SYPOLE: A MOBILE ASSISTANT FOR THE BLIND

Vincent Gaudissart, Silvio Ferreira, Céline Thillou, Bernard Gosselin

Laboratoire de Théorie des Circuits et de Traitement du Signal, Faculté Polytechnique de Mons
Bâtiment Multitel – Parc Initialis, 1, Avenue Copernic, 7000, Mons, Belgium
phone: + (32) 65 37 47 61, fax: + (32) 65 37 47 29: gaudissart@tcts.fpms.ac.be
web: http://tcts.fpms.ac.be/projects/sypole/

ABSTRACT

This paper describes an embedded device designed to provide textual information to the blind or visually impaired. For them, all that information, which exists in the daily life as banknotes, schedule of train, books, postal letters, is not easily accessible. The main aim of this system is to build an automatic text reading assistant which combines small-size, mobility and low cost price. Moreover, to be as efficient as possible, a specific human-machine interface has been created and a lot of additional modules have been developed to answer blind people requests. The goal of this system is to improve their autonomy by the use of a mobile device anywhere and anytime.

1. INTRODUCTION

To take a jar in the refrigerator without knowing if it is mustard or jam. To get dressed without being certain that the colour of your sweater is harmonized with your trousers. To catch and to listen to more and more CDs until you finally found the one you wanted to listen to. Not to be certain of the value of a banknote that a shop assistant holds out to you, because the tactile reference marks are erased... Misfortunes of this kind are daily encountered by several partially-sighted persons. Their family circle can assist them, but they are not always there. Above all, they want to gain a maximum autonomy like as all disabled people.

Currently, the only technical existing solution they can use to access to textual information is to couple a scanner with a computer, an optical character recognition (OCR) software and speech synthesis. But those systems are cumbersome and not handy. Moreover, textual information is everywhere, not only in the user's living room, and can exist under different supports such as schedules of trains or labels in stores. It is not possible to put them on a table and therefore to use those existing devices. A mobile, handheld device is thus one of the important keys which will enable the independence of blind people.

From another point of view, some companies produce handheld assistants designed for the blind [1,2]. These systems use technologies developed for PDAs (Personal Digital Assistants), but the touch screen (the main input and output device on palm-sized device), unusable by the blind, is replaced by a set of buttons and Braille cells. These modifications have a cost, and as many other systems designed for disabled people, such systems are very expensive. These devices have the same functionality that common digital assistant, such as agenda, mail reader, address book, etc. but none have functionality to transform textual information into sound.

The main challenge of the SYPOLE project is to find a solution to all the problems described above, and many others, by the realization of a prototype device, which will be portable, autonomous, small-sized and easy to use, especially designed for these people. As well as providing all the functions present on a PDA, this device will be able to recognize text and coloured forms, and to auto-generate a speech signal that reads the detected text or describes an object.

Three key technologies are mainly necessary: text and object detection, optical character recognition and speech synthesis. All these technologies are designed and developed at the TCTS Laboratory at Faculté Polytechnique de Mons [3,4,5,6]

This paper is organized as followed. Section 2 discusses some of the challenges of developing a PDA-based reading assistant for the blind. Section 3 details the choice we made for the realization of the application kernel and the human-machine interface. In Section 4, we describe the applications currently available on our prototype device and their use context. Next, we give more information about the prototype that will be demonstrated and conclude the paper.

2. IMAGE PROCESSING ON A PDA

PDAs have become increasingly popular during the last few years. The main benefit of using a PDA is combining small size, lightweight, computational resources and low cost price. It can be easily programmed and does not require heavy and costly hardware development. However, we have to take into account a lot of limitations to produce an efficient system on a PDA.

