Swiss Army Knife meets Camera Phone: Tool Selection and Interaction using Visual Markers

Christian Kray
Informatics Research Institute
Newcastle University, UK
c.kray@ncl.ac.uk

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin, Germany
michael.rohs@telekom.de

ABSTRACT

A key issue in ubiquitous computing in general and public display research in particular is how to enable interaction. Oftentimes, it is not clear how users can interact with a system and what functionality it provides. In the case of public displays, several methods have been suggested such as touch-enabled surfaces, gesture recognition, voice input, or text messaging. However, all these methods have some inherent flaws such as being unreliable, limiting the number of concurrent users or requiring complex configuration. In this paper, we introduce a novel approach based on visual markers that users can photograph using their mobile phones. By displaying the marker snapshots on the screen of their phone and moving them over a public display, an external camera can track the position of the device and identify the marker being displayed. We introduce a prototype making use of this idea, and highlight key benefits of our approach such as no need to install custom software on the phone, the elimination of network configuration and exposure of system functionality.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*input devices and strategies, interaction styles*

General Terms

Design, Human Factors

Keywords

situated interaction, mobile phones, marker-based tracking, public displays, large displays, visual communication

1. INTRODUCTION

In recent years, the technical infrastructure surrounding us and the number of devices/services at our disposal have been growing at a steady pace. This trend towards ubiquitous computing raises a number of research questions, which have not yet been fully addressed. For example:

- how can we discover what services are available to us in a particular ubiquitous environment?
- how can we interact with a ubiquitous infrastructure?
- how can we make this interaction intuitive and accessible to untrained users?

In this paper, we discuss issues related to these questions in the context of a specific type of ubiquitous technology, namely public displays, which are rapidly proliferating in areas such as airports, transport hubs, and shopping malls. We introduce a particular interaction technique based on *purely visual* communication using mobile phones that are equipped with a camera. Figure 1 outlines the basic idea underlying our approach. The remainder of the paper is structured as follows. In the following section, we briefly discuss related work, and then describe our approach in detail. We will demonstrate its feasibility in Section 4 using an example application, and will discuss benefits and drawbacks of our approach in more detail in Section 5. A brief summary of the main contributions will conclude the paper.

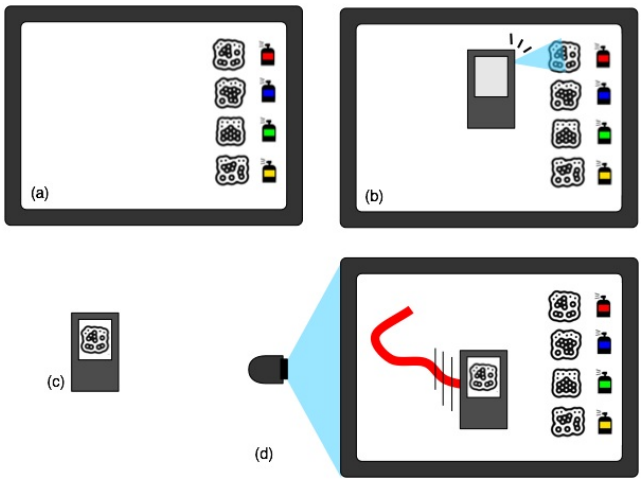

Figure 1: Basic idea: (a) toolbar displayed on public screen alongside corresponding visual markers (b) user takes photograph of marker next to tool to be used (c) user displays photograph of marker on phone screen (d) tool activates when phone is brought in front of public screen and into the field of view of the associated camera.

2. RELATED WORK

Users can interact with a (public) display in different ways. The traditional approach is to use a keyboard and a mouse (trackpad, trackball). More recent approaches include voice or gesture recognition, or a combination of these [7]; touch-enabled surfaces are another option [2]. If the public screen

is a tabletop display, custom-made tokens [4] or custom-made devices are another option [3]. Frequently, personal devices such as mobile phones are also used to interact with a public display (see [1] for a survey). One common way to realize this is through a network connection between the phone and the public display, e. g. via Bluetooth, wireless LAN (802.11) or infrared. If interaction is asynchronous, mechanisms such as text messaging (SMS) can be used. Some interaction mechanisms make use of an external or built-in camera. The built-in camera of a mobile device can help to sense the optical flow resulting from moving the device in space while the camera is recording video footage. Oftentimes, visual markers are used either in conjunction with this approach or on their own. Optical markers being shown on a large public display can be recognized using the built-in camera of a mobile phone, and help to select an item or to track the (relative) location of the mobile device with respect to the marker/large display [1]. The system most closely related to the approach discussed in this paper is C-Blink [6]. It enables mobile phone-public display interaction through a visual marker (i. e. a particular sequence of colors), which is shown on a the screen of a mobile device, and an external camera that tracks these markers. There are however several key differences compared to our system, e. g. C-Blink relies on custom software being installed on the mobile device and does not provide its users with a clear idea of the available functionality or the current state of the system. One advantage of C-Blink over our approach is that it only requires a display on the mobile device whereas our approach also depends on a camera to be present.

Without embarking on a fine-grained analysis of the alternative methods to enable interaction with a public display, it is safe to say that they all have a number of drawbacks. These range from requiring a lot of processing power (e. g. speech recognition) and the need for custom-made hardware (e. g. active tokens) to privacy problems (e. g. tracking facial expressions). Some approaches require extensive configuration, for example, to set up a network connection or to download/install a custom piece of software in order to be able to interact. Furthermore, some systems do not support multi-user interaction (e. g. most touchscreens), while others are fully opaque, i. e. users do not know what services are available to them and how to use them (e. g. speech recognition). Our approach addresses several of these problems: it does not require any custom-made hardware; the only requirements for a mobile device are that it features a display and a camera. We do not require a network connection or specific software on the mobile device either. Hence, a standard camera phone can serve as an interaction device straight away. Furthermore, our approach is user independent and inherently supports the simultaneous interaction of multiple parties. Another key advantage is its immediacy: users can easily tell what functions are available and which one they are currently using. Finally, even though we rely on an external camera, it is possible to anonymously interact with the system.

3. VISUAL INTERACTION

The basic concept behind the approach proposed in this paper is to use a *purely visual channel* for communication between a mobile device and a ubiquitous infrastructure. While we will focus on the interaction between a mobile phone and a public screen, the principle can be applied to other device combinations as well.

One challenge in ubiquitous computing is how to make users aware of available services and how they can be used. In the case of a public display, an obvious solution to this problem is to simply display a list of available functions on the screen. Figure 1a shows an example in the context of a painting application (cf. Section 4), where spray cans for different colors symbolize the different tools available in the application. Note that next to the actual tool, visual markers are displayed (in our proof of concept prototype, we are using the reacTVision toolkit and the corresponding markers at the moment [5]).

Using the visual display of available functions on the screen, users can immediately select the tool they want to use by taking a photograph of the associated marker (Figure 1b). They can then use the built-in photo browsing application of their mobile phone to show this marker on the display of their phone (Figure 1c). If they have taken pictures of several markers they can select the tool they want to use by flicking through the corresponding photographs that are stored on their mobile phone. In the example case shown in Figure 1, the markers and tools being displayed on the public display can serve as a visual aid to identify which function is currently selected on the phone.

Once the display on the mobile phone shows a marker associated with an application-specific function, it will activate the function while it is in the field of view of a camera pointed at the public screen. For example, in the graffiti application we describe in Section 4, the phone will become a spray can (see Figure 1d). To de-activate the tool, a user can either move the phone out of the camera view, change the content of the screen on the phone, cover the screen of the phone or tilt the phone (so that the camera can no longer track the marker). To switch between different functions on the fly, users can either take a photograph of another tool or select a marker corresponding to the desired function from a list of previously photographed markers.

4. EXAMPLE APPLICATION

In order to explore and demonstrate the idea of a purely visual means of interaction between a mobile phone and a public display, we built a simple example application (see Figure 2). It enables users to 'spray' paint onto a virtual canvas using their mobile phones. In the figure, the canvas is shown in the lower left hand area of the plasma screen. On the right hand side of the screen, spray cans for different colors are shown alongside visual markers [5]. To select a particular color, users can photograph the marker shown next to it using their mobile phones (e. g. the topmost one to select red paint). In order to spray paint on the canvas, a user displays the marker corresponding to the desired color on the screen of their mobile phone. Once the marker is visible on the phone's display, moving the phone inside the canvas area will activate the spray can and the user can paint by moving their phone in front of the canvas (see Figure 2). To stop spraying users can either cover the phone screen, remove the marker on the phone display, move the phone out of the canvas area or move it in a way so that the phone is not parallel to the canvas anymore.

