Towards Universal, Direct Remote Interaction with Distant Public Displays

Sven Gehring, Florian Daiber, Christian Lander
German Research Center for Artificial Intelligence (DFKI)
Campus D3 2
Saarbrücken, Germany
{sven.gehring, florian daiber, christian.lander}@dfki.de

ABSTRACT
The digital augmentation of urban landscapes with digital public displays rapidly increases. Although they offer great potential for interactivity, the majority of public displays do not allow direct interaction. Prominent reasons are missing touch sensitivity and the placement of the displays since digital public displays are often placed outsides a user's reach or they protected behind a transparent surface. In this paper, we describe a framework that allows universal, direct interaction with distant displays. We describe a vision-based approach to detect rectangular displays within the live camera view on a mobile device. We utilize the TUIO protocol to forward touch input on the display, which appears in the live video, to the actual public display.

Keywords
Mobile device, interaction technique, input device, public display

INTRODUCTION
Nowadays, the digital augmentation of public spaces rapidly proceeds. More and more public displays, video walls and media façades find their way into the urban landscape. Unfortunately, most of these displays do not offer interaction to passersby. Furthermore, they are often placed outside the users reach (see Figure 1) and in case of video walls and media façades, they require a certain viewing distance. Due to their placement and the required viewing distance, interacting with them directly (e.g., by touching them) is normally not possible.

Recent advances in mobile computing allow users to directly interact with such displays and façades through live video on mobile devices. With *live video*, we denote the constantly captured view of the mobile device's built-in camera. In [4], Boring et al. introduce the Touch Projector, which is a vision-based system that allows users to interact with remote screens through live video on their mobile device by direct touch input. They give the user the impression that he directly touches the remote display. The remote display is tracked in the live video and the touch input of the user on the live video is projected onto the display itself. In [5], Boring et al. adapted and extended the Touch Projector interface to accommodate multiple users interacting with a media façade in an urban outdoor setting with dynamic environmental conditions. However, in the Touch Projector approach, the individual frames of the live video on the mobile device are transferred to a server

Figure 1: Public displays in urban environments.

application, in which the tracking of the screen and all computations happen. In terms of generalizability, this approach is rather limited since (1) a dedicated server application is necessary and (2) it demands a certain bandwidth to transfer the video frames from the mobile device to the server.

In this paper, we pick up the concept of Touch Projector and describe a framework, which allows universal interaction with public displays through live video on mobile devices. We use the mobile device as a *Magic Lens* [3] for interacting with the display. We describe how to detect and track a display in live video on a mobile device without sourcing out computation to a server. Furthermore, we describe how to use the TUIO protocol [7] to transfer the touch input from the mobile device to a distant display. The TUIO protocol is the de-facto standard for multi-touch applications and thus, our approach can be easily used to enable distant interaction with any multi-touch application.

RELATED WORK
Several interaction techniques have been proposed for using mobile devices to interact with distant displays. In the following, we introduce the approaches that are closely related to our work. The most prominent techniques are relative and indirect pointing, as well as augmented reality approaches. In [1], Ballagas et al describe how a camera-equipped mobile device can be turned into a mouse-like device when using relative and indirect pointing

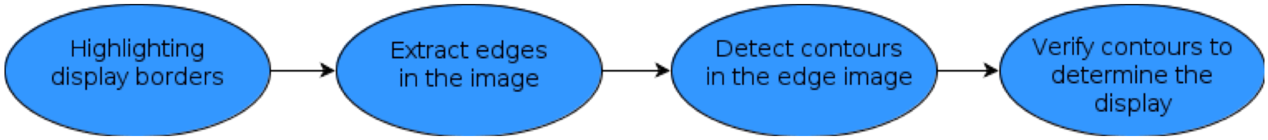

Figure 2: The four steps of the image detection algorithm.

