Ubiquitous Shortcuts: Mnemonics by Just Taking Photos

Mikko J. Rissanen
mjrissanen@ntu.edu.sg

Owen Noel Newton Fernando
ofernando@ntu.edu.sg

Horathalge Iroshan
iroshan@ntu.edu.sg

Samantha Vu
sgtvu@ntu.edu.sg

Natalie Pang
nlspang@ntu.edu.sg

Schubert Foo
sfoo@ntu.edu.sg

Centre of Social Media Innovations for Communities
Nanyang Technological University
14 Nanyang Drive, 638332 Singapore

Abstract

Ubiquitous Shortcuts is an image processing based method for making and using mnemonics set onto the real world using smartphones or other computing systems. The mnemonics can be created by taking photos of the user's vicinity and by binding them onto command sequences. The mnemonic is triggered every time a similar photo is taken. Our method uses natural feature matching algorithms and end-user programming approaches. The mnemonics can be concatenated into more complex command sequences. Thus, limited user input is realized by just taking photos with a camera embedded into a finger-ring, which enables rapid, subtle and socially acceptable user interaction. Our method can be used as semi-automatic way of achieving location and context sensitive services, activity recognition or tangible interaction.

Author Keywords

Ubiquitous Computing; Intuitive Interfaces; Natural Feature Matching; End-user Programming

ACM Classification Keywords

H.5.2 User Interfaces: Interaction styles, Input devices and strategies; I.4.7 Feature Measurement
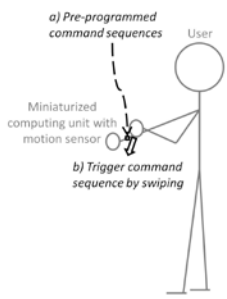
Figure 1: Related work: Example of micro interactions based user input as proposed by Ni & Baudisch [1].
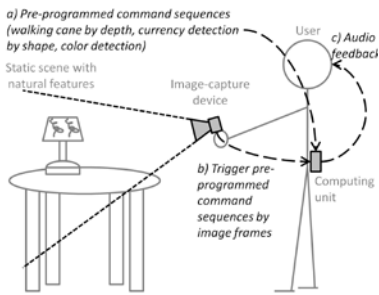

Figure 2: Related work: Free-air pointing based user interaction as proposed by Nanayakkara et al. [9].

Introduction

The miniaturization of mobile devices is making steady progress. This affects the way user interactions have to be designed in future products and services. Ni & Baudisch [1] expected tiny wearable, mounted or implanted devices to enable "only the most basic interactions" and proposed gestures made on tiny interfaces without using any display, e.g. using fingers to swipe over a sensor to produce Morse code-like input (Figure 1). This kind of ubiquitous computing methods include sensors on the user and in the environment and/or complex algorithms for gesture or context recognition. Many prototypes demonstrate gesture-based micro interaction (see [2]). Even a simple implementation requires a tag and a device to read it, e.g. WebStickers [3].

Social acceptance, subtlety, speed and low cognitive load have been understood as key factors in mobile user interaction. Williamson et al. [4] have shown that even simple gestures such as wrist rotation can be uncomfortable in public places such as trains. EMG-based interfaces worn under clothes could achieve motionless gesture interaction thus being unnoticeable by spectators [5]. Montero et al. [6] defined two levels of social acceptance, one of the user, and the other of the spectator, which both affect appropriateness of mobile user interfaces. The most socially acceptable gesture interaction classes are *secretive* (manipulation and effects not perceived by spectators) or *expressive* (manipulation and effect perceived and easily understood by spectators) interactions [6]. The study by Oulasvirta et al. [7] showed that most mobile interactions happen in 4 to 8 second bursts on a bus stop, escalator, metro car, etc., which lets us understand that mobile interactions suffer from high

fragmentation and serious attentional limitations. Moreover, high cognitive load is an obstacle for user groups with special needs such as the elderly that require fewer functions and easy menu paths [8].

This paper describes our pursuit for creating a practical ubiquitous user interface that enables rapid, subtle and socially acceptable interactions that would also be easy to learn. The work on EyeRing [9] has demonstrated how a finger-worn camera can enable free-air pointing interaction (Figure 2), for applications useful to the blind. We aim at enabling free-air pointing and even more subtle interactions for smartphones and other computing devices. Our thinking is influenced by the *body mnemonics* concept [10], finger-worn user interfaces [11, 9], object tagging [3] and end-user programmed devices that have emerged in mobile device markets. For example *Sony Xperia™ SmartTags*[1] and Samsung's *TecTiles*[2] are user-programmable NFC tags that user can hang in places such as at home, in a car or on clothes similar to using QR codes or barcodes. These tags, however, are markers that require modifications to the environment. Interaction with the tags as well as with the *body mnemonics* is neither *secretive* nor *expressive*, but rather *suspenseful* (as in [6]), i.e. not optimal in theory. Our proposed user interaction method achieves secretive interaction by allowing selection type of operations to be embedded into any static part of the user's surroundings, clothes or body. We describe work-in-progress on the method, its expected benefits to various user groups as well as applications in a markerless real world environment.

---

[1] www.sonymobile.com/gb/products/accessories/xperia-smarttags/

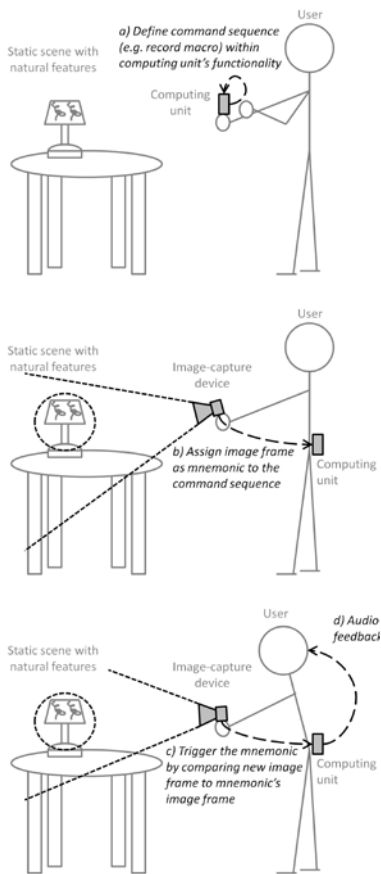[2] www.samsung.com/us/microsite/tectile/

Figure 3: Making and triggering of ubiquitous shortcuts. Natural features enable the image frames to be taken from slightly different angles and yet be matched logically.

Ubiquitous Shortcuts

We propose Ubiquitous Shortcuts (ubicuts), mnemonics set to the real world that can be created on any static object or combination of objects by arranging them into an image that is meaningful for the user. Also they can be concatenated into more complex command chunks.

Concept

Figure 3 demonstrates how ubicuts can be created by defining a command sequence (e.g. by recording a macro), assigning a reference image frame to it (e.g. by taking a photo) and how they can be triggered by free-air pointing using the body part that the image capture device is attached to (finger, wrist, earlobe). Mnemonics database contains reference images and the command sequences they are associated with. Image processing methods are used to recognize similarity of the reference image frame and the triggering image frame. Ubicuts can be implemented using an embedded video camera of a modern smartphone or a wearable still photo or video camera, e.g. *Looxcie*[3]. Ubicuts do not require any physical markers such as RFID tags or NFC-enabled devices for creating mnemonics. Therefore, ubicuts can be created nearly anywhere in the user's static physical environment, clothes or body. To enable feedback, speech synthesis provides information about what command sequence is about to be executed, during which the user can cancel the command manually using the computing unit. In theory, it is possible to trigger ubicuts without pressing any physical button. This can be done for example with a *Looxcie*-like device that has 10h battery life for continuous recording and an acceleration sensor that toggles the video recording on and off.

Implementation

We have implemented the first wearable user interface device that enables ubicuts. The current version still requires physical buttons. The system consists of an embedded camera, a smartphone and an app, and a server-side remote image matching and training services as shown in Figure 4. Design variations include a bracelet and a finger-ring shaped image capture devices (Figure 5) as well as a head-worn video camera (Figure 6). The main components of image capturing device consist of a Bluetooth module and a serial camera (Figure 7). When the user presses the trigger button, the captured JPEG image transmitted to the smartphone via Bluetooth module. The microcontroller is used to initialize, determine the data format, synchronize the serial camera, and read the status of the trigger button. The serial camera operates at 57600 baud rate with resolution of 320 x 240 px.
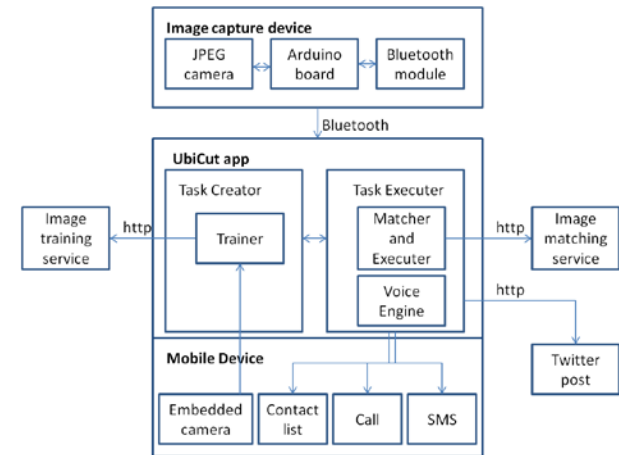

Figure 4: Overall system architecture.

---

[3] www.looxcie.com

3

Figure 5: Sketches of bracelet and ring designs with pin-hole cameras.

Figure 6: Concept of using a head-worn video camera and a desktop wallpaper as a ubicut image to send an SMS about reaching work safely.
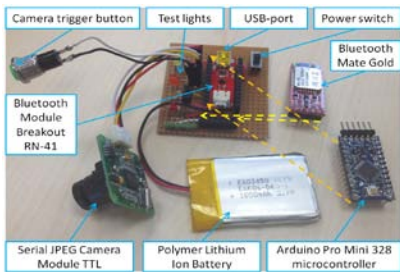

Figure 7: Technical components.

UbiCut app is implemented on Android. It has two main components: Task Creator and Task Executer. Task Creator is used to define command sequences (macros) for the user. The Trainer sub-module provides image training service for real-time still images. Predefined images are trained to detect previously taken image frames. Task Executer uses the image matching service to find the most similar image among ubicuts references images stored in a local SQLLite database. Also, it produces feedback to the user via Android's text-to-speech function. Image matching and image training services are done through IQ Engines[4]. Each image is cropped so that the centre part is used for matching. Current implementation supports making phone calls and sending SMS to a contact and use of an external web service (tweeting on Twitter).

Advantages in user interaction
Ubicuts are fast to trigger, subtle nearly to the point of being unnoticeable especially to others and therefore most likely to be socially acceptable. They are especially beneficial in two main types of usage: 1) when the user is not able to give the commands using conventional user interfaces (because of computer-illiteracy or physical or cognitive deficiency), and 2) when the user gives the same computer commands repetitively. Therefore elderly, disabled and computer illiterate users are the main user group benefiting from using ubicuts, since the mnemonics can be assigned for instance to parts of the daily environment in a way that is easy to remember. For instance, the user can trigger a phone call by just pointing a page of a photo album with the image capture device. A ubicut could be created on a lamp to increase lighting connected to a

smart home system operated via a smartphone app. No interaction with the actual smartphone is needed. Thus, use of a smartphone's functionality in "4-second bursts" (as in [7]) is feasible. Also, it is possible to trigger ubicuts by masquerading the act to body language without interrupting the social situation by breaking eye contact, which achieves subtlety (as in [4, 5]).

Current technical limitations
There are several limitations in the current implementation. Arduino Pro Mini 328 microcontroller was chosen for its small size and configurable image size but it has limited processing power. These are not yet optimal for real-time use. Currently, we are using IQ Engine's web interface to achieve image matching separately from the mobile device. It can cause some negative detection for blurred images and noise images. To enable somewhat accurate and fast matching of images, at least three reference images are needed for each ubicut. Also, the UbiCut app only supports a few basic functions because Android does not permit direct macro-recording without the use of surveillance methods such as key logger software. Our implementation will be extended to include more built-in functions of the smartphone as well as more external applications and web services, e.g. Facebook. Although the technical performance of the system still has room for improvements, we highlight that the speed is in the free-air pointing based user input. A quick snapshot with the camera can be done in a few seconds. After taking the photo, it is not time-critical for the user if the command sequences are triggered later, as long as the image-matching accuracy is reliable. Moreover, the current implementation of a bracelet form suffers from too large size to fully achieve the aim of subtlety.

[4] www.iqengines.com

Figure 8: Use of ubicuts as assistive technology. End-user programming is done by a family member to enable the senior citizen call for more rice.

Figure 9: Ubicuts as context sensitive application to set smartphone into silent mode in a meeting.

Applications

There are several practical applications for ubicuts. Although numerous advanced approaches have been proposed by various researchers to achieve similar functionality, we argue that ubicuts could be a low-cost solution for many of the following applications given that a smartphone is already present.

Assistive technology
The functionality of ubicuts can help in practical scenarios to facilitate computer use for people who are not familiar with or able to use normal computer input methods (elderly, disabled). Basic functions of a smartphone can be mapped into ubicuts by a computer-literate person (e.g. caregivers) to meet specific communication needs of the actual user as in Figure 8. By separating the access to the computer system's functionality from its built-in user interface, e.g. a smartphone's menu-structure and touchscreen which currently can act as a limiting factor [8], each user's most common functions can be accessed through ubicuts. Cognitive load due to complex menu structures can also be avoided.

Context and location sensitivity
As ubicuts can be created onto any real world scene from which a unique image frame can be taken of, they can be used as a semi-automatic way of triggering location or context sensitive command sequences. For instance, a jogger can set a static location-based ubicut on his usual jogging route to mark lap times without touching his smartphone. And as illustrated in Figure 9, a context-sensitive ubicut could be created for instance onto the user's shoe to trigger an SMS reply message to anyone calling during an important meeting.

Child-computer interaction
Ubicuts can be used for creating mnemonics based on arrangement of physical objects such as toys. A child could set toys in various arrangements to trigger funny reactions on the smartphone or to call parents (Figure 10). This could be a workaround to tangible interaction as Blackwell's positional, orientational, temporal, structural, material variables and surface markings [12] could be detected by natural feature tracking of objects arranged in a specific way. This, however, is limited only to static moments of interaction and object arrangements that can be photographed after the act.

Personalized activity sensing
Semi-automatic activity sensing can be implemented using ubicuts that trigger labeled messages that act as an activity classification mechanism (see [13]). It would require a large number of ubicuts, but it would not produce much raw data nor it would require physical tags or sensors if a wearable video camera is used. This approach would avoid the challenges in raw data filtering and processing required in most activity sensing systems. As every ubicut is user-defined, privacy concerns are minimal by default. For example, when the user reaches toward newly printed papers on an office printer, an entry is created, as illustrated in Figure 11. In this way, ubicuts could be placed strategically around the user's environment to trigger specific entries to the activity stream for instance for the purpose of senior home monitoring. On the other hand, if ubicuts are shared among a certain community, e.g. in a old folks home, a sensing platform is rather fast to set up. This approach, however, is not fully invisible to the user as some conscious effort would be needed to trigger each ubicut.

5

Figure 10: Ubicuts for tangible interaction. A child using soft toy arrangements for calling parents by pointing one toy and triggering funny reactions on phone with three of them.

Figure 11: Activity sensing application of ubicuts that create an activity stream entry from the printer's logo every time the user reaches for newly printed papers.

Conclusion and Future Work

We presented Ubiquitous Shortcuts, a method for creating mnemonics in the user's static environment and its proof-of-concept implementation that enables use of some of smartphone's functions by free-air pointing. Ubicuts can be used as a workaround to location-based and context-sensitive services, activity recognition and child-computer interaction. Ubicuts could be a generic method for using some of the Ni & Baudisch's "disappearing mobile devices" [1].

We are currently investigating how the user could enhance the image matching process by defining which features would fit which context, i.e. shapes regardless of color. This would enable more generic ubicuts that would work also with scenes that have not been previously photographed. Empirical testing of the method's subtlety, social acceptance and general usability are essential parts of our future work.

Acknowledgements

References

[1] Ni, T. and Baudisch, P. Disappearing mobile devices. In *Proc. UIST 2009*, 101-110.

[2] Tan, D., Morris, D. and Saponas, T.S. Interfaces on the go. *XRDS 16, 4 (June 2010)*, 30-34.

[3] Ljungstrand, P. and Holmquist, L.E. 1999. WebStickers: using physical objects as WWW bookmarks. In *CHI EA '99*, 332-333.

[4] Williamson, J.R., Crossan, A. and Brewster, S. Multimodal mobile interactions: usability studies in real world settings. In *Proc. ICMI 2011*, 361-368.

[5] Costanza, E., Inverso, S.A., Allen, R. and Maes, P. Intimate interfaces in action: assessing the usability and subtlety of EMG-based motionless gestures. In *Proc. CHI 2007*, 819-828.

[6] Montero, C.S., Alexander, J., Marshall, M.T. and Subramanian, S. Would you do that?: understanding social acceptance of gestural interfaces. In *Proc MobileHCI 2010*, 275-278.

[7] Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proc CHI 2005*, 919-928.

[8] van Dyk, T., Renaud, K. and van Biljon, J. Moses — method for selecting senior mobile phones: supporting design & choice for the elderly. In *Proc. SAICSIT 2012*, 277-285.

[9] Nanayakkara, S., Shilkrot, R. and Maes, P. EyeRing: a finger-worn assistant. In *Proc CHI EA* 2012, 1961-1966.

[10] Ängeslevä, J., O'Modhrain, S., Oakley, I., Hughes, S. Body mnemonics. In *Proc. PI 2003 Workshop on Real World User Interfaces*.

[11] Fukumoto, M. and Suenaga, Y. 1994. "FingeRing": a full-time wearable interface. In *Proc. CHI '94*, 81-82.

[12] Blackwell, A.F., Cognitive dimensions of tangible programming techniques. In *Proc. of PPIG 2003*, 391-405.

[13] Choudhury, T., Consolvo, S., Harrison, B., LaMarca, A., LeGrand, L., Rahimi, A., Rea, A., Borriello, G., Hemingway, B., Klasnja, P., Koscher, K., Landay, J., Lester, J., Wyatt, D. and Haehnel, D. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing, 7(2) (April-June 2008)*, 32-41.