WatchMe: A Novel Input Method Combining a Smartwatch and Bimanual Interaction

Wouter Van Vlaenderen
Expertise Centre for Digital Media
Hasselt University – tUL – iMinds
first.last@student.uhasselt.be

Jo Vermeulen
Expertise Centre for Digital Media
Hasselt University – tUL – iMinds
&
HCI Centre
University of Birmingham, UK
first.last@uhasselt.be

Jens Brulmans
Expertise Centre for Digital Media
Hasselt University – tUL – iMinds
first.last@student.uhasselt.be

Johannes Schöning
Expertise Centre for Digital Media
Hasselt University – tUL – iMinds
first.last@uhasselt.be

CHI'15 Extended Abstracts, Apr 18-23, 2015, Seoul, Republic of Korea
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732789

Abstract

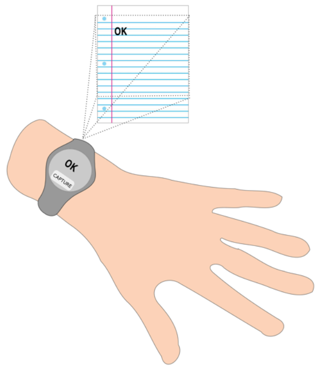
Smartwatches can facilitate several tasks that are performed on mobile devices. However, due to their limited size, touch interaction can be cumbersome. Although alternative input modalities such as speech input could be used, these can also introduce other issues regarding privacy or ease of use. Consequently, HCI researchers have been exploring novel input techniques for smartwatches. In this paper, we introduce WatchMe, a smartwatch application that uses the smartwatch camera and image processing techniques to allow for providing input on a drawing canvas composed of everyday objects and surfaces. We rely on a cloud OCR engine to retrieve text from captured images. Combining these characteristics, we illustrate some scenarios in which WatchMe could be used, such as a novel method for two-factor authentication.

Author Keywords

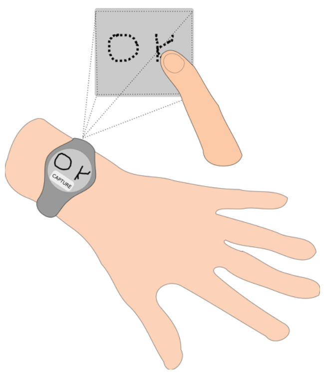
Smartwatch Interaction; Wearable Computing; Input Methods; Image Processing; Text Through Display.

ACM Classification Keywords

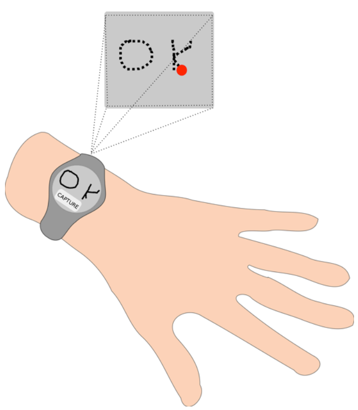
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.



(a) Text input using OCR



(b) Text input using finger tracking



(c) Text input using a laser pointer as an input device

Figure 1: Different input modalities scenarios for the WatchMe smartwatch app.

Introduction and Motivation

Modern smartwatch display sizes range from around 1.26” (the Pebble smartwatch) to 2.4” (the Pine smartwatch by Neptune) using different aspect ratios. Even though touch interaction is common in smartwatches, their limited display size introduces several issues that are well-documented, such as the fat finger problem [12].

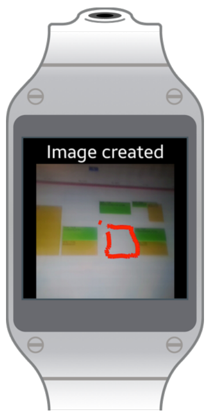
To address these problems, HCI researchers have explored input techniques that make better use of the limited screen space available [2,4,13,16]. Zoomboard [1] for example, enlarges keys using a tree-like zooming technique. H4-writer [9] uses 4 buttons, each at one side of the screen and containing multiple letters, in combination with a prediction algorithm. The 1Line keyboard [8] offers multiple horizontal aligned buttons for text entry and correction. TouchSense [5] exploits the ability of the human finger to touch an object using different contact points. These techniques are all aimed at improving the use of the touchscreen as the main input channel.

In this paper, we introduce WatchMe, a smartwatch application that uses the smartwatch camera as the main input channel. Image processing techniques are used to detect input provided by users drawing on everyday surfaces and objects. We rely on a cloud OCR engine to retrieve text from the captured images. Combining these characteristics, we illustrate some scenarios in which WatchMe could be used, such as a novel method for two-factor authentication.

WatchMe Concept

Instead of improving touch interaction on the small display, WatchMe focuses on using the smartwatch camera to provide input. WatchMe was inspired by the collection of examples for mobile devices by Ballagas [1]. A number of smartwatches, such as the Samsung Galaxy Gear, already include a camera. WatchMe is a fully functional smartwatch app that creates a large input canvas from images captured by camera. Essentially, the smartwatch camera acts as a ‘peephole display’ [15] on this larger input canvas.

WatchMe can recognize input provided in various ways. For example, a pen can be used to draw strokes on a piece of paper, which is then processed by an OCR engine (Figure 1–a). Additionally, it is possible to track the movement of a laser pointer or of the user’s finger to create an input channel where everyday objects are augmented with tracked strokes (Figure 1–b, 1–c). The camera provides functionality to track an input device and creates a resulting picture from it, combining its preview and tracked strokes. Depending on the user’s preference, this image can then be sent to other users as an attachment, or can be transformed into text using OCR techniques. Each different input modality has its advantages and disadvantages. For example, it is easy to interact with objects at a distance with the help of the laser pointer (see also Myers et al. [10], who introduced a similar technique for handhelds). The user’s finger, on the other hand, can serve as a ubiquitous input device [4]. Nevertheless, these three basic input techniques can open up a large number of different applications. We envision that these techniques can be quite useful for quick responses to text messages, missed calls or other notifications.



gure 2: WatchMe smartwatch app in use. user creating a calendar entry with a ser pointer.

gure 3: WatchMe smartwatch app in use. user performing a two-factor ɹthentication on a website again using a ser pointer as an input device.

The following three scenarios illustrate the concepts of WatchMe in detail:

Scenario 1
Upon receiving a message, a user can choose to reply using WatchMe. She uses her pen to write a message on a piece of scratch paper, and takes a picture of the piece of paper by tapping the screen (Figure 1–a). The OCR engine then processes the picture, and eventually allows the user to confirm her input and send the message. In absence of a piece of scratch paper the user can also chose to write her text on any other surface, like a wall or table, using a tracked laser pointer or finger (Figure 1–b, 1–c).

Scenario 2
A message is received and the user is being asked to provide a timeslot to arrange a meeting. While checking her calendar, she notices multiple possibilities. She annotates these timeslots on her calendar using her laser pointer and replies by sending over the composite image (Figure 2). This technique can also be used to annotate documents. The user could, for instance, underline, cross out or annotate certain words on a document with her laser pointer.

Scenario 3
The user is being asked to login to her account using two-factor authentication. When entering her username, a numeric keypad appears on the website, possibly containing multiple markers. She points the camera of her watch at the keypad and uses her laser pointer, or finger to enter a PIN number (Figure 3). This PIN number is then tracked, processed, and sent to the server. Finally the user is able to login securely (Figure 4). Just as traditional two-factor authentication requires the user to confirm a code sent to her mobile device (which is assumed to be personal), this authentication method assumes the user's smartwatch is a personal device which only the user should have access to.

Implementation

We have implemented a couple of scenarios with WatchMe, focusing on the interaction concepts *a* (OCR) and *c* (laser) shown in Figure 1. We used the Samsung Galaxy Gear smartwatch, as it provides us with a built-in 1.9 megapixel camera. Given our aim for the watch to act as an independent device and the fact that the standard Galaxy Gear software has no in-built network connection, we flashed the smartwatch to a custom Android ROM to enable this feature. We used the *NULL_* ROM to provide the watch with an Internet connection using Bluetooth tethering.

We rely on OpenCV[1] for image processing. Frames that are captured by the camera are converted from RGB to the YUV range. Next the laser dot is extracted using thresholds on the Y, U, and V channels. Combining a median blur and calculations of coordinates, tracking functionality is provided. To reduce false positives, a threshold is set on the radius of the tracked laser dot combined with calculations of average distances between consecutive tracking points.

WatchMe relies on Abbyy[2] as a cloud-based OCR engine. Images are uploaded using a HTTP POST request, and an XML file containing the result is downloaded using a HTTP GET request. This result

[1] http://opencv.org/

[2] http://www.abbyy.com/

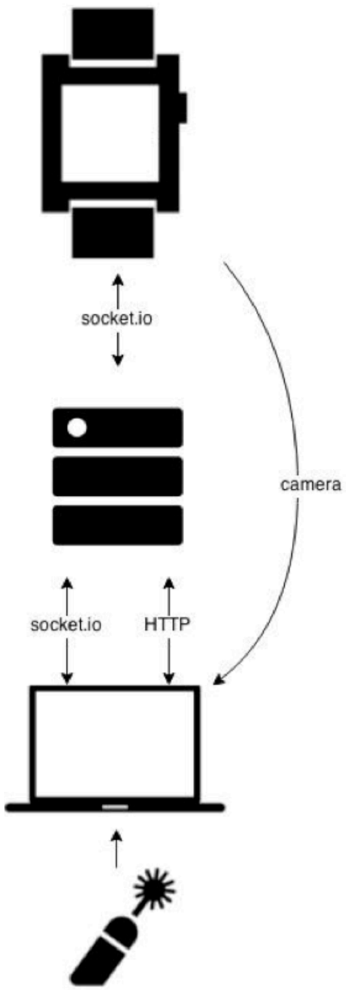


Figure 4: Two-factor authentication.

contains the recognized text, as well as the credibility of every recognized character. Multiple lines and paragraphs can also be recognized as Abby returns information about lines and fields on a page. Based on these measurements, it is possible to make WatchMe can interact with the built-in Android dictionary, which learns the user's personal words and phrases over time.

To implement two-factor authentication, the application needs to communicate with a website. We have opted for socket.io[3] since this implies cross-platform cross-device real-time bidirectional event-based communication. A Node.js server implementing socket.io was set up to serve web pages for authentication. Upon requesting a webpage, server side JavaScript is used to connect to this socket.io server. Gottox' socket.io java client is used on the watch to connect to this same socket.io server, providing us with the opportunity to communicate. Input is recognized on the watch, and then sent to the server. The server retransmits this input to the client to provide the client with feedback about pressed keys. When the server recognizes a valid PIN number being entered, the client is notified, and the user can securely log in (as can be seen in Figure 3 and 4).

Conclusions and Future Work

In this paper, we introduced WatchMe, a smartwatch application that uses the smartwatch camera and image processing techniques to detect input provided by users drawing on everyday surfaces and objects. We feel there are interesting opportunities in using the smartwatch camera for providing input. Next to overcoming the difficulties in using touch interaction on small displays, this also opens up the possibility to use the smartwatch to annotate and augment real-world objects.

Results of a first informal study with the WatchMe app are promising. Even though it took participants some time to become familiar with the interactions (mostly due to the indirect viewing angle and the absence of feedback on the input canvas), participants found the app to have a gentle learning curve overall.

In the future, tracking could be further improved by filtering out reflections of the laser dots on reflective materials. We would also like to explore different combinations of the smartwatch display and camera (similar to the work of Cauchard [3]), as the current angle of about 90 degrees is not optimal for the prosed interactions. An additional limitation is the relatively high power consumption when using the camera.

Finally, we would like to perform a formal user study with our application to evaluate the effectiveness and intuitiveness of our interactions in different usage contexts, similar to the studies conducted by Rohs [14]. In addition, it could also be interesting to compare WatchMe with traditional smartwatch input techniques.

[3] http://socket.io/

References

[1] Ballagas, R., Borchers, J., Rohs, M., & Sheridan, J. G. (2006). The smart phone: a ubiquitous input device. *Pervasive Computing*, IEEE, 5(1), 70-77.

[2] Bernaerts, Y., Druwé, M., Steensels, S., Vermeulen, J., & Schöning, J. (2014, June). The office smartwatch: development and design of a smartwatch app to digitally augment interactions in an office

environment. In *Proceedings of the 2014 companion publication on Designing interactive systems* (pp. 41-44). ACM.

[3] Cauchard, J. R., Löchtefeld, M., Irani, P., Schoening, J., Krüger, A., Fraser, M., & Subramanian, S. (2011). Visual separation in mobile multi-display environments. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 451-460). ACM.

[4] Harrison, C., Tan, D., & Morris, D. (2010, April). Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 453-462). ACM.

[5] Huang, D. Y., Tsai, M. C., Tung, Y. C., Tsai, M. L., Yeh, Y. T., Chan, L., ... & Chen, M. Y. (2014, April). TouchSense: expanding touchscreen input vocabulary using different areas of users' finger pads. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 189-192). ACM.

[6] Kim, J., He, J., Lyons, K., & Starner, T. (2007). The Gesture Watch: A Wireless Contact-free Gesture based Wrist Interface. In *Proc. of Wearable Computers* (pp. 15-22). IEEE.

[7] Lamping, J., Rao, R., & Pirolli, P. (1995). A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 401-408). ACM Press/Addison-Wesley Publishing Co.

[8] Li, F. C. Y., Guy, R. T., Yatani, K., & Truong, K. N. (2011, October). The 1line keyboard: a QWERTY layout in a single line. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 461-470). ACM.

[9] MacKenzie, I. S., Soukoreff, R. W., & Helga, J. (2011, October). 1 thumb, 4 buttons, 20 words per minute: Design and evaluation of H4-Writer. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 471-480). ACM.

[10] Myers, B. A., Peck, C. H., Nichols, J., Kong, D., & Miller, R. (2001, January). Interacting at a distance using semantic snarfing. In *Ubicomp 2001: Ubiquitous Computing* (pp. 305-314). Springer Berlin Heidelberg.

[11] Oney, S., Harrison, C., Ogan, A., & Wiese, J. (2013, April). ZoomBoard: a diminutive QWERTY soft keyboard using iterative zooming for ultra-small devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2799-2802). ACM.

[12] Siek, K. A., Rogers, Y., & Connelly, K. H. (2005). Fat finger worries: how older and younger users physically interact with PDAs. In *Human-Computer Interaction-INTERACT 2005* (pp. 267-280). Springer Berlin Heidelberg.

[13] Rekimoto, J. (1996). Tilting Operations for Small Screen Interfaces. *In Proc. of UIST 1996*. ACM.

[14] Rohs, M., Schöning, J., Raubal, M., Essl, G., & Krüger, A. (2007, November). Map navigation with mobile devices: virtual versus physical movement with and without visual context. In *Proceedings of the 9th international conference on Multimodal interfaces* (pp. 146-153). ACM.

[15] Yee, K. P. (2003). Peephole displays: pen interaction on spatially aware handheld computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1-8). ACM.

[16] Xiao, R., Laput, G., & Harrison, C. (2014). Expanding the Input Expressivity of Smartwatches with Mechanical Pan, Twist, Tilt and Click. In *Proc. of CHI 2014* (pp. 193-196). ACM.