MOBILE APPLICATION FOR TEXT RECOGNITION (OCR)

Ondrej Krejcar
Department of Information Technologies, Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, Hradec Kralove, 500 03, Czech Republic

Keywords: OCR, MDA, Built-in Camera, Image Processing, Virtual Keyboard.

Abstract: The aim of this project is to develop a simple application for capturing digital photography and its subsequent processing of OCR technologies. The application operates as an alternative for manual rewriting of parts of text from printed data using the keyboard on the screen of the device. It is mainly targeted at short text sections, such as hypertext references and addresses which are hard to be copied and at more excessive texts which would take long to copy. The emphasis is laid on the simple, fast and intuitive manipulation. The end platform is represented by PDA device, more precisely MDA which is based on Windows Mobile operating system.

1 INTRODUCTION

The Smart Phones, such as cell phones and PDA, especially MDA are the phenomenon nowadays (Mikulecky, 2009). The number of cell phone users over 16 years old in the Czech Republic for the year 2009 climbed up to 91%. For the population in the age group from 16 to 54 years the number is equal to 98%. A great boom in the field of cell phones and their performance was caused by using the operating system, such as Symbian, Android or Windows Mobile. Many of these devices use big colourful displays with touch screen and with a connection to 32bit processor on which the OS runs and therefore may be considered a PDA. Moreover, the GSM module is usually integrated within the standard PDA together with WiFi module (Brida et al., 2010). The result is the incorporation of cell phones and PDA. The name communicator or MDA is sometimes used for these devices. The usage of efficient standardized 32bit processors which support OS makes it possible to develop for these platforms even complicated applications for computation.

The primary input system of these devices is the keyboard in a classic "physical" design or in the form of virtual keyboard on a display in the casesi of touch screen, also called the on-screen keyboard. These types of keyboards provide a comfortable method of information inscription. Nevertheless, the typing is approx. 4 times slower than in the case of computer keyboards. However, this typing speed may be insufficient if using the PDA as a tool for fast information recording (e.g. business card copying or copying parts of text). Most commonly integrated CCD chip enables the photographing or recording of a video-sequence. Therefore, it is a convenient and instant way of capturing information. Moreover, if this information is time-limited (e.g. it must return within certain time limit or it is only displayed for short time period) then it is the only method.

Nevertheless, sometimes there is a need to further process this captured text (Liou and Cheng, 2007). The text retyping from these images is lengthy. Furthermore, if it is necessary to retype using the PDA it should be accounted for switching often between an application with displayed image and the text editor. Every such switching "produce" time delay which resulted in money losses (Lefley et al., 2004).

In these cases the usage of OCR technology is the best solution. The first mobile application OCR was released to the market already in 2002 (Tariq, Nauman and Naru, 2010). Certain factors complicate the usage of OCR in PDA which mostly originated from the low quality of copies acquired by CCM. Finally, it is necessary to mention that the common source for OCR application is a scanner.

A PDA which is supplied by OCR has many options in a way of utilization. If the user notices an URL address in some printed document, he can look at it by taking a picture which consequently opens

Krejcar O..
MOBILE APPLICATION FOR TEXT RECOGNITION (OCR).
DOI: 10.5220/0003947002620265
In *Proceedings of the 2nd International Conference on Pervasive Embedded Computing and Communication Systems* (PECCS-2012), pages 262-265
ISBN: 978-989-8565-00-6

the link in a browser. After this picture the business card with user's data is saved into contacts, etc.

2 PROBLEM DEFINITION

The accuracy of OCR depends mainly on the quality of recognizable under layer. The most common usage of OCR on scanned documents achieves quite satisfactory results. Using of OCR in PDA with CCM as a data source recognizer carries number of problems (Hymes and Lewin, 2008), especially:

- Relatively low computational performance
- Low quality of images for OCR
- Tilt (perspective deformation), skew and rotation
- Incoherent lighting and shadows

Mainly due to these complications is OCR in PDA limited to just small parts of text. Therefore, the insufficient quality of acquired images is compensated by the size proportion of symbols in the overall resolution. The existing applications may be good examples, because they are usually specialized on business card scanning.

2.1 Existing Applications

2.1.1 Nokia Multiscanner

Nokia Multiscanner is an application designed for cell phones with Symbian system and spread as a freeware. The application supports picture taking and consequently sending it through MMS, Bluetooth or via infrared (Bodnarova et al., 2010). It is possible to transfer the image into a text and save it and at the same time the selection of certain area can be made by dragging. Another possibility is to send the image for business card recognition. This option automatically recognizes contact details on the business card and fills in the details for adding a new contact. The OCR engine supports post-processing on the basis of language dictionaries, including the Czech language.

2.1.2 CameraDictionary OCR for Moto

It is an application for cell phones with Android, Symbian and Windows Mobile systems. It operates on the basis of recorded text recognition and its immediate translation to another language. Even though, this recorded language is available in Chinese or English, the translation is extended by couple of other languages. Furthermore, it enables the text recording with consequent signing of the translated text or so called "Video" regime during which the cursor appears on the screen. The text below the cursor is immediately translated. However, the main disadvantages are the price and the necessity of internet connection when used.