techniques. One disadvantage of this approach is the required virtual pointer, which wastes screen real estate. Baur et al. explore the metaphor of optical projection to digital surfaces in an administrated multi-display environment [2]. They use image features to determine the position and orientation of the mobile device in relation to the display. In [10], Tani et al. describe how users could interact from a distance with real-world objects through live video. The aforementioned Touch Projector [4] revisits this metaphor and extends it for mobile use. Our approach for detecting and tracking a display is similar to the one that is used by Touch Projector. Touch Projector detects white rectangular frames, which are displayed around the potentially interactive content. With the relative position of the white rectangles to each other, it retrieves the ID of the current display from an environment manager to retrieve and access the actual content. This is the first of two general aspects, in which our approach is fundamentally different. While Touch Projector detects the display based on the relative positioning of its content (managed by an environment manager) and offers the possibility of directly interacting with the content and moving it from one display to another by drag and drop, our approach detects the display as a whole, based on its contours, and simply transfers plain touch input on the mobile device to the display. This can be compared to simply controlling a pointer from remote. The second fundamental difference is the computation of the data. While Touch Projector only grabs video frames and sends them to a server for computation, in our approach, all computations are handled on the mobile device itself, and the touch information is the only data that is constantly send. Hence, in contrast to Touch Projector, to work in real time, our system requires much less bandwidth.

DISPLAY TRACKING

The central part of our framework is the recognition and the tracking of the display, both achieved by analyzing the frames of the live video from the mobile device. To be independent of a server application, we implemented the tracking component of our framework completely on the mobile device. The tracking component detects the rectangular borders of a display that occurs in the live video.

The applied algorithm utilizes the Open Source Computer Vision Library OpenCV[1]. The process of detecting a public display in the live video on the mobile device consists of the following four steps (compare Figure 2):

1. The current frame of the live video is converted into a gray-scale image. This basic step is necessary to achieve optimal results. In order to enhance dark structures in the gray-scale image, we apply Erosion [9] as a morphological filter. This sharpens the edges of the display in the image.
2. We apply a Canny Edge Detector [6] to extract the edges of the filtered image. To improve the result and to close gaps between the detected edges, we apply Dilation [9] as an additional morphological filter. This expands bright structures in the filtered image.
3. In this step, we apply a contour detection on the filtered image. Since the dominant shape of public displays is rectangular, all detected contours are verified accordingly. Contours, which form a rectangle, are classified as a possible display.
4. To complete the display recognition, the rectangle is declared as a display, whose center is nearest to the camera picture center.

The algorithm is initially applied to every frame of the live video until the first display is successfully detected. At this point, the display detection is paused to save resources and to speed up the tracking. When moving the smartphone, the display tracking continues to update the position of the detected display in the live video. The motion of the smartphone is detected with the help of the accelerometer.

As mentioned before, we use the TUIO protocol for sending the touch information from the mobile device to

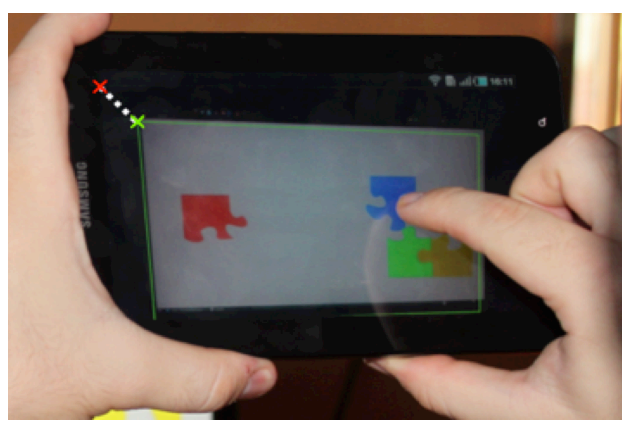

Figure 3: A detected display (green rectangle). The red x denotes the zero point of the touch capable area. The green x denotes the zero point of the coordinate system, which is used for mapping the touches.

