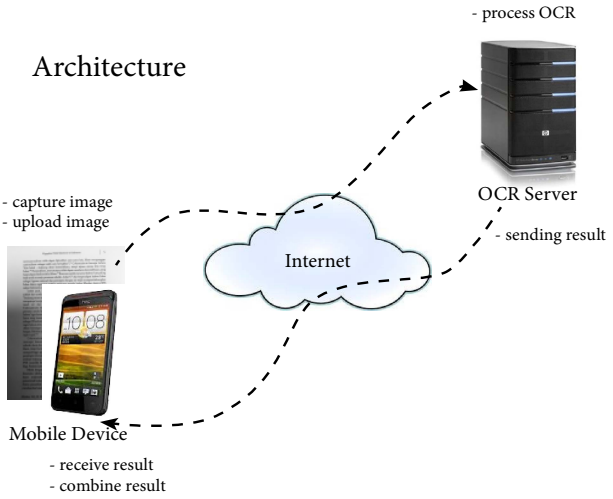


Figure 1. Figure 1. Server-based mobile OCR Framework

After capturing all the images, there is no need to wait a long time to see the result, because when the first was picture taken, the image file was directly sent to the server, and the OCR server processed it directly. So, when picture number 2 was captured, actually the first picture is being sent to the server, and the server processes it directly with the OCR server. Text result is also directly sent back to the mobile device. An application on a mobile device, receives the result from the server, and combines OCR result of page 1 with next page that will be arrive from the OCR server.

Below, the illustration of schematics order of time of architecture method:

TABLE I. Illustration of Schematics Order of Time

Time	Mobile devices			OCR Server	
1	cap page 1	-	-	-	-
2	cap page 2	up page 1	-	-	-
3	cap page 3	up page 2	-	pro page 1	-
4	cap page 4	up page 3	-	pro page 2	sen page 1
5	cap page 5	up page 4	rec page1	pro page 3	sen page 2
6	cap page 6	up page 5	rec page2, cwp	pro page 4	sen page 3
7	cap page 7	up page 6	rec page 3, cwp	pro page 5	sen page 4
8	cap page 8	up page 7	rec page 4, cwp	pro page 6	sen page 5
9	cap page 9	up page 8	rec page 5, cwp	pro page 7	sen page 6
10	cap page 10	up page 9	rec page 6, cwp	pro page 8	sen page 7
11	-	up page 10	rec page 7, cwp	pro page 9	sen page 8
12	-	-	rec page 8, cwp	pro page 10	sen page 9
13	-	-	rec page 9, cwp	-	sen page 10
14	-	-	rec page 10, cwp	-	-
15	-	-	display OCR result	-	-

* cap : capture
up : upload
rec : receive
sen : sending
pro : process
cwp : combine with previous page

The illustration above does not describe the real time process. The real time process could be varied depending on several variables, such as: speed of Internet connection, specifications of both mobile or server device and speed of attached device camera.

IV. Standalone Mobile OCR

Under this section, a few applications on the iPhone, iPad and Android market and the ability to perform text recognition from images taken from the devices' attached camera will be discussed.

A. OCR Kit

OCRKit released for iPhone and iPad. It is a simple application easiest to convert and extract text from photos captured from iPhone and iPad's camera. So, the text can be edited, copied, and shared with other applications, including Mail.

It features Optical Character Recognition performance, which allows the taking of photos from the iPhone and iPad attached camera on the go. Save the text and edit it, if and when it is needed.

This OCR Kit has features as follows:

- Includes editable history of previously recognized texts.
- Allows for editing, copying or pasting the text into third party applications.
- Use an existing photo or take a new one with your built-in camera.
- Designed for iPhone, iPod touch and iPad.

The OCR requires a sharp image with macro focus, and thus works best with the iPhone 3G, iPhone 4/4S and iPad 3. Older iPhones will only yield recognition results if used with a close-up lens case, such as Griffin's Clarifi.

The 5 mega-pixel camera starting with the iPhone 4 and OCRKit can potentially replace an entry-level portable scanner for road warriors working with paper documents abroad.

B. Mobile OCR

Mobile OCR is an OCR application released by Smart Mobile Software. It runs on both android and iOS platform. The OCR processes page by page by taking pictures directly from the attached camera or loading it from image album. This application does not support multiple page scanning yet. If there are 2 pages or more, it will be processed page per page, and the result can be joined manually.

Mobile OCR is compatible with more than 25 languages [4] including: Arabic, English, etc.

C. Scanthing

This application runs on the android platform, and also captures images to process on an OCR system page by page, and does not supported multiple pages of OCR scanning yet.

The features of the applications are:

- Image captured can be rotated and cropped to the exact direction and position.
- Support for 33 languages, just set your default version in the settings.
- Using a Text-To-Speech Engine of your choice (IVONA is recommend) you can play back the text, sentence by sentence directly through Scanthing.
- Scanthing will automatically recognize and extract contact information, such as Telephone Numbers, Web URLs and Email Addresses for easy, one click linking to other features on your phone such as web browser, email client or your contacts database.
- Works with all printed typefaces, but please note it will not be able to read your handwriting.
- Use with Google Translate on your phone to convert from one of the 33 languages supported.
- Fully secured HTTPS connection to the server.
- After Scan, search for your documents by just using a few keywords.
- Share image or extracted text by Email, Dropbox, Evernote, Google Translate and other apps.
- Associate Tags with your documents and then use Tags to find documents fast when you are on the move [5].

V. TESTING AND EVALUATION

The applications testing focused on process speed and accuracy of text recognition compared with original printed text. The original printed-paper that will be captured with the mobile camera is a sheet that was removed from a book. The test uses 10 sheets removed from a white printed book. Every page contains about 2.300 characters.

The device that was used to examine the framework related to OCR is as follows: (1) HTC Desire VC (Mobile Devices, Android 4.0); (2) Apple iPad (Mobile devices, Apple iOs 6.1); and (3) MacBookPro (Desktop Devices, as server).

Figure 2 shows that an Internet connection provide by First Media Corporation (local provider) with an average of 3.79 Mb/s download speed and an average of 380 Kbp/s was used to communicate between the mobile device and the OCR server. The speed of the connection is described below:

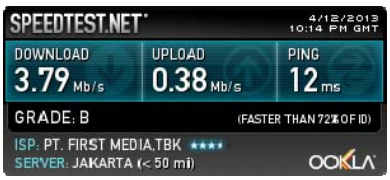


Figure 2. Internet connection speed

In order to measure the accuracy of OCR, this paper adopts the metric proposed by the Information Science Research Institute at UNLV for the Fifth Annual Test of OCR Accuracy [1].

First formula to count text accuracy:

$$\text{Accuracy} = \frac{NC - EC}{NC}$$

NC= Number of character
EC= Error of read character

Second formula to count format accuracy:

$$\text{Accuracy} = \frac{NW - EF}{NW}$$

NW= Number of word
EF= Error of read format

Format accuracy measures format detection such as *italic*, sub/superscript, **bold**, and underline format. In this case, most of these formats are ignored. When the value of format accuracy is 0%, this indicates that no text format was detected by the application.

When performing OCR on a single page, the application uses the directly captured image and continues to the recognition process. But, on multiple pages, images are captured page by page using the camera application, then

begins running the OCR application and uses the saved images to perform the recognition process.

The scanned image or images captured by mobile device camera are often noised that can make it difficult to perform the OCR process [6]. Noised image need to go through a de-noise process for further processing [7].

To get maximum image quality, keep in mind the following things when capturing objects with a mobile device camera:

- Focus. Position the object to focus on the camera to avoid image blur. Detailed images are easier to recognize the characters. Some mobile device cameras are equipped with auto focus feature, it makes it easier to focus on objects.
- Position object tilt. Objects should be placed on a flat surface and the camera is positioned parallel to the object to avoid curved objects.
- Lighting. The lighting is very influential on shooting speed, and the results obtained. When there is less light, the camera will automatically slow the diaphragm aperture. This slow the diaphragm aperture produces a movement when recording an image and the resulting image with shade.

There was a discussion regarding skew angle detection of the picture as well as the solution [2]. This discussion proposed an algorithm to correct the skew angle of a picture that was recorded by a digital camera phone. To correct the angle of the picture, there was discussion about image auto rotation [8].

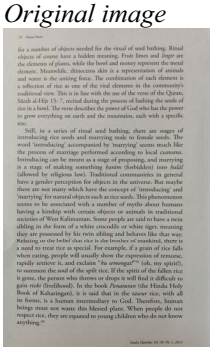
Although some of the above disadvantages can be eliminated with image processing software, but the process will take time and will slow down the OCR process. To save time, should keep in mind the process of scanning images to get maximum results.

Below are the results from the test of each application by format and text accuracy:

A. *OCR Kit*

The test using Apple iPad 4 to examine the application, and using attached camera to capture images. After performing OCR on 10 sheets of text image, the result is described as below:

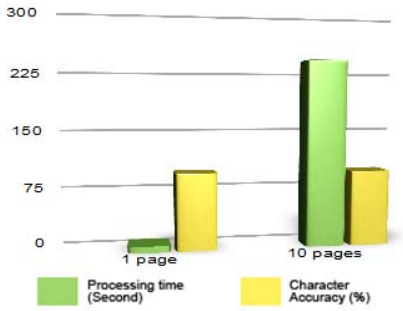
Original image



Result sample

76 l'aiaal A sni n
for a number of objects needed for the ritual of seed bathing. Ritual objects of course have a hidden meaning. Fruit limes and lingir are the elements of plants, while the bowl and money represent the metal element. Meanwhile, rhinoceros skin is a representation of animals and water is the uniting force. The combination of each element is a reflection of rice as one of the vital elements in the community's traditional view. This is in line with the use of the verse of the Quran, Surah al-Hijr 15: 7, recited during the process of bathing the seeds of rice in a bowl. The verse describes the power of God who has the power to grow everything on earth and the mountains, each with a specific size.
Still, in a series of ritual seed bathing, there are stages of introducing rice seeds and marrying male to female seeds. The word 'introducing' accompanied by 'marrying' seems much like the process of marriage performed according to local customs. Introducing can be meant as a stage of proposing, and marrying is a stage of making…….

Using the above formula, the results of the test is described in Figure 3.



Number of pages	Processing time	Accuracy	Format Accuracy
1	8 second	98.69%	0%
10 - capture	20 second		
10 - process	120 second		
Total 10	240 sec/ 2.33 min	98.10%	0%

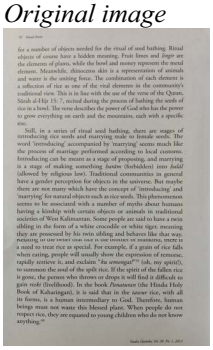
Figure 3. Test Result with OCR Kit

B. *Mobile OCR*

Apple iPad 4 was also used to examine the application, and the image was captured when the sun was shining bright. The result is described as below:

Sample of original and the result:

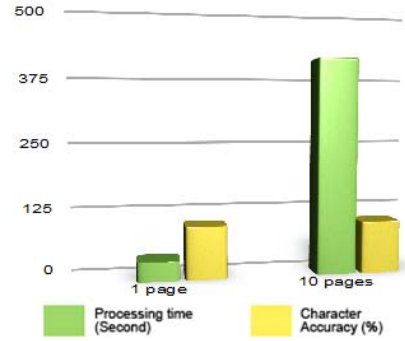
Original image



Result sample

for a number of objects needed for the ritual of seed bathing. Ritual
objects of course have a hidden meaning. Fruit limes and lingir are
the elements of plants, while the bowl and money represent the metal
element. Meanwhile, rhinoceros skin is a representation of animals
and water is the uniting force. The combination of each element is
a reflection of rice as one of the vital elements in the community's
traditional view. This is in line with the use of the verse of the Quran,
Surah al-Hijr 15: 7, recited during the process of bathing the seeds of
rice in a bowl. The verse describes the power of God who has the power
to grow everything on earth and the mountains, each with a specific
size…….

The calculation of accuracy result is as follows:



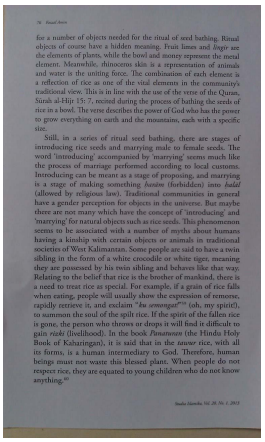
Number of pages	Processing time	Text Accuracy	Format Accuracy
1	37 second	99.92%	0%
10 - capture	20 second		
10 - process	390 second		
Total 10	6.83 minute	99.90%	0%

Figure 4. Test Result with OCR Mobile OCR

C. *Scanthing*

The test result of the application is as follows:

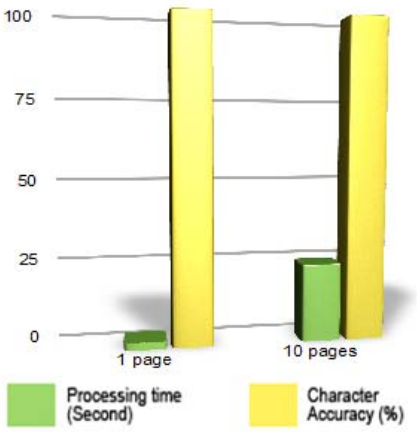
Original image *Result sample*



"6 hiwt At* j/p

for a number of objects needed for the ritual of seed bathing. Ritual
objects of course have a hidden meaning. Fruit limes and lingir are
the elements of plants, while the bowl and money represent the metal
element. Meanwhile, rhinoceros skin is a representation of animals
and water is the uniting force. The combination of each element is
a reflection of rice as one of the vital elements in the community's
traditional view. This is in line with the use of the verse of the Quran,
Surah 15: 7, recited during the process of bathing the seeds of
rice in a bowl. The verse describes the power of God who has the power
to grow everything on earth and the mountains, each with a specific
size.

Still, in a series of ritual seed bathing, there are stages of
introducing rice seeds and marrying male to female seeds. The
word 'introducing* accompanied by marrying1 seems much like
the process of marriage performed according to local customs.
Introducing can be meant as a stage of proposing, and marrying
is a stage of making something hardm fforhidden) into halat
(allowed by religious law). Traditional communities in general …….



Number of pages	Processing time	Text Accuracy	Format Accuracy
1	2.4 minute	99.91%	0%
10	24 minute	99.89%	0%

Figure 5. Test Result with OCR Scanthing

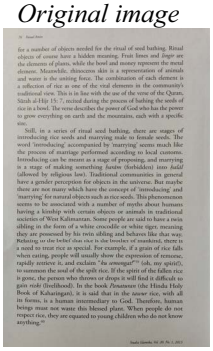
Time of process of this application is longer than the 2 applications above, but character accuracy is also better, as shown on the result below:

Like two application above, this also does not support capturing character format. The captured character also displayed on plain format, there is no italic, bold or underline format.

D. *Proposed framework*

Sample of text recognized with the method as follows:

Original image *Result sample*



for a number of objects needed for the ritual of seed bathing. Ritual objects of course have a hidden meaning. Fruit limes and *lingir* are the elements of plants, while the bowl and money represent the metal element. Meanwhile, rhinoceros skin is a representation of animals and water is the uniting force. The combination of each element is a reflection of rice as one of the vital elements in the community's traditional view. This is in line with the use of the verse of the Quran, Surah al-Hijr 15: 7, recited during the process of bathing the seeds of rice in a bowl. The verse describes the power of God who has the power to grow everything on earth and the mountains, each with a specific size.

Still, in a series of ritual seed bathing, there are stages of introducing rice seeds and marrying male to female seeds. The word 'introducing' accompanied by 'marrying' seems much like the process of marriage performed according to local customs. Introducing can be meant as a stage of proposing, and marrying is a stage of making something *hardm* (forbidden) into *haldl* (allowed by religious law). Traditional communities in general have a gender perception for objects in the universe. But maybe there are not many which have the concept of 'introducing' and 'marrying' for natural objects such as rice seeds. This phenomenon seems to be associated with a number of myths about humans having a kinship with certain objects or animals in traditional societies of West Kalimantan. Some people are said to have a twin sibling in the form of a white crocodile or white tiger, meaning they are possessed by his twin sibling and behaves like that way. Relating to the belief that rice is the brother of mankind, there is a need to treat rice as special. For example, if a grain of rice falls when eating, people will usually show the expression of remorse, rapidly retrieve it, and exclaim *"ku semongat!"*[59] (oh, my spirit!), to summon the soul of the spilt rice. If the spirit of the fallen rice is gone, the person who throws or drops it will find it difficult to gain *rizki* (livelihood). In the book *Panaturan* (the Hindu Holy Book of Kaharingan), it is said that in the *tawur* rice, with all its forms, is a human intermediary to God. Therefore, human beings must not waste this blessed plant. When people do not respect rice, they are equated to young children who do not know anything.[60]

Studia Islamika, Vol 20, No. 1, 2013

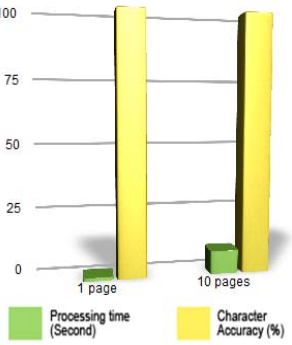
Estimation of real time of process is described follows:

No	Per 10 second	Scan	Upload	OCR	Sending	Finish
1	Minute 1		Page 1			
2						
3						
4						
5						
6			Page 2	Page 1		
1	Minute 2					
2					Page 1	
3						
4						
5			Page 3	Page 2		
6						
1	Minute 3				Page 2	
2						
3						
4			Page 4	Page 3		
5						
6					Page 3	
1	Minute 4					
2						
3			Page 5	Page 4		
4						
5					Page 4	
6						
1	Minute 5					
2			Page 6	Page 5		
3						
4					Page 5	
5						
6						
1	Minute 6		Page 7	Page 6		
2						
3					Page 6	
4						
5						
6			Page 8	Page 7		
1	Minute 7					
2					Page 7	
3						
4						
5			Page 9	Page 8		
6						
1	Minute 8				Page 8	
2						
3						
4			Page 10	Page 9		
5						
6					Page 9	
1	Minute 9					
2						
3				Page 10		
4						
5					Page 10	
6						

TABLE II. ESTIMATION OF REAL TIME OF SERVER-BASED OCR PROCESS

Table 2 shows the Total time of process of 10 sheets printed page using our proposed architecture is estimated by 9 minutes.

Figure 6 shows the result of the test with server-based mobile OCR.



Number of pages	Processing time	Text Accuracy	Format Accuracy
1	1.3 minute	99.99%	99.8%
10	9 minute	99.98%	99.8%

Figure 6. Test Result with server-based mobile OCR

Figure 7 presents the comparasion of four methods including OCR kit, Mobile OCR, Scanthing and cloud OCR based on time of processing, character recognition accuracy and format accuracy. This experiment also shows that Mobile OCR method delivers a better results in both character and text format.

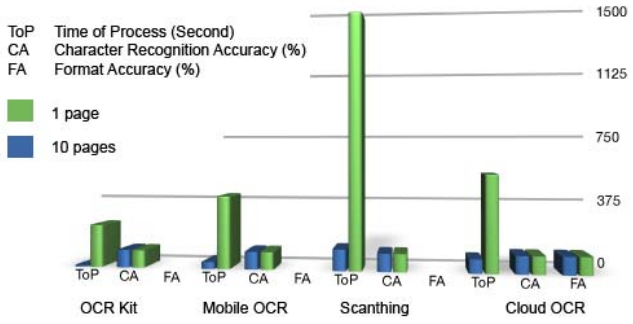


Figure 7. Result Comparison

OCR accuracy either format or character accuracy is more complex. The test did not recognize different formats using this method. All character formats is captured by the best way, such as italic, bold and sub script.

VI. CONCLUSION

This paper presents the method and framework for performing multipage OCR on mobile devices. The discussion focused on how to produce a format and character accuracy from OCR application and also its time of process.

The format and character accuracy is very useful for example in the book reprints. In terms of cost, using OCR with good accuracy would also be more efficient. For example, the reprint of 100 pages of printed book, the typing cost is around USD 20.00 and can be done in about 4 days [10]. Using this method, it will only takes 90 minutes to complete the digitization of the 100 page book, and can be done while drinking coffee at Starbucks.

Although the time required is longer than using standalone OCR application, our method delivers a better results in both character and text format.

The problem on longer time required, however can be solved with reducing captured image file before uploading to server.

This paper also has not yet built the application that can run all the processes automatically, starting from the capturing of images on mobile device until combining the text resulting from the OCR server.

REFERENCES

[1] A. Joshi, M. Zhang, R. Kadmawala, K. Dantu, S. Poduri and G. S. Sukhatme, *OCRdroid: a Framework to Digitize Text Using Mobile Phone*, University of Southern California, Los Angeles, CA 90089, USA.

[2] M. Laine and O.S. Nevalinen, *A Standalone OCR System for Mobile Cameraphones,* The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06), 2006.

[3] Canedo-Rodríguez, Adrián and Kim, Jung Hyoun, etc, Efficient Text Extraction Algorithm Using Color Clustering for Language Translation in Mobile Phone, *Journal of Signal and Information Processing, vol 3,pp. 228-237,* 2012.

[4] Smart Mobile Software – Mobile OCR. Internet: http://www.smartmobilesoftware.com/mobile-ocr.html [Accessed: April 13, 2013].

[5] Scanthing. Internet: http://www.scanthing.com [Accessed: October 11, 2013].

[6] K. S. Bae, K. K. Kim0, Y. G. Chung, W. P. Yu, *Character Recognition System for Cellular Phone with Camera,* Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC'05), 2005.

[7] Priya Sharma and Randhir Singh, *Survey and Classification of Character Recognition System*, International Journal of Engineering Trends and Technology- Volume 4 Issue 3- 2013.

[8] Wojciech Bieniecki, Szymon Grabowski and Wojciech Rozenberg, *Image Preprocessing for Improving OCR Accuracy*, MEMSTECH'2007, May 23-26, 2007, Lviv-Polyana, UKRAINE.

[9] Jessica Dolcourt, *Camera megapixels: Why more isn't always better (Smartphones Unlocked)*. Internet: http://www.cnet.com/8301-17918_1-57423240-85/camera-megapixels-why-more-isnt-always-better-smartphones-unlocked/ [Accessed: October 21, 2013].

[10] Pengetikan Murah dan Berkualitas, Daftar Tarif Jasa Ketik. Internet: http://jasaketikjogja.blogspot.com/2013/04/daftar-tarif-jasa-ketik.html [Accessed: November 28, 2013].