The application was built by modifying software created by the reactable project, in particular the reacTIVision toolkit [5]. The hardware setup consisted of a 50" plasma screen, a Nokia N95 mobile phone (display size: 40×55mm), an

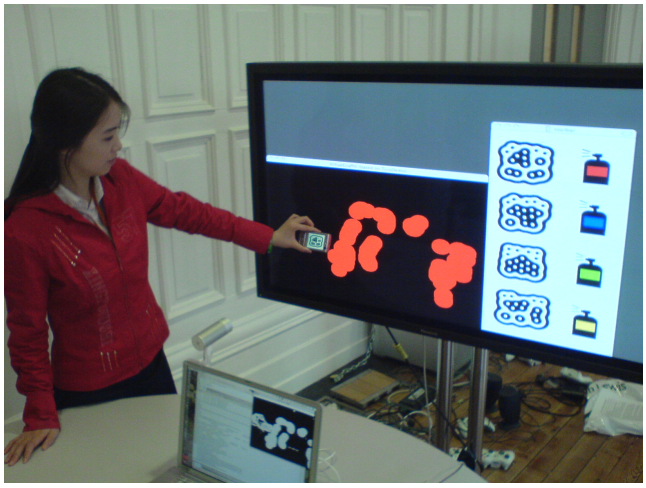

Figure 2: Virtual Graffiti prototype: users can spray paint on a virtual canvas using their mobile phone; photographing a marker next to spray can and then displaying it on the screen selects the color.

Apple iSight webcam (resolution: 640 × 480 pixel, 30 fps) and a G4 PowerBook (1.67Ghz G4, 1GB RAM). The camera was positioned in front of the display (ca. 40-50cm away) and was able to cover about a third of the display area of the plasma screen (i. e. the canvas area in Figure 2). Both the plasma display and the camera were connected to the PowerBook. No custom software was installed on the phone; we relied solely on the built-in capture and photo browsing applications.

Using this setup (even though it was far from ideal) we were able to paint on the screen when the phone displayed the photographed markers on its screen. While our prototype illustrates that the basic idea of purely visual, marker-based interaction is feasible, there are several aspects, which can be significantly improved in future versions. Most importantly, the reacTIVision toolkit is not optimized for this setup (we mainly used it because it is easy to quickly build applications). It was developed to handle a lot of different markers (89 are provided with the software), which is more than would be needed for mobile phone-public display interaction. Reducing the number of different markers would make the recognition more reliable and can also help to make them smaller (thereby affording a larger webcam-to-screen distance). In our experience, the reacTIVision toolkit is also not very forgiving with respect to tilting: when phones are not held parallel to the canvas/camera plane, the recognition rate drops sharply (it was built for use with tabletop applications, where it is safe to assume that objects are aligned with the tabletop due to gravity).

In addition, if markers were iconic representations of the function they are associated with, interaction would be more intuitive as users would not have to take photographs of abstract markers but of the actual tool itself (see Figure 3 for an example design). In order to cover a larger area and to improve recognition, a higher-resolution camera (or multiple cameras) would be beneficial.

5. DISCUSSION

While the prototype application presented in the previous section is fairly simple and could be improved in many ways, it nevertheless illustrate the potential of the basic concept. There are several key benefits that result from using a purely visual communication mechanism based on markers to enable mobile phone-public screen interaction:

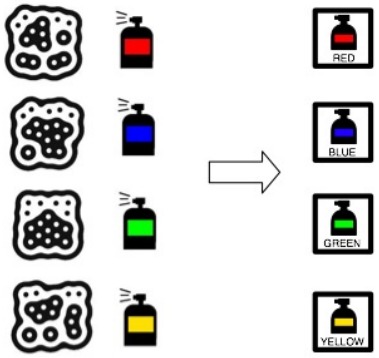

Figure 3: Toolbar exposing available functions: (left) photographing a marker next to a spray can and then displaying it on the phone's screen activates the corresponding color, (right) design for merging markers and tools.

- **configuration-free operation**
  As the system does not require any network connection between the public display and the mobile device, it is not necessary to configure any network settings (unlike Bluetooth or WLAN based approaches). In addition, as there is no need for any special hardware (such as accelerometer or ultrasound sensors often required by custom-made interaction devices) besides the built-in camera, no configuration of such add-ons is required.

- **no custom client software**
  The proposed approach can be implemented using solely the software already available on a camera-equipped device, i. e. the capture and photo-browsing application. Users thus do not need to familiarize themselves with any new software (as they frequently have to using, for example, Bluetooth based systems).

- **inherent support for multi-party interaction**
  The camera(s) can easily track multiple markers at the same time - thus enabling multiple people to interact simultaneously (in contrast to most touchscreen solutions). The reacTIVision toolkit we used to build the prototypical application was designed to track multiple markers simultaneously. However, camera placement with respect to the screen, mobile phones and users needs further investigation, i. e. regarding occlusion issues resulting from more than one user.

- **transparency of interaction**
  Through the markers/icons being displayed (see Figure 3, right) it is immediately apparent to the users what functions are available to them (which is frequently not the case with voice-based interaction). We would argue that there is a fairly straightforward correspondence between 'picking up a tool' and taking a photograph of the icon representing the tool (which

can be a problem with gesture-based systems). The same can be said for activating a tool by displaying it on the screen of the mobile device.[1] A beneficial side effect of this is the fact that users can always tell which tool their mobile device currently incorporates by looking at its screen. Moreover, the large display could highlight the currently selected tool.

- **anonymous interaction**
  Since there is no network connection between the mobile device and the public screen, interaction is by default anonymous, i. e. the system cannot determine *who* is interacting – it can only tell which tools are active at any given point in time. Many alternative approaches such as those using text messages or Bluetooth disclose at least the identity of the device to the display software.

- **ownership/control/empowerment**
  By relying on unmodified mobile phones, our approach enables users to employ their personal devices in a way that does not require them to give up control: they do not need to install any custom software and they do not have to connect to a third party server. In addition, they can transfer their knowledge about their personal devices (e. g. how to take photographs with them) to control an application running on the public screen. Other systems oftentimes require users to learn new mechanisms such as a set of gestures or voice commands.

There are also a few open questions that need further investigation. In addition to the technical shortcomings discussed in Section 4, there is a need to research the particular properties of tracking visual markers being displayed on the screen of a mobile phone (e. g. the impact of reflections, robustness against tilting, optimization for being photographed). The same applies for evaluating the approach with a larger number of users (e. g. to determine a suitable set of markers, to identify appropriate metaphors, to provide interaction-related feedback). Furthermore, different hardware setups need to be explored, in particular different display technologies (e. g. back/front projection) and their properties as well as camera configurations (e. g. multiple cameras, different camera positions).

6. SUMMARY

In this paper, we have introduced a purely visual mechanism to enable interaction between mobile devices and ubiquitous infrastructure in general and public displays in particular. It is based on visual markers that users photograph using their mobile devices and then display on the screen of these devices to activate a particular tool. We briefly described a prototypical drawing application using this approach, and outlined key benefits of this form of interaction. The advantages include ease of use, configuration-free operation, multi-party interaction as well as the preservation of anonymity. Based on this initial research, we attribute considerable future potential to the basic idea of purely visual communication. We will hence further investigate both technical aspects (e. g. how to optimize the markers and the display-camera setup) and issues related to interaction (e. g. suitable metaphors, user studies with a set of applications).

[1] We have not yet conducted proper user studies to test these conjectures but intend to do so in the near future.

7. ACKNOWLEDGMENTS

We would like to acknowledge a hardware grant by Nokia Research Labs, Finland that provided the N95 handsets used. We would also like to thank the organizers and participants of the FTIR workshop in Münster, Germany, for inspiring discussions. Furthermore, we want to thank the authors of the reacTIVision toolkit for making their software publicly available.

8. REFERENCES

[1] R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 05(1):70–77, 2006.

[2] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In P. Baudisch, M. Czerwinski, and D. Olsen, editors, *Proceedings of UIST 2005*, pages 115–118, New York, 2005. ACM Press.

[3] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn. A relative positioning system for spatial awareness of co-located mobile devices and users. In K. G. Shin, D. Kotz, and B. D. Noble, editors, *Proceedings of the third international conference on mobile systems, applications, and services (MobiSys 2005)*, pages 177–190, New York, 2005. ACM Press.

[4] S. Jordà, M. Kaltenbrunner, G. Geiger, and R. Bencina. The reactable*. In *Proceedings of the International Computer Music Conference (ICMC 2005)*, Barcelona, Spain, 2005.

[5] M. Kaltenbrunner and R. Bencina. reactivision: a computer-vision framework for table-based tangible interaction. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 69–74, New York, NY, USA, 2007. ACM Press.

[6] K. Miyaoku, S. Higashino, and Y. Tonomura. C-blink: a hue-difference-based light signal marker for large screen interaction via any mobile terminal. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 147–156, New York, NY, USA, 2004. ACM Press.

[7] W. Wahlster. Smartkom: Fusion and fission of speech, gestures, and facial expressions. In *Proceedings of the 1st International Workshop on Man-Machine Symbiotic Systems*, pages 213–225, Kyoto, Japan, 2002.