Touching the Untouchables: Vision-based Real-time Interaction with Public Displays through Mobile Touchscreen Devices

Matthias Baldauf, Peter Fröhlich, and Peter Reichl

FTW - Telecommunications Research Center Vienna
Donau-City-Strasse 1, 1220 Vienna, Austria
{baldauf, froehlich, reichl}@ftw.at

Abstract. Omnipresent public displays are mainly passive devices presenting non-interactive slide shows. Techniques to easily and quickly interact with these displays could help mobile users to gather personalized and detailed information. In this paper, we present *VirtualTouch*, a vision-based approach to enable a direct interaction with remote public displays. Applying a mobile phone as a see-through interface, the user is able to control the screen's mouse cursor via the mobile's touch screen in real-time. Thus, our technique touch-enables physically inaccessible displays through the mobile's live video view. We describe the interaction technique, present the design of our current prototype, and finally outline our future work.

1 Introduction

Recently, large public displays have become omnipresent in urban environments. While we are on the move, they inform us about the latest news, the public transport or local bargain offers. Yet, these displays present passive slide shows and do not provide users with any means to fetch more detailed, personalized information. This is because they are either physically not accessible or they do not offer any input hardware - or in many cases even both, e.g. an advertising screen behind a window of a closed shop.

Mobile phones as our steady companions are increasingly investigated mediators to interact with such pervasive devices. Lorenz et al. [5] proposed a general framework which sends key and touch screen events invoked on the mobile phone via WiFi and transforms them to corresponding desktop user interface events. A more direct interaction approach is presented by Seewoonauth et al. [6]. Their technique named *touch & select* is based on Near Field Communication (NFC): the touch of a display by a NFC-enabled phone triggers the exchange of files or folders. The selection of specific content is made possible by several NFC tags arranged in a grid at the display's backside. Alternative vision-based approaches utilize a mobile device's built-in camera and analyze the delivered video stream to infer actions and desired content. For example, Ballagas et al. [1] presented a technique to control a remote mouse cursor by moving the mobile phone or to select on-screen objects with the help of displayed visual markers. Boring et al. [3]

introduced another visual approach to literally gather information from a screen by automatically identifying photographed icons. Only recently, Boring et al. [4] presented an approach most related to the work described in this paper. Their system enables touch interaction with displays through a mobile's live video. The presented architecture assumes the availability of a WiFi connection and relies on a centralized 'environment manager' handling the involved displays.

In this paper, we present a mobile interaction technique named *VirtualTouch* to control public displays. Our approach combines the advantages of existing approaches while omitting their drawbacks: *VirtualTouch* enables a vision-based, real-time screen interaction without any required modification of the displayed content (such as including visual markers) or the involved hardware (such as adding NFC tags). *VirtualTouch* augments the device's live video stream with a control layer: Touches on the device's screen are mapped to the corresponding captured screen positions and events. The underlying idea was first investigated by Tani et al. [7] in the context of stationary monitored industrial plants. Their so-called 'object-oriented video' allowed inspectors to control machines by directly interacting with the live video. Applied on mobile devices, this technique figuratively may open a portable hole through a glass window and touch-enable the screen behind.

2 User Interaction

VirtualTouch touch-enables remote, maybe physically inaccessible displays. A user's mobile is pointed at the target, i.e. the display to interact with, and used as a see-through device comparable to mobile augmented reality use cases. The user then is able to directly touch sensitive areas on the live video such as they would click with a mouse attached to the remote computer (Figure 1).

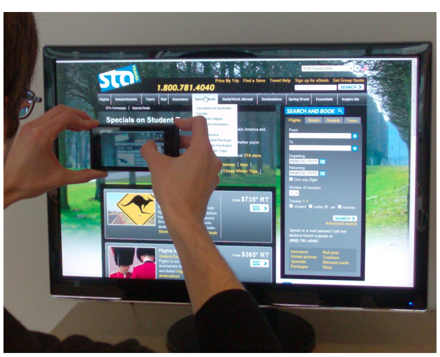

Fig. 1. A user controls the display by touching captured sensitive screen areas.

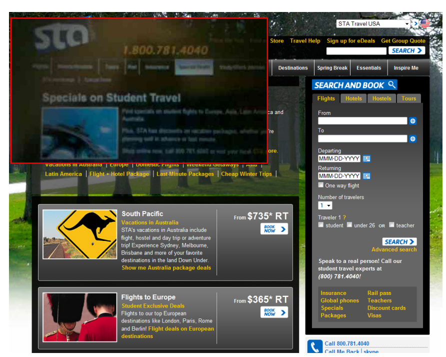

Fig. 2. In the background, the mobile's video is matched with the screen content.

Our technique not only supports simple clicks but also more complex actions such as drag'n'drop operations in real-time. Obviously, the mobile device should be held as still as possible to accurately hit the desired areas. Possible options to ease interacting with the live video include a zoom feature and temporary pausing the live view as suggested in [4]. When using *VirtualTouch* with displays behind shopping windows, the mobile device can be placed directly onto the window and shifted over the surface to have access to different screen areas. Thus, shaking is prevented anyway. Additionally, in this scenario the distance between the mobile phone and the display is constant. Placing the display between five and ten centimeters away from the window was found out to be a reasonable distance during first trials.

3 System Architecture

Our current prototype architecture consists of a server application which is installed on the computer serving the public display as well as a mobile application for a smartphone.

The mobile application is installed on an Android-powered Motorola Milestone smartphone featuring a high resolution touchscreen. On startup, the application triggers a Bluetooth discovery for nearby devices with names ending on *-VTouch* (e.g. MyTravelAgency-VTouch). If one is found, it establishes a connection; if several devices are found, the user is prompted to select the one she wants to interact with according to the meaningful device name. During usage, our application shows the live video as captured by the phone's built-in camera in full screen. When the user touches the screen, the latest video frame is compressed and sent to the server application via Bluetooth followed by a sequence of detected touch events. One touch event consists of the position relative to the smartphone's screen size as well as of the occurred event type (*pressed*, *moved* or *released*).

Our server application is implemented in Java and offers the corresponding Bluetooth service waiting for incoming data. Concurrently, the application periodically captures the public display's current content and precalculates its appropriate feature points applying the SURF algorithm [2]. When the service receives a photo its feature points are also extracted and matched with the latest set of detected screen feature points. Comparing the matching screen/photo feature point pairs in twos, the photo's rotation angle and scaling factor can determined. Using two feature point pairs and the calculated rotation and scaling values, the transmitted photo can be mapped to the corresponding screen portion (Figure 2). Thus, submitted event positions can be converted to screen positions where the attached mouse event then is simulated.

4 Conclusions and Outlook

In this paper, we introduced *VirtualTouch*, an innovative technique to interact in real-time with a remote display through a mobile's camera and its touch screen. This direct interaction with a real-world view provides a novel experience: In contrast to augmented reality applications, the user's view is not only extended with overlaid information but real-world objects can be even modified or controlled through *VirtualTouch*.

Our technical approach focuses on universal applicability and both easy and instant deployment. Due to the used advanced computer vision algorithm, *VirtualTouch* works for arbitrary existing applications shown on public displays and is not limited to specific presentations. The required server component is an autonomous, tiny piece of software and can be easily installed on the computer serving the display. Furthermore, *VirtualTouch* does not require a WiFi or 3G connection but instead works via Bluetooth resulting in no additional costs for the user.

Besides arising usability and efficiency aspects of this novel mobile interaction approach (partly discussed in [4]), we are especially interested in possible new use cases. For example, the basic control interaction can be extended by the integration of personal information stored on the mobile phone. Leaving or fetching contact details or transferring media files by dragging and dropping items between the mobile phone and special areas of the remote display are only a few possible applications.

Acknowledgments. This work has been carried out within the project U-0 financed in parts by the Austrian Government and by the City of Vienna within the competence center program COMET.

References

1. Rafael Ballagas, Michael Rohs, and Jennifer G. Sheridan. Sweep and point and shoot: phonecam-based interactions for large public displays. In *Ext. Abstracts of CHI'05*, pages 1200–1203, 2005.
2. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer Vision ECCV 2006*, Lecture Notes in Computer Science, chapter 32, pages 404–417. 2006.
3. Sebastian Boring, Manuela Altendorfer, Gregor Broll, Otmar Hilliges, and Andreas Butz. Shoot & copy: phonecam-based information transfer from public displays onto mobile phones. In *Proceedings of Mobility'07*, pages 24–31, 2007.
4. Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. Touch projector: Mobile interaction through video. In *Proceedings of CHI'10*, 2010. To appear.
5. Andreas Lorenz, Clara Fernandez De Castro, and Enrico Rukzio. Using handheld devices for mobile interaction with displays in home environments. In *Proceedings of MobileHCI'09*, pages 1–10, 2009.
6. Khoovirajsingh Seewoonauth, Enrico Rukzio, Robert Hardy, and Paul Holleis. Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone. In *Proceedings of MobileHCI'09*, pages 1–9, 2009.
7. Masayuki Tani, Kimiya Yamaashi, Koichiro Tanikoshi, Masayasu Futakawa, and Shinya Tanifuji. Object-oriented video: interaction with real-world objects through live video. In *Proceedings of CHI'92*, pages 593–598, 1992.