2.1.3 CamCard - Business Card Reader

CamCard is an application specialized on reading business cards. It is targeted at cell phones which run on OS Android, iOS, or Windows Mobile and BlackBerry phones. The main disadvantage is the narrow specialization on business cards and its price.

2.1.4 Babel Reader-LE

Babel Reader-LE is a particular version of Babel Reader for Windows Mobile distributed as a freeware. It enables capturing of an image and subsequent storing of this image in a form of text. Babel Reader-LE is a very simple application. Moreover, it is possible to adjust the captured image before the actual recognition e.g. by background noise removal.

2.2 Conclusion

Nokia Multiscanner is the closest application to the one being developed. However, it is designed only for OS Symbian. CameraDictionary OCR and CamCard are commercial applications which are very specialized and not free. Finally, the last mentioned application called Babel Reader was only invented for text recognition. The selection of these applications with OCR for cell phones is significantly limited and the broader application with OCR which would work as an alternative for a virtual keyboard is still missing. The newly developed application which is described in this article is supposed to fill in these blanks.

2.3 Selection of OCR Engine

Due to the extent of this application, it is planned to use the existing OCR engine. Following types of engines were chosen as the most suitable:

- **Tesseract OCR** – OCR Engine developed by HP Company in since 1985 until 1995. Nowadays, it is being improved by Google. It is offered in C/C++ language.
- **Ocrad** – another open-source OCR engine. One of his main advantages is mainly an automatic transformation of an input image. It does not accomplish post-processing on the basis of language dictionaries. It is written in

C/C++ language.

- **Puma.NET** – an engine for implementation in C# projects with .NET framework.
- **ABBYY Mobile OCR Engine** – a commercial engine used here just for comparison of results.

The Greek letter ω is going to represent the number of symbols in a reference text and ω_{err} is the number of errors (substituted, missing symbols or additional symbols). Then the accuracy of match γ_{acc} is defined as:

$$\gamma_{acc} = \frac{\omega - \omega_{err}}{\omega} 100 \quad [\%]$$

In order to identify the accuracy of recognition, the reference text [Fig. 1] was us\ ed.

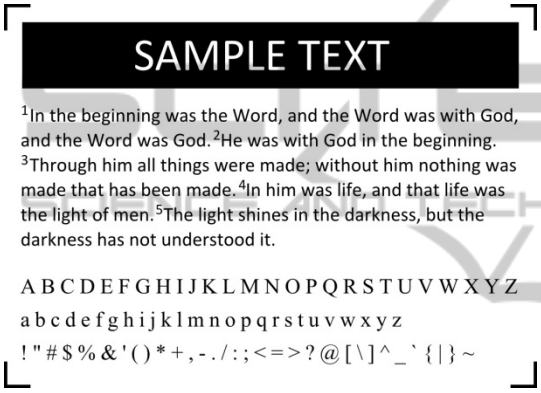

Figure 1: Reference image.

Furthermore, this sample was photographed by Canon PowerShot S3 IS and MDA HTC Touch 2. Consequently, this sample was transferred back to the text form using the above mentioned OCR and compared to the reference text. The accuracy of the match is expressed in percentages in [Tab. 1].

Table 1: Comparison of OCR engine accuracy.

OCR	Original	Canon	HTC
Tesseract	89,78 %	94,72 %	85,25 %
Ocrad	93,30 %	92,71 %	74,36 %
Puma.NET	92,05 %	90,41 %	25,55 %
ABBYY	95,96 %	94,41 %	87,77 %

The comparison shows that the most exact engine is ABBYY Mobile OCR Engine. At the same time it may be noticed that the decreasing quality of sample results in a gradual increase in number of errors. The most significant is the rapid increase of errors when using the Puma.NET engine. In this case, the application was able to correctly recognize approximately one quarter of the text from an image taken by HTC Touch 2. On the other hand, the least sensitive engine considering the quality is Tesseract OCR. Even though, the assumptions were different, Ocrad was proven to be very precise.

3 IMPLEMENTATION

The programming language C# with a connection to the developing environment Microsoft Visual Studio 2008 was designed for the development of the described application. Microsoft Windows Mobile 5.0 runs as the end platform. Therefore, the application should function well on a PDA with this OS or a higher type of OS (Zelenka, 2009).

3.1 Ouciary Application

The application is practically created in form of a guide. After the initiation user is able to turn on the camera and capture the recognized image or simply choose from a file. This is displayed on the following images of the application [Fig. 2].