[1] http://opencv.willowgarage.com/

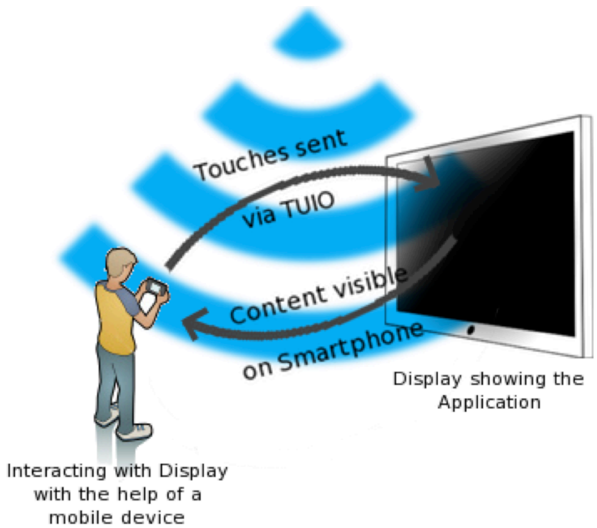

Figure 4: The interaction framework. The geometry of the distant display is detected and tracked in the live video on the mobile device. Touch coordinates on the mobile device are mapped onto the geometry of the distant display and the touch information is transferred utilizing the TUIO protocol.

the public display. As can be seen in Figure 3, touches are only perceived, if they are recognized within the borders of the detected display in the live video. Before transmitting the touch information to the display, the touch coordinates need to be mapped on the dimensions of the detected display. Hence, the upper left corner of the rectangle presents the zero point in a two dimensional coordinate system. As depicted in Figure 3, the mapping of the touch coordinates from the screen of the mobile device to coordinates on the display can simply be achieved by subtracting the difference between the zero point of the above explained coordinate system and the zero point of the complete touch area from the touch input coordinates.

When touching the display of the mobile device, the display tracking is paused if active. With this approach, a stable computation of the mapping can be achieved. In this case the last detected display is used to compute the coordinates mapping.

INTERACTION FRAMEWORK

We propose a simple, yet versatile framework to enable interaction with digital public displays, which are not capable of handling touch input or which are located out of the user's reach (see Figure 4). The described framework enables designers of digital signage and information displays to introduce novel means of interaction for this class of devices. This allows for example the design of interactive advertisements that enable users to interact with a digital product or simply purchase it.

The proposed approach makes use of the TUIO protocol, which is the de-facto standard for multi-touch applications. Hence, the framework allows simply porting of existing multi-touch applications to public displays and allowing applications such as Photo-Browsing (e.g. the "City Wall"[8]) or Games on distant digital displays.

Interactions Scheme

In the first step, the display needs to be detected and identified. The mobile device automatically detects the display in the live video, if both mobile device and display are within the same network. Then, simply scanning a QR-code, which is displayed on or near the public display, can pair the devices (see Figure 5).

Second, a user can interact with the content on the display by touch input on the occurrence of the display in the live video on the mobile device. As already mentioned, the touch input can be forwarded from the mobile device to the display by sending TUIO messages from the mobile device to the display. To forward the touch information, we use a special TUIO cursor message. Such a remote cursor message contains the following parameters:

- The mobile device ID to associate the remote cursor. The key part of the device's IP address is used as ID.
- The mapped touch coordinates and the velocity of the touches to realize accelerated interaction, like quick scrolling or sliding through content.
- A value for the ratio between the dimension of the public display and the recognized display in the live video. This value is used to compute the distance between the interacting user and the display.

With this approach, existing multi-touch applications can be easily ported to unreachable or touch-incapable displays.

