S. Boll, S. Maaß & R. Malaka (Hrsg.): Workshopband Mensch & Computer 2013
München: Oldenbourg Verlag, 2013, S. 527–529

Detecting of and Interacting with Text in Free Space

Frank Wippich, Christian Graf, Daniel Drewes

Blindsight Europe GmbH

Abstract

Computer Vision technology has seen a significant push in popularity in consumer applications, whether for face recognition in today's smartphone camera apps, or server-based object recognition that identifies products, text or contexts in images within seconds. And thanks to powerful smartphone and tablet devices, computer vision application can now perform the processing locally, providing real-time feedback to the user, about the object they are looking at just by pointing the camera lens at it – without the need of a remote server.

Hence, new ways of interacting with one's environment become possible in particular for people with specific access needs, like vision-impaired users. Blindsight's text detection algorithm showcases a way on how to detect and speak out text in virtually real time using a smartphone.

In this demo, users will be able to try out text detection and immerse into a "non-visual" user-experience showing the state of the art for text detection and assistive technology for the visually impaired. Based on the user's experiences, this demo shall also discuss challenges and new ways to interact with text in free space. How can technology guide a user to find the desired text, how can technology filter the right information or group the information based on context, conventions or user preferences?

Providing ideas and answers for these question will not only enrich the advancements of assistive technology, it will also inspire new application in a world where visual experiences are more and more supplemented by tactile, audible and vocal user interfaces.

1 Background

Technology is advancing at a breakneck speed – thanks to Moore's law processing power is doubling every 18 months making smartphones powerful devices that can run sophisticated algorithms to determine locations, identifying objects, shooting high resolution photographs and videos. The merging of the virtual world and the real world has already happened and it will further expand as devices get even more powerful, data networks even faster and storage more affordable.

This development has also had an impact on assistive technology. Computer vision technology for example can be deployed to detect objects and text to help people with vision impairment navigating through daily life and interacting with their environment. Combined with accessible user interfaces – where Apple has played a pioneering role to make a universal design mainstream – vision impaired people now have access to a wealth of applications on one single device, where they needed dozens of point-products earlier.

Reading text is a particular example to help people with vision impairment - with today's smartphones, optical character recognition software can scan documents on the fly just by pointing the phone's camera at the desired object. Applications are commercially available for around US$10 or less, such as TextGrabber from ABBYY or TextDetective by Blindsight, as compared to several hundred or thousand dollars that dedicated or stationary text readers cost.

Nevertheless the process of finding the text in the environment without sight, remains a big obstacle and requires usually a lot of retries to take an image, if not visual assistance.

Even for sighted users, it is often difficult to properly frame a document such that the OCR app can correctly identify its boundaries. The apps rely on the sighted user to correct the boundary before starting the long process of scanning the document for text. While this is only mildly cumbersome for a sighted user, it is awful for a blind user, who only discovers at the end of scanning whether they have properly imaged the document. Even worse, often the blind user has to run through the full process several times to get a good scan, only to discover at the end that the text that was scanned was not worth reading. Thus, text remains the most critically inaccessible part of the environment to people with vision impairment.

2 The Demo

Blindsight has developed a proprietary text detection algorithm that can address exactly the problem of text acquisition. Thanks to powerful processors and high-resolution cameras in smartphones and tablets, it is possible to finding areas of text from images in natural environments and processing it in real time.

"Finding and reading text in free space through the eyes of a vision impaired person" will thus be the theme of the demo.

A smartphone (iPhone 5) or tablet (iPad) with the app TextPeriscope will enable the user to explore text for different use cases:

- Finding text on a sign (a room number or name plate for example)
- Identifying a product (a softdrink, can of food or a candy bar)
- Finding text on a document (a restaurant menu or envelop)

The demo will showcase two stages in text interaction: First, the user can explore the environment for text, where the text is read to the user as it is detected in real-time. Second, the user can search for a specific word or text area and the smartphone will point the user in the direction where the text is. Essentially, the smartphone will act like a "reading wand" that points you to text though audible or tactile feedback.

In addition, users may also try out the demo with blindfolds and/or specific simulation glasses, to simulate vision impairment.

In order to get a comparison to conventional text reading apps, the demo will also include state of the art applications that are already commercially available, such as TextDetective (a Blindsight product) or TextGrabber (an ABBYY product). Users may do a comparison test between the commercially available apps and the real-time reader TextPeriscope.

TextPeriscope will perform the following steps in real time on the phone (no remote server connection required) when pointing the camera lens at text:

- Selecting frames from the camera's video stream and detecting / skipping blurry frames
- Detecting text slices and cluster them within & between scales
- Send the detected text snippets for "OCR Preparation" to perform noise removal (such as perspective transforms, removal of artifacts, etc.)
- OCR – Tesseract
- Filtering text on a specific keyword and calculate the position of the word to a reference point on the image
- Text-To-Speech and text output OR audio/tactile output that communicates the distance (e.g. sought-for word is spoken out loudest when it is in the center of the screen and fades as it moves to the edges of the field of view).

3 Objective of this Demo

The goal of this demo is simple: showing how computer vision software, deployed with today's smartphones can enable anybody to read – regardless of visual or even cognitive abilities. And, more importantly, it shall discuss future ways and challenges in interacting with text in free space – how to search for and find the *right* text, filter it, and process it; whether with smartphones, tablets or wearable devices such as Google Glass.

While the problem of finding the text in free space is solved, the challenges of accuracy and the way how to present information to the user are now to be tackled.

In order to bring this smart accessible technology to the 285 million vision impaired people in the world (WHO Media Center 2012), the discussion shall focus on exploring new ways of user interaction in terms of communicating and educating the user of spatial relations of text areas or objects as well as filtering information based on the user context - because only an easy user experience makes a great technology a success.

Bibliography

WHO Media Center (2012). *Visual Impairment and Blindness – Facts Sheet*, World Health Organization,http://www.who.int/mediacentre/factsheets/fs282/en/