A PDA-based reading system does not only share common challenges that the previous OCR systems have met, but also some particular problems. Commercial OCR softwares perform well on "clean" documents, but they fail under unconstrained conditions, or need the user to select the kind of documents, for example forms or letters. The challenge is at different levels of character processing. Firstly, the document image is degraded by taking a picture with a low quality camera. We have tested most of the solutions available on the market with various and hazardous results. The resolution of cameras for Pocket PC is limited to 1.3

Megapixel and the image quality is altered by JPEG compression and by poor or uneven lighting conditions. These problems are very limitative for character recognition. Moreover, these cameras do not have auto focus. In our system, pictures must be taken at a distance of 20-30 cm and the new cameras integrated in the PDA cannot be used without deep hardware modifications because their focus is fixed at infinite. Currently, we are searching for additional lens that can adapt the focus to reach our needs.

Secondly, our system must be free-context, like scene images and free-layout with a wide diversity of characters or text layouts.

Another limitation to take into account when developing algorithms comes from the limited resources of PDA. Indeed, current PDAs are much less powerful than desktop computers in terms of computational power and memory size. In addition, these do not have hardware supporting floating point computation, and heavy algorithms run much slower.

And last but not least, as the user is a blind or a visually impaired person, he cannot use the touch screen of the PDA for reading and entering data. Other modalities must be used or adapted to give a total access of this system to his handicap.

We will address these challenges in both algorithm development and system implementation.

3. SYSTEM DESIGN

The main innovative functionality of our system is the automatic text reading. However, the system needs to be modular, so other functionalities can be easily added to answer users request. Indeed, we have adopted a user-centred design in close relationship with low vision people [7]. These potential users give continuously their evaluations and constructive critics, in order to modify what is considered as ineffective and to submit new ideas and new needs.

Figure 1 gives an overview of the system. Users interact with a dedicated human-machine interface, specifically created for their disabilities. For camera-based applications, the image taken by the embedded camera is sent to the image processing module. When the desired outputs are available, the text-to-speech module synthesizes the information to the user. The principle is the same for all other modules: a request is made by the user, the interactive interface routes the data through the appropriate module and then outputs the information with the TTS Engine.

The main problem encountered when developing this human-machine interface was to allow data introduction by the visually impaired people. The small number of buttons and the PDA's touch screen make their use almost impossible. The idea consists in adding new buttons on the device, without any hardware modification (Figure 2.). For that, we add an overlay on the touch screen: this way turns its drawback into advantage. Upon pressing a key on this overlay, the user applies pressure on a specific part of the display, and that information can be understood by our application.

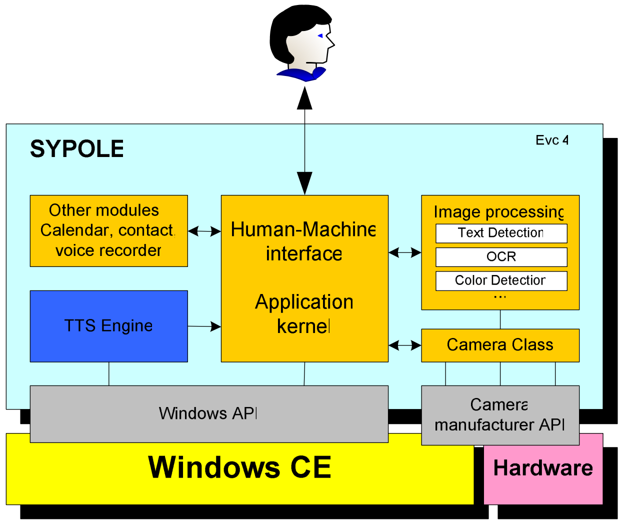

Figure 1. System architecture

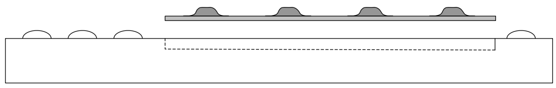

Figure 2. Overlay for PDA touch screen

Using speech synthesis combined with this virtual keyboard provides the necessary key to create an efficient human-machine interface for the blind.

4. APPLICATIONS

The communication with the user is based on natural language. The navigation throughout the interface is inspired by voice server such as ones found in GSM mailboxes. When entering in a new menu, the TTS engine synthesises an enumeration of the available sub-functions or a small tutorial. This navigation is easy to understand, intuitive and well suited for blind users. The vocal interface can also describe the system status and information about the current function on demand. So, the blind user can know what the PDA is doing anytime.

Originally, the only objective of the Sypole project was to produce an automatic and generic text reading device. But the project has been redefined towards a multifunctional assistant and others modules were born to answer user requests as best as possible.

4.1 Automatic reading machine

For example, we can cite the blur due to the camera sensor but the motion during the acquisition time with the hand shaking as well. Different lighting conditions can occur in daily life (interior or not, day time, weather…) and the very large diversity of use situations of Sypole must also be taken into account.

Due to main degradations and limited resources described in Section 2, a mobile text reading system becomes an

interesting challenge to solve. To be automatic and as versatile as possible, several key issues and acquisition conditions must be clearly thought to be then handled.
Firstly, text areas need to be detected to find relevant information to recognise. By “relevant”, we mean especially the compromise between quality, computation time and easiness of listening. Sypole is not designed to detect and recognize all available text in an image when the amount of textual information is large. Text is detected (see [5] for more details) and a quick layout of the document is performed to read automatically the main information in the image such as titles for example. The aim is twofold: to be able to give an understandable answer (smaller the text is, more errors do occur) to visually impaired quickly and to avoid to listen to the answer during a large amount of time.
Secondly, after text areas detection and information sorting, text extraction can occur. Colour information is used in this step to be more robust with respect to the wide heterogeneousness of backgrounds. Technical details are given in [3].
Thirdly, in order to reach again the compromise quality-computation time which is a very strong focus in an embedded application, segmentation into individual characters is then performed [4]. To recognise degraded documents, some techniques, experienced with a scanner-based acquisition, use word recognition without segmentation. With the number of handled fonts, these methods are not efficient here and moreover very computationally expensive.
Finally home-made character recognition is applied to convert textual information from image pixels to ASCII data to be read by a speech synthesizer.

4.2 Objects recognition

This specific request has been received during the preliminary step of study of their major needs. Even if blind people are generally used to recognising objects by touch, this sense does not resolve all their problems to find a desired object. Daily life objects such as CDs, bottles, books, tinned tomatoes or tinned beans are not easily discernible (in fact all similar objects by touch). We have developed an image processing application based on stick-on labels to remedy this problem. The idea is to record a vocal message for each labelled object to keep in memory (recorded only one time when the book is bought for example). Then, each time the user wants to distinguish one of them, he takes a picture, the device identifies the label ID and outputs the correct message. The pattern of the label (a “big” barcode) was designed to be highly detectable by our system and the label ID is recognised by OCR technology. Figure 3.a illustrates the pattern of these labels.

4.3 Colour recognition

This demand is mainly used to help them to dress up. Sorrowfully, we are technically limited by the reconstitution of the colours of the image sensor. As the illumination is unknown, a same object can have different coloured aspects in the image according to the moment of the day when the picture is taken, the orientation of the source of illumination, etc. We are trying to fix the illumination conditions (by adding an adapted flash system) in order to improve the efficiency of this application.

4.4 Banknotes recognition

This euro banknotes recognition application could be treated with the generic reading machine but with a lower efficiency. Indeed in this case we have a priori information about the position of the text region of interest (always in the same zone for all banknotes) and also about the text we want to recognize (only numbers of five, ten, twenty and so on).
By using one-dimensional signals (image profiles) the system scans the image firstly vertically and then horizontally to find the right zone to apply OCR recognition. Figure 3.b shows a sample of images used in this application.

(a) (b)

Figure 3. (a) Sample image of object recognition application (b) Sample image of euro banknotes recognition application

4.5 Training system

We have just presented several application using the image processing capacities of our system. However, there exists a common problem for all these applications when the blind have to take a picture. Indeed, most of them have never used a camera before, hence they have important problems to centre, estimate the distance between the apparatus and the object to be analyzed, and do not have any idea of the ambient luminosity.
To enable these people to use efficiently all the image processing functions described above, we have included a training module: a speaking test card.
Hence, the user will be able to take correct pictures in an autonomous way. For each snapshot, this system answers by indicating the quality of the sight: "the image is fuzzy, bring the camera closer to the object"; "slightly swivel the PDA clockwise"; "orientation and illumination are correct".

4.6 Personal assistant

In order to respond the users request at best, other basic assistive modules have been developed in order to provide a complete digital assistant. Amongst them, we can find a calendar, an address book, a calculator, a news reader…

5. PROTOTYPE

The current prototype uses a Fujitsu-Siemens Pocket Loox 720, which includes an Intel XScale PXA 272 (520 Mhz) processor. Among the many available Pocket PCs, we have selected this model because it combines high performance and wireless communication technologies, and has got built-in 1.3MegaPixels colour camera. Unfortunately, the API of the integrated camera has not been released and for reasons previously described, we had to plug a HP Photosmart camera. Moreover, one goal was to design a system that could be used on different platforms, hence the system can be installed on nearby all off-the-shelf PDAs equipped with a camera.

All algorithms are developed using Microsoft embedded development tools on a desktop computer and then downloaded to the pocket PC.

Figure 4. Prototype using the overlay

6. CONLUSIONS

We have presented a complete assistant with unique functionalities, specially designed to visually impaired people. Based on recent scientific breakthrough about character recognition and speech synthesis and using recent hardware evolutions and powerful data processing, this device can assist visually handicapped users to increase their environmental awareness and improve their autonomy and social integration.

A key idea of our assistant is to be modular to continuously integrate new image processing technologies we develop, but also third-party technologies, such as GPS positioning, other input/output modalities, …

Our aim is to build the most complete and adapted talking assistant for the blind.

Once the first results appeared, complete versions have been given to the blind people to test the efficiency, stability, rapidity and usability of the system. These future users have been chosen by the CRETH (Namur, Belgium), one of our partners. A precise study of satisfaction for the overall system is currently in progress. Detailed results are still not known, but first opinions are satisfying and promising.

7. ACKNOWLEDGEMENT

The authors would like to thanks ACAPELA group for providing the high quality English Text-to-Speech engine used in this demonstrator and Multitel ASBL for his voice recognition module (implementation of this functionality is in progress).

This project is funded by the Ministère de la Région wallonne in Belgium under "Recherche Initiatives 5".

REFERENCES

[1] Braillnote, http://www.pulsedata.com, visited in June 2005.

[2] Pac Mate, http://www.freedomscientific.com, visited in June 2005.

[3] C. Mancas-Thillou, B. Gosselin, "Color Text Extraction from Camera-based Images ─ the Impact of the Choice of the Clustering Distance", in *Proc. of IEEE International Conference on Documents Analysis and Recognition (ICDAR 2005)*, Seoul, South Korea, 2005.

[4] C. Mancas-Thillou, M. Mancas, B. Gosselin, "Camera-based Degraded Character Segmentation into Individual Components", in *Proc. of IEEE International Conference on Documents Analysis and Recognition (ICDAR 2005),* Seoul, South Korea, 2005.

[5] S. Ferreira, C. Mancas-Thillou, B. Gosselin, "An embedded application for degraded text recognition", *in EURASIP Journal on Applied Signal Processing, Special Issue on Advances in Intelligent Vision systems: methods and applications*, 1st quarter 2005.

[6] T. Dutoit, Y. Stylianou, "Chapter 17: Text-to-Speech Synthesis", in *Handbook of Computational Linguistics*, R. Mitkov, ed., Oxford University Press, 2003, pp. 323-338.

[7] J.P. Peters, C. Mancas-Thillou, S. Ferreira, "Embedded Reading Device for Blind People: a User-Centred Design", in *Proc. of Emerging Technologies and Applications for Imagery Pattern Recognition (AIPR 2004)*, Washington DC, USA, 2004.