SmartFinger: An Augmented Finger as a Seamless 'Channel' between Digital and Physical Objects

Shanaka Ransiri [1]
shanaka@sutd.edu.sg

Suranga Nanayakkara [1]
suranga@sutd.edu.sg

[1] Singapore University of Technology and Design, 20 Dover Drive, Singapore, 138682

ABSTRACT
Connecting devices in the digital domain for exchanging data is an essential task in everyday life. Additionally, our physical surrounding is full of valuable visual information. However, existing approaches for transferring digital content and extracting information from physical objects require separate equipment. SmartFinger aims to create a seamless 'channel' between digital devices and physical surrounding by using a finger-worn vision based system. It is an always available and intuitive interface for 'grasping' and semantically analyzing visual content from physical objects as well as sharing media between digital devices. We hope that SmartFinger will lead to seamless digital information 'channel' among all entities with a semblance in the physical and digital worlds.

Categories and Subject Descriptors
H.5.2. [**User Interfaces**]; Interaction styles, Input devices and strategies, Graphical user interfaces (GUI), Screen design

General Terms
Human Factors

Keywords
Finger augmentation, Media sharing, Smart copy-paste, Device identification, Intuitive Interface.

1. INTRODUCTION
We live in a ubiquitous computing era, where the use of multiple personal digital devices such as computers, smart phones and tablet computers plays a significant role in our daily lives. As a result, connecting these devices seamlessly has become more important. Furthermore, extracting information from the physical world (such as taking a photo, copying information from a book etc.) is a common task [17]. As such, intuitive ways of extracting data associated with physical objects into the digital domain remains an important challenge.

Sensors such as cameras, microphones, accelerometers, Global Positioning Systems (GPS) etc. are integrated into digital devices for understanding the characteristics and properties of the physical environment we live in. Existing systems designed to extract information from the physical world [13, 14], and to transfer data between digital devices [11] require specially designed equipment for each purpose. However, we notice that most of the objects in the physical and digital worlds have a unique appearance. Motivated by this observation, we decided to use a camera as the primary sensor to extract and assemble necessary information from physical and digital worlds. Our research focuses on exploring methodologies to create a 'channel' that seamlessly links personal digital devices and physical objects. This would ideally be an immediate and always accessible interface, without creating an extra burden to the user. In this paper, we describe related work, details of our proposed prototype: SmartFinger, and SmartFinger enabled interactions.

AH'13, March 07 - 08 2013, Stuttgart, Germany

2. RELATED WORK
Commercial products such as *2-D Digital Scanners* (graphic based information), *Digital cameras* (light and color), *Microphones* (sound) and *Radar* (distance) are available to capture physical information from the environment into the digital domain. *I/O Brush* [13, 14] has been developed to extract color and texture information from physical objects. *Magic Finger* [20] also extracts textural data on physical surfaces through a finger-worn micro camera, and allows a user to carry out actions based on the texture of a surface. *Slurp* [21] provides a Tangible User Interface (TUI) to extract digital information from physical objects, and transfer them to the digital domain with the assumption that in the future, every real object is created with a digital body attached to it. Linking physical documents with its digital content by a smart-phone camera using embedded hotspots [4], and feature points belong to textual and graphical content [1] have been studied comprehensively. *PACER* [6, 7] is a platform which links a physical document to its digital version by matching visual features captured through a smart-phone camera. It allows a user to interactively handle content of the document through pre-defined hotspots in the document. *Embedded Media Markers (EMM)* [8] are icons printed on physical document to provide a way to extract digital content associated with those marks through a smart-phone camera. The local features of the captured image containing the *EMM* are matched with a database in order to retrieve its associated digital content. Gormish *et al.* studied the capturing of information from white boards [2] into the digital domain. In this work, we want to augment the finger with a finger-worn camera [10] as an always available input device to capture information.

Traditional hardware solutions such as *Floppy Drive*, *Universal Serial Bus (USB) sticks* and *External Hard Disk Drives* have existed for a long time as a method for transferring digital data between multiple digital devices. However, these removable media devices have limited capacity and speed [18]. Synchronizing data between a smart mobile device and a Personal Computer (PC), has been traditionally achieved through the use of *USB cables*. *Bluetooth* and *Infrared* were subsequently developed as a method for wireless data transfer [12, 15]. In addition to these widely recognized standards, some researchers have also come up with novel methods of performing these tasks (*Slurp* [21], *mediaBlock* [18], *Stitching* [3] and *Tool device* [5]). *Pick-and-Drop* [11] emulates the drag-and-drop interaction for transferring digital information over multiple computer interfaces using a stylus-like device. *Tool device* [5] extends Pick-and-Drop by augmenting the hardware device with haptic and thermal feedback as a way of conveying the quantity and the 'freshness' of its acquired data. *mediaBlock* [18] is another TUI that consists of electronically tagged wooden blocks that are used for transferring multimedia files among computers. In terms of software-only solutions for sharing digital media between devices, *File Transfer Protocol (FTP) Server*, *Media Server*, *Dropbox* and *Google Drive* are well known. *SPARSH* [9] takes on an interactive approach of transferring data through the idea of conceptually 'copying and storing data on the human body', and transferring information 'into and out of the body' by touching the devices with the index finger. *Bump*[1] is an application developed for smart phones to exchange information such as contacts, images and applications. When two devices are 'bumped' together closely, the application starts to exchange selected data over the network. *Swÿp*[2] is developed to transfer data between two touch screen enabled devices by a swiping gesture out of the edge of a device's screen and into the screen of another device.

Most of the above mentioned systems are designed either to extract data from the physical world [2, 6, 7, 8, 13, 14, 20] or transfer media between digital devices [3, 5, 9, 11, 18]. Some of those applications use extensive dedicated hardware components such as IR nodes [12, 21], RFID tags [19], stylus like accessories [3, 11, 16], communication techniques like Ultrasound (*JAJA*[3]) and intermediate tools for interfacing digital ports with different protocols [18]. Few methodologies such as *Bump*[1] and *Swÿp*[2] perform pairing between digital devices based on network location information and account information such Google or Facebook. As a result, those methods could be unreliable in a situation where four or more users at same location try to initiate a paring between devices. Some systems are designed to work with futuristic scenarios such as every physical object will be created with an attached digital body [21]. We wanted to introduce a method useful for present-day which can be incorporated as a solution for both extracting information in our physical surroundings, as well as connecting to digital world seamlessly.

3. SMARTFINGER

SmartFinger uses a finger worn camera device that leverages on the pointing gesture [10]. It is an always available and intuitive interface; with that we aim to create a seamless 'channel' for capturing visual information from the physical surrounding into the digital domain, and for sharing data between digital devices. We use the metaphorical concept of 'channel', as our system provides a continuous and unified information stream between physical and digital objects. SmartFinger enable users to interact with the physical world and digital world concurrently.

3.1 SmartFinger Enabled Scenario

Jessica is writing a case study for her biology class about a bacterium that is immensely harmful for willow trees. She goes to the library to read more about the anatomy of this bacterium. She points her SmartFinger to capture some relevant information found in books. On her way back from the library, she sees a bacteria infected willow tree. As this is an important example for her assignment, she points at the tree with SmartFinger to capture that information. While meeting her friend Kevin, Jessica sees some relevant research articles on bacteria and its repercussions in his tablet. She again uses SmartFinger to create a virtual connection between her mobile phone and Kevin's tablet, copying the research articles to her mobile phone. In return, Jessica transfers some of pictures of the infected willow tree to Kevin's tablet. Jessica goes back home and selectively paste the copied content at appropriate locations; (1) points at the case study document with SmartFinger to paste text content extracted from the library books, and images of the infected tree; (2) points at a folder in her tablet PC to paste research articles she collected from Kevin.

3.2 Interactions and Implementation

The overall architecture of the SmartFinger system is illustrated in Figure 1. We discuss the following four interactions associated with SmartFinger and their corresponding implementations:

1) Linking digital devices
2) Copying and pasting digital data
3) Capturing information from physical objects
4) Interaction between SmartFingers

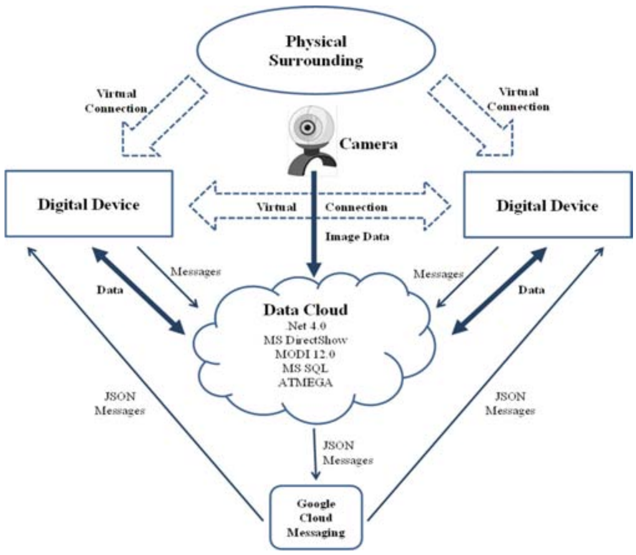

Figure 1. Architecture of the SmartFinger system.

Current implementation consists of (a) a standard webcam and a push button switch attached to the index finger using Velcro; (b) a backend server implemented on a computer running Windows 7 64-bit. Each camera has its own unique ID for differentiating itself from other cameras and acts as an input device, providing an image stream of the object/scene pointed at. The backend server is accountable for (i) managing the data cloud and device data, (ii) capturing the button events to operate the camera, (iii) communication with Google Cloud Messaging (GCM) service, and (iv) segmentation and extraction of information from the image stream.

3.2.1 Linking Digital Devices

SmartFinger generates a unique semblance and blends it with the existing appearance of a digital device for embodying it distinctively. This allows the system to recognize the device uniquely, and to produce a virtual association between devices. Thus, it eliminates the necessity of attaching external identifies to the device. Ideally, a gestural interaction initiated by a user, generates a unique appearance which is then captured by SmartFinger and processed to discern the device. This is illustrated in figure 2.

[1] https://bu.mp/
[2] http://fluid.media.mit.edu/projects/sw%C3%BFp
[3] http://hex3.co/products/jaja

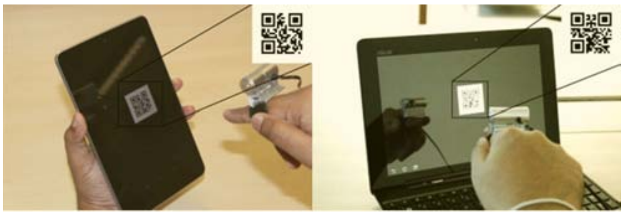

Figure 2. QR code appears with the long press.

In the current implementation, the SmartFinger client application was developed in two tablets running on Android OS 4.0. When the SmartFinger Android application is installed in the device, the application registers with the GCM service, and generates a Quick Response (QR) code based on the GCM registration details. The GCM registration data is then transmitted to a server program (developed using C# and .Net Framework 4) through a Transmission Control Protocol (TCP) connection, and stored in a Microsoft Structured English query Language (SQL) Server database. A push button switch on SmartFinger provides various user input actions. The push button states are sent to the server via the serial port. A double click action on the push button switch starts the video capturing. Microsoft DirectShow API was used to capture image data from the webcam. A long press gesture on the SmartFinger application window displays a QR code on the screen (Figure 2). The server analyses the image stream from the camera that contains the QR code to identify the digital device. Subsequently, the server sends a JavaScript Object Notation (JSON) message to the device through GCM. The application vibrates and changes the brightness of the screen on the device to indicate a successful identification.

3.2.2 Copying and Pasting Digital Data

This interaction allows a user to copy data from one 'information container' to another once two devices are linked together through SmartFinger. A touch event initiated by the user on any file located on the touchscreen of the device produces a replicated instance in the cloud. A subsequent touch event (on a folder or another media file) on another device then opens up a preview of the available data (semantically extracted data from previously captured images) for pasting. This allows the user to paste the data selectively according to their preference.

In the current implementation, the copied data is transmitted to the cloud via TCP and tagged to a specific ID associated with the device. Images and documents can be copied as a single file. Moreover, the user is allowed to copy only the text content in a document. To transmit data to a second device, a user performs a long press on the SmartFinger application window or an opened document. The user will then see a preview of the data and be given the option to paste it on any location on the screen. Figure 3 illustrates both copying and pasting actions.

3.2.3 Capturing Images of Physical Objects

SmartFinger provides a methodology to extract contents from the physical environment into the digital domain through a vision-based system. The system semantically analyses the visual content and allows the user to selectively 'grasp' and identify information such as textual data, human presence, colors/textures from objects etc.

In the current implementation, the user points SmartFinger at an object and performs the single click as shown in Figure 4. The backend system identifies the single click via the serial port and captures the visual data. If the captured image consists of textual content, both the original image and its extracted textual content are stored in the cloud under the camera's device ID. For Optical Character Recognition (OCR), in order to extract textual data in captured content, we used Microsoft Office Document Imaging (MODI) Library 12.0.

3.2.4 Interaction between SmartFingers

This interaction is currently conceptual and being implemented. Users can tap their SmartFingers against each other to establish a connection between them. This allows them to exchange data without having to use an intermediate device. We are exploring integrating different technologies such as IR and ultrasound into future camera devices. This would then enable device to device paring and exchanging of camera ID for transferring data through the cloud.

4. DISCUSSION & FUTURE WORK

SmartFinger focuses on creating a seamless 'channel' between digital and physical objects. One of its purposes is to uniquely identify any object in the physical and digital domain using visual information. Digital contents are physically tagged to the user's SmartFinger, thus avoiding tedious login/logout functions to distinguish between different users (e.g. *SPARSH* [9]). SmartFinger also allows a user to paste contents selectively; it understands the context and shows appropriate paste-able data. For example, when a user touches on a blank space of a document, SmartFinger only shows images and text in the

(a) Copying images (b) Preview of images before pasting

(c) Copying documents (d) Preview of documents before pasting

(e) Copying text from a document (f) Preview of text before pasting on a document

Figure 3. Copying data to the cloud and pasting selectively.

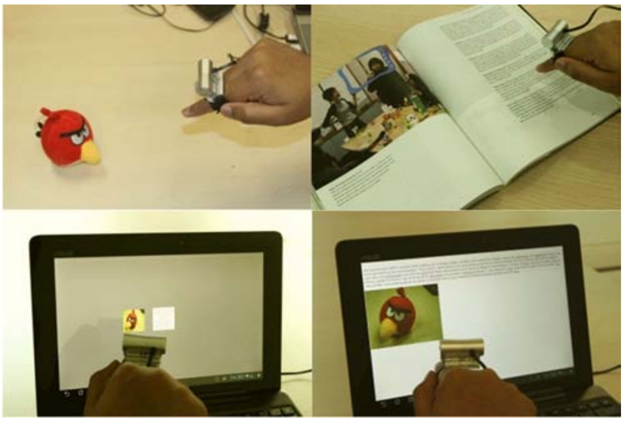

Figure 4. Capturing visual information form physical environment.

preview. It will not show a file in the preview as this is a non paste-able item into a document.

In the current prototype, we generate a unique QR code on the screen of a digital device as a unique identifier. However, ideally, SmartFinger should be able to identify a device based on its unique personal features (user's finger prints, personal sticker, etc.) on a device. In addition, it should be capable of recognizing physical objects via an attached physical tags (e.g. barcode) or visual feature points associated with the object. We did not focus on implementing a customized wireless camera system as it might not have provided a significant contribution at this point [20]. However, a wireless-camera based SmartFinger would be a more portable and convenient product. The current system only extracts images and text from an image. We will work on segregating information semantically and retrieving further data from the Internet with respect to both the identified physical objects and segregated visual information. Furthermore, a user study will be conducted to investigate the current prototype and to get more insights into how people would interact with this kind of seamless connection between physical and digital worlds.

5. CONCLUSION

We have proposed an always available and intuitive system, which allows users to interact with the physical and digital worlds seamlessly. It creates a metaphorical digital information 'channel' for exchanging data between digital devices and excerpting visual information from the physical world into digital domain. We describe four interactions enabled through SmartFinger: linking digital devices, copying and pasting digital data, capturing images of physical objects and interaction between SmartFingers. Finally, we discussed the limitation of the system and future enhancements. Despite some limitations of the current implementation, we believe SmartFinger is a step towards creating an always available seamless 'channel' between digital and physical worlds.

6. ACKNOWLEDGMENT

The International Design Center (IDC) of the Singapore University of Technology and Design (SUTD) has supported this work with IDC grants IDG31100104 and IDD41100102.

7. REFERENCE

[1] Erol B., Graham J., Antúnez E., and Hull J.J. 2008. HOTPAPER demonstration: multimedia interaction with paper using mobile phones. In *Proc. MM'08*, pp.983-984.

[2] Gormish M. J., Erol B., Daniel G., Olst V., Li T., and Mariotti A. 2011. Whiteboard sharing: capture, process, and print or email. In *Proc. SPIE'11 Electronic Imaging*.

[3] Hinckley K., Ramos G., Guimbretiere F., Baudisch P., and Smith M. 2004. Stitching: pen gestures that span multiple displays. In *Proc. AVI'04*, pp.23-31.

[4] Hull J.J., Erol B., Graham J., Ke Q., Kishi H., Moraleda J., and Olst D.G.V. 2007. Paper-Based Augmented Reality. In *Proc. ICAT '07*, pp.205-209.

[5] Ikeda Y., Kimura A. and Sato K. 2003. TOOL DEVICE: Handy Haptic Feedback Devices Imitating Every Day Tools. In *Proc. HCI International'03*, pp.661-665.

[6] Liao C., and Liu Q. 2009. PACER: Toward a Cameraphone-based Paper Interface for Fine-grained and Flexible Interactions with Documents. In *Proc. Multimedia'09*, pp. 969-970.

[7] Liao C., Liu Q., Liew B., and Wilcox L. 2010. Pacer: fine-grained interactive paper via camera-touch hybrid gestures on a cell phone. In *Proc. CHI'10*, pp.2441-2450.

[8] Liu Q., Liao C., Wilcox L., Dunnigan A., and Liew B. 2010. Embedded Media Markers: Marks on Paper that Signify Associated Media. In *Proc. IUI'10*, pp.149-158.

[9] Mistry P., Nanayakkara S., and Maes P. 2011. Touch and copy, touch and paste. In *Proc. CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp.1095-1098.

[10] Nanayakkara S., Shilkrot R., and Maes P. 2012. EyeRing: a finger-worn assistant. In *Proc. CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp.1961-1966.

[11] Rekimoto J. 1997. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proc. UIST'97*, pp.31-39.

[12] Rekimoto J., Ayatsuka Y., and Hayashi K. 1998. Augment-able Reality: Situated Communication through Physical and Digital Spaces. In *Proc. ISWC'98*, pp.68- 75.

[13] Ryokai K., Marti S., Ishii H. 2004. I/O Brush: Drawing with Everyday Objects as Ink. In *Proc. CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pp.303-310.

[14] Ryokai R., Marti S., and Ishii H. 2005. Designing the world as your palette. In *Proc. CHI'05 Extended Abstracts on Human Factors in Computing Systems*, pp.1037-1049.

[15] Sairam K., Gunasekaran N., and Redd S.R. Bluetooth in wireless communication. IEEE Communications Magazine. 40, 6, pp.90-96, 2002.

[16] Song H., Benko H., Guimbretiere F., Izadi S., Cao X., and Hinckley K. 2011. Grips and gestures on a multi-touch pen. In *Proc. CHI'11*, pp.1323-1332.

[17] Thakur A., Gormish M., and Erol B. 2011. Mobile Phones and Information Capture in the Workplace. In *Proc. CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 1513-1518.

[18] Ullmer B., Ishii H. and Glas D. 1998. mediaBlocks: physical containers, transports, and controls for online media. In *Proc. SIGGRAPH'98*, pp.379-386.

[19] Want R., Fishkin K.P., Gujar A., and Harrison B.L. 1999. Bridging physical and virtual worlds with electronic tags. In *Proc. CHI'99*, pp.370-377.

[20] Yang X., Grossman T., Wigdor D., and Fitzmaurice G. 2012. Magic finger: always-available input through finger instrumentation. In *Proc. UIST'12*, pp.147-156.

[21] Zigelbaum J., Kumpf A., Vazquez A., and Ishii H. 2008. Slurp: tangibility spatiality and an eyedropper. In *Proc. CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pp.2565-2574.