Design Considerations

With this approach and distant touch enabled public displays, various design considerations can be deduced:

- Interaction zones can be defined in such a way, that the possible interactions and the displayed content depends on how far the user is away from

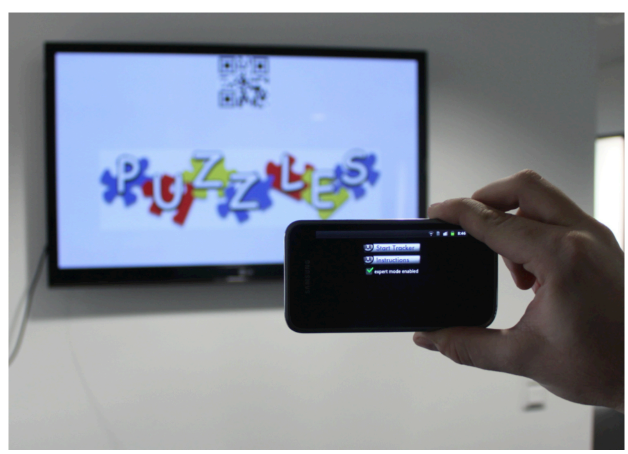

Figure 5: Device pairing. One possible way of pairing of the mobile device and the distant display by scanning a barcode on the display in order to send TUIO messages to the intended display.

the display. E.g., the closer a person is to the display, the more fine-grained interactions are possible. Hence, the further the user is from the display his interactions will be coarse.

- Proximity: The position of the user - relative to the display - can be determined when detecting a display. If a display appears in the live video of the mobile device, we can use the size of the visible part in combination with the properties of the camera to compute the users distance to the display. When detecting a display, we can determine the user's orientation to the display (the angle from which he is looking at the display) by the distortion of the detected rectangle. This information could be communicated back to the display, in order to enable services according to the user's relative position.
- Magic Lens-/AR-approach allows indirect interaction in a direct manipulation manner.

This framework can be seen as a simple input device and driver model (see Figure 3). We see this as an initial step towards universal mobile interaction with public displays. We like to use this approach to investigate and discuss the interaction with digital signage and public displays in general from the question how to design user interfaces for public displays, how to entice people to interact with public displays or how to collaborative interact with such displays in a multi-user scenario. As a first proof of concept in order to do initial testing we implemented a small puzzle app (see Figure 6). Additionally various TUIO-based multi-touch applications have been successfully operated via distant interaction on the mobile device.

Technical restrictions like accuracy, e.g. the big fat projected finger problem should be also investigated. The system might be extended in order perform in the wild studies in public display interaction.

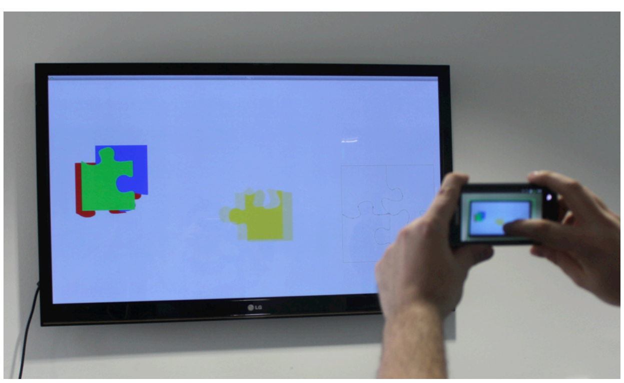

Figure 6: Example App. Puzzle Game. The user drags a puzzle tile by touching its representation in the live video on the mobile device.

CONCLUSION

In this paper, we described a framework that allows universal, direct interaction with distant displays. We described a vision-based approach to detect rectangular displays within the live camera view on a mobile device. The applications maps touch input from the display of the mobile device onto the geometry of the distant display. To transfer the touch information to the public display, we utilize the TUIO protocol. The described system allows supporting direct interaction with distant displays with a rectangular form factor, which are capable of handling TUIO information.

For future work, we plan to further refine the tracking of the display and a release of the application as a universal TUIO client via the TUIO community.

REFERENCES

1. Ballagas, R., Borchers, J., Rohs, