

A Service Platform for the Visually Impaired

Sibel Teker¹, Hayrunnisa Pektaş², O. Raif Önvural³, Erdal Güvenoğlu⁴

^{1,2,3,4}Maltepe Üniversitesi, Bilgisayar Mühendisliği Bölümü, İstanbul

buketsibel@gmail.com, hayrunnisa.pektas@gmail.com, onvural@maltepe.edu.tr, erdalg@maltepe.edu.tr

Abstract: In this paper, we present a system architecture to provide services for the visually impaired. The system consists of a client software in a smartphone or a thin client device used to capture and display/playout data, a computational platform such as a virtual server used to process the received data and a communication mechanism to transfer data between the thin client and the server. Although the focus of this project is to provide services to impaired people, the platform can be used to commercialize various services for the general public as well.

This work was supported by Turk Telekom under Grant Number 440 000 016 9.

Keywords: Visually Impaired, Image Distortions, Digital Camera, Android.

Görme Engelliler İçin Servis Platformu

Özet: Bu çalışmada görme engellilerin günlük yaşamlarını kolaylaştırıcı hizmetleri sağlama amacıyla geliştirilen bir sistem mimarisi sunulmaktadır. Geliştirilmekte olan bu sistem mimarisi kullanıcı akıllı bir telefonda geliştirilen bir istemci yazılımı ile cihaza gelen metinler veya Braille alfabesi ara yüzü veya kamera ile çekilen üzerinde metin olan resimler ile verilerin girişini yapabilmektedir. Bu veriler geliştirdiğimiz bir protokol ile sunucuya iletilmektedir. Tüm görüntü işleme, metinleri sese dönüştürme, resimlerde bulunan metinleri veri haline dönüştürme işlemleri sanal sunucularda yapıldıktan sonra üretilen ses dosyası kullanıcıya gönderilmektedir. Bu projenin odak noktası engelli insanlara hizmet sunmak olmasına rağmen, platform olarak hem genel halk için çeşitli hizmetler ticarileştirmek için kullanılabilir. Bu çalışma 440 000 016 9 hibe numarası ile Türk Telekom tarafından desteklenmiştir.

Anahtar Sözcükler: Görme Engelliler, Görüntü Bozulmaları, Sayısal Kamera, Android.

1. Introduction

In this paper, we present a new services platform to visually impaired towards providing tools that would make their daily life easier. Although its main focus is for visually impaired population, the proposed platform and produced research can be used to commercialize new services for the general public.

While there are hundreds of thousands of applications available for smartphones today and perhaps thousands of new ones are being deployed every week, there are very few applications available for the impaired in the world and much fewer in Turkey. Pretty much none of the applications developed for the general public can be easily used by the impaired. We believe the proposed platform will help these challenges uniquely.

The system consists of a server farm accessed via internet by thin clients. In practice, the clients are envisioned to be mostly smartphones but could be tablets or PCs as well. The term thin client is used to emphasize the fact that the processing required by an application will be done in a server farm. This is due

to the fact that the memory and processing capacity at envisioned user devices, e.g., smartphones, is rather restricted.

How can visually impaired people use smartphone towards easing their daily activities? Starting with the phone itself, it is pretty hard if not impossible for such a person to use the keyboard even just to dial a phone number. Voice activated dialing systems may be available at some phones but they still are not accurate enough and often frustrating. Can a smartphone be used to read short documents such as menus at restaurants, articles from a regular newspaper? Can a smartphone be used to read book pages? Can a smartphone be used by a visually impaired person to get directions? Can these services be provided in real time or almost in real time?

One possible solution to these problems may be to take a picture of a written material of interest, such as a restaurant menu. An application at the phone can do some minimal processing over the picture file and transmit over the internet to a server. At the server, the text is converted to speech and transmitted to the phone for play out. One of the challenges to enable

such an application is to retrieve text from the received image. In particular, the text material cannot be expected to be received like you would scan a page in a scanner. Instead, perhaps words will be scratched (similar to how the words in middle pages of a thick book), the text material is at an angle to the borders, etc. Hence, a significant amount of image processing may be required to be able to convert the received material to text.

Another possible use of such a solution could be to provide walking directions to visually impaired. Using the GPS capabilities at the smartphone, the user location information together with the destination information can be transmitted to the server. The location information can be also available from a picture taken by the user (a street name or a well-known building). The user can enter the destination information using the Braille alphabet. The smartphone application can update the location information to the server and the server can provide near real time voice commands using its own map together with the information from the user device.

Some basic applications at the smartphones may be available in a rather limited capability, to the visually impaired, such as reading and sending SMS messages, knowing who the caller is, dialing the phone number. However, accessing internet, social sites from a smart phone is a dream for a visually impaired. Text-to-speech (TTS) software capabilities has shown significant progress in recent years but these programs become rather restricted when used in smartphones. This is mainly due to the processing capabilities and amount of memory available at these devices. Similarly, image processing is a process and memory intensive task and what can be done in smartphones is expected to be rather limited for years to come. The basic contribution of the proposed architecture is to do the processing of most applications at servers where amount of memory and processing power is plenty while minimizing the amount of processing and memory requirements at smartphones.

1. 2. Services Architecture

The Braille system is a method widely used by blind people to read and write. Each Braille character is made up of six dot positions, arranged in a rectangle containing two columns of three dots each. A dot may be raised at any of the six positions to form sixty-four (2^6) possible subsets, including the arrangement in which no dots are raised [1, 2, and 3]. Figure 1 illustrates an example of letters defined in this alphabet.

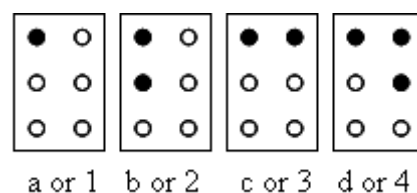


Figure 1. Examples of Letters in the Braille Alphabet

In the proposed architecture, the key entry system for text is the Braille Alphabet implemented in a smartphone. The phone screen is divided into six areas and finger touching in an area corresponds to a dot in the Braille Alphabet.

The second entry system is the camera. Assuming the user can take a picture of a document of interest, the document, after some minimal processing at the client, is sent to a server, the server will have the intelligence to process the received image and convert the text retrieved from the image to speech, which then can be played out at the user device

Enabling these two entry systems at the phone together with the processing capabilities that will be developed at the server can enable a variety of applications and services for the visually impaired.

User devices have limited processing capacity and memory. Hence, its use for computationally intensive applications is rather limited. Furthermore, IP addresses change as the user moves from one location to another, potentially introducing challenges to phones accessing servers. On the other hand, cloud computing in general or server farms in particular have relatively unlimited processing and memory capacity and, fixed IP addresses. Hence, it is a logical step to enhance smartphone applications with cloud computing to reduce the problems related to limited battery, processor speed, memory size, data storage, and changing addresses. In this framework, the main challenge is to decompose an application execution between the client and the server seamlessly. We propose to address some of these challenges by introducing a service enabler layer between the applications and the operating system.

In this framework, we envision there is a thin service layer between the applications layer and the phone operating system, referred to as service enabler, as illustrated in Figure 2.

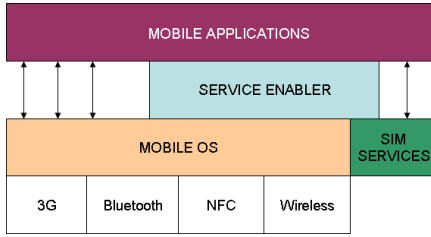


Figure 2. Service enabler layer

The objective of the service enabler level is to provide an interface from mobile applications that will be developed for the visually impaired and provide common functions that will be required by such applications. Examples of services that will be implemented in the service enabler include the common user interface implemented as a phone book, the use of Braille Alphabet among various applications, text retrieval (such as SMS, email, etc.), communication services with the server including login information.

As discussed previously, the services to the user are provided at a server. The architecture will support a large number of users and possibly tens of services supporting this large number of users. A high level view of the server architecture is illustrated in Figure 3. When a client starts the application, it automatically connects to a web service and logs in to the Login/Admin unit. After the qualifications of the client

are checked, the unit creates a virtual machine to manage the traffic between server applications used by clients and the clients themselves. In this framework, each user is assigned to a virtual machine configured to support one user only. Each virtual machine is assigned to a local IP address and communicates with external services via a router in its segment. The required communication services such as DNS, NAT, DHCP, etc. are provided by this router as well.

Once its user virtual machine is created, the client only communicates with this machine. As a simple example, if client y wants to read an e-mail message, the message is sent to user y virtual machine. The virtual machine then sends the received data to e-mail services virtual machine which retrieves the text in the e-mail and some information such as the sender. The e-mail services then informs client y services machine that the required operation is completed and the address where the output file is stored. Client y services then sends a message to TTS (Text-to-Speech) services virtual machine requesting the text to be converted to a speech file and the address where the file is stored. The TTS services retrieves the file, converts it to a speech file and then informs the client y services the operation is complete and the address of the voice file. The client y services then tells the client the service is complete and provides it with the address of the file.

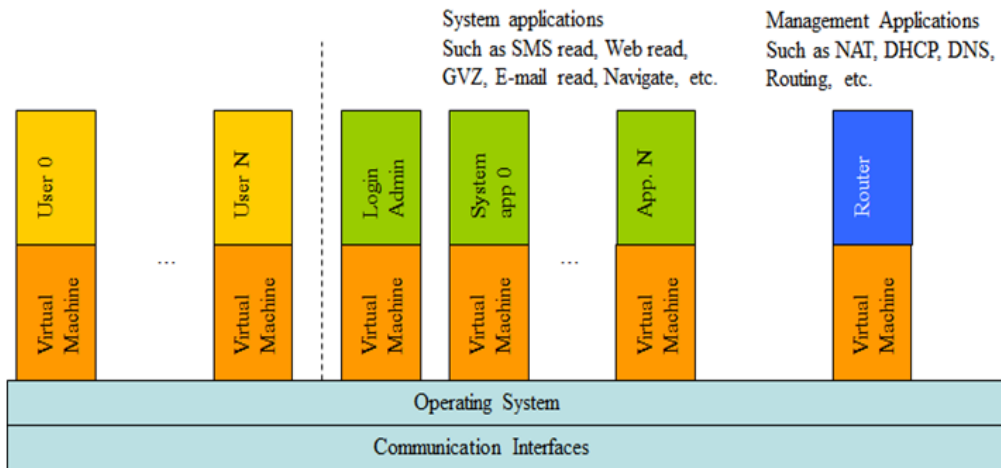


Figure 3. Server architecture

All communication in this system uses http. In particular, clients and their user services virtual machines, user services virtual machines and applications services virtual machines, all the communication to and from router uses http to communicate. We developed a proprietary user level control protocol for messaging that uses http to send and receive. This framework presents us the flexibility of distributing user services virtual machines and application services virtual machines without physical limitations. It also provides scalability of the system.

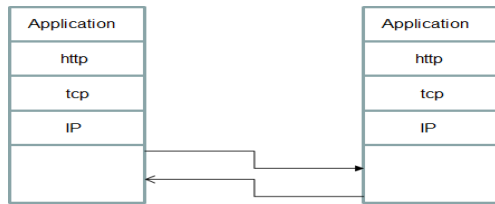


Figure 4. Communication Architecture

One of the main challenges in providing voice services to visually impaired is to retrieve text from the received image. Figure 5 illustrates examples of a typical image of a text when the image is taken by a camera. As seen in this figure, it is necessary to process these received images before the text in these pictures can be retrieved. In general, there are three types of distortions of interest to our application: angled distortion, geometric distortion, and perspective distortion. In addition to processing required for these types of distortions, it may be necessary to enlarge the received image with as little additional distortion as possible caused as part of this process. This is known as scaling problem in image processing. It is also necessary to identify and mark the area of interest to the user in a received image [4].