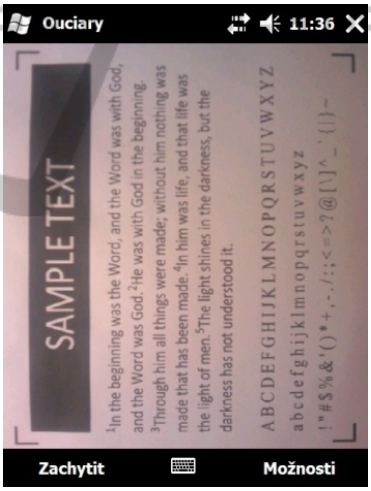

Figure 2: Application after the initiation and text capturing.

Consequently, the image is modified. According to the settings, the normalization, automatic rotation and saturation removal take place.

Furthermore, there is the area selection screen for recognition. Here it is possible to rotate the image manually and choose an area for recognition. Character recognition (described above) is very helpful during the text selection if this function is allowed. Moreover, this screen might be absolutely left out (again according to the settings). The selection happens by dragging ("rectangle drawing"). If it is necessary to cancel the whole selection it is enough to press anywhere on the image. If no area is chosen, the application

automatically calculates with the whole image.

After the selection of the area it is possible to establish individual recognition process. The progress of recognition is shown here. After the termination of recognition process the application automatically moves on to the form for storage.

The resulting text can be seen here in a textbox and at the same time may be saved into a file or Windows mailbox.

As it was already mentioned, there is a space here for adding the functionality in a form of automatic events in relation to recognized text, eventually to a "templates" usage for contact creation according to a business card etc.

The application was tested continuously. The comparison of Tesseract with other OCR engines can be found in the table [Table 1].

4 CONCLUSIONS

The output of this work is OCR for devices with Windows Mobile system. OCR on this platform is capable of speeding up the work when there is no requirement for manual transfer of text from an image. Its usage is mainly connected to integrated CCM for faster business card capturing or other short texts.

This solution may significantly speed up the work after the completion with automatic actions on the basis of recognized text. An example might be URL address capturing and consequent its display in the browser which is much faster than rewriting the address manually (especially in case of long and complicated addresses) (Tucnik, 2010). However, it might be used in the case of business cards and practically every other printed material which need to be rewritten. The mobile devices are always nearby and therefore this method brings instant capturing of printed texts.

ACKNOWLEDGEMENTS

This work was supported by „SMEW – Smart Environments at Workplaces", Grant Agency of the Czech Republic, GACR P403/10/1310. We also acknowledge support from student Ales Kurecka in development of testing application and in several technical problems they grown during development phase.

REFERENCES

Kobryn, C., 2000, Modeling components and frameworks with UML, In *COMMUNICATIONS OF THE ACM*, Volume: 43 Issue: 10 Pages: 31-38 DOI: 10.1145/352183.352199

Bodnarova, A., Fidler, T., Gavalec, M., 2010. Flow control in data communication networks using max-plus approach, In *28th International Conference on Mathematical Methods in Economics*, pp. 61-66.

Brida, P., Machaj, J., Benikovsky J., Duha, J., 2010. An Experimental Evaluation of AGA Algorithm for RSS Positioning in GSM Networks, In *Elektronika ir Elektrotechnika*, No. 8(104), pp. 113-118.

Hymes, K., Lewin, J., 2008. OCR for Mobile Phones. Stanford University

Labza, Z., Penhaker, M., Augustynek, M., Korpas, D., 2010. Verification of Set Up Dual-Chamber Pacemaker Electrical Parameters. In *2nd International Conference on Telecom Technology and Applications*, March 19-21, 2010, Bali Island, Indonesia, Volume 2, NJ. IEEE Conference Publishing Services, p. 168–172.

Lefley, F., Wharton, F., Hajek, L., Hynek, J., Janecek, V., 2004. Manufacturing investments in the Czech Republic: An international comparison, In *International Journal of Production Economics*, Volume: 88 Issue: 1 Pages: 1-14 DOI: 10.1016/S0925.5273(03)00129-4

Liou, CY., Cheng, WC., 2007. Manifold construction by local neighborhood preservation, In *Springer LNCS, Volume 4985,* pp. 683-692.

Mikulecky P., 2009. Remarks on Ubiquitous Intelligent Supportive Spaces, In *15th American Conference on Applied Mathematics/International Conference on Computational and Information Science*, Univ Houston, Houston, TX, pp. 523-528.

Popek, G., Katarzyniak R., 2008. Measuring similarity of observations made by artificial cognitive agents, In *LNAI*, Vol. 4953, pp. 693-702.

Tariq, J. Nauman, U. Naru, M.U., 2010. α-Soft: An English Language OCR. In *Second International Conference on Computer Engineering and Applications (ICCEA 2010)*, IEEE Xplore, DOI 10.1109/ICCEA.2010.112

Tucnik P.,2010. Optimization of Automated Trading System's Interaction with Market Environment, In *9th International Conference on Business Informatics Research,* Univ. Rostock, Rostock, Germany, LNBI, Vol. 64, pp. 55-61.

Zelenka, J., 2009. Information and Communication technologies in tourism – influence, dynamics, trends, In *E & M EKONOMIE A MANAGEMENT*, Vol. 12, Issue 1, pp. 123-132