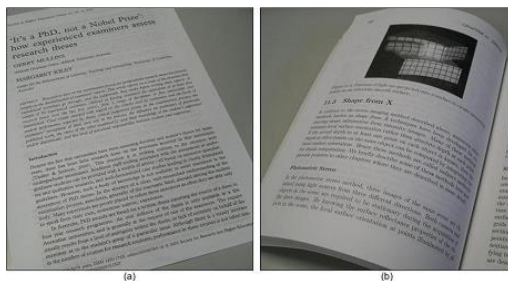


Figure 5. Images with perspective distortion

Next, we illustrate various examples of problems that need to be satisfactorily addressed.

At times, it is possible that the text in an image is too small to be identified. Hence, it is necessary to enlarge the text first as shown in Figure 6 (a) and (b). After it

is enlarged, it is possible that angled and/or perspective distortions are detected. Hence, further processing may be necessary before the text of interest can be retrieved.

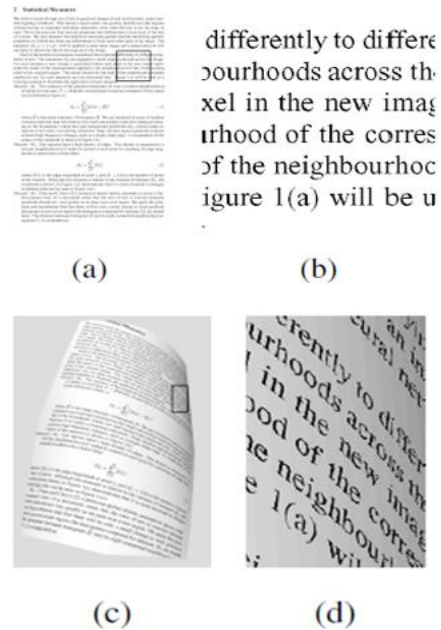


Figure 6. Examples of enlarging partial images for further processing [5]

In various images, first we need to identify the area of interest to the user, as illustrated in Figure 7. For example, if we are interested in identifying the traffic plate of a car, the plate needs to be identified as accurately as possible and separated from the rest of the image. It is then possible to further process the retrieved image of interest to read the interested text. In this figure, we illustrated other examples from a chess board. In this figure, we also notice a picture of a text taken with an angle and distorted perspective. In order to retrieve the text in this picture, first the angled distortion needs to be corrected. It is then necessary to correct the perspective distortion before the individual words can be detected.

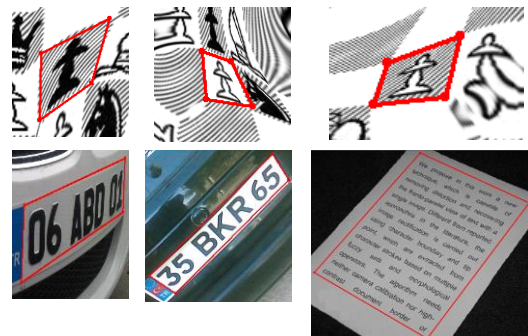


Figure 7. Distorted images for correction [6]

2. 3. Conclusions

The main contribution of the proposed architecture is to use smartphones as client devices and server farm for the computational tasks to support the development of applications that could help impaired people in their daily lives. While the applications that will be developed with this architecture would help more than the 700,000 visually impaired alone in their daily activities, they can also be extended to be used by the general public, serving millions of wireless customers and Internet customers.

4. References

3.

[1]http://wiki.answers.com/Q/How_does_the_Braille_system_work, Date of access: 21.06.2012

[2]<http://www.newworldencyclopedia.org/entry/Braille>, Date of access: 21.06.2012

[3]<https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Braille.html>, Date of access: 21.06.2012

[4] Clark, P., Mirmehdi, M., “Estimating the Orientation and Recovery of Text Planes in a Single Image”, in Proc. of BMVC'01, pp. 421.430,2001.

[5] Liang, J., Dementhon, D., Doermann, D., “Geometric Rectification of Camera-captured Document Images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, Issue. 4, Pages: 591-605, 2008.

[6] Güvenoğlu, E., “Optik görüntü Bozulmaların Yazılımla Düzeltmesi İçin bir Yöntem”, Trakya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Doktora Tezi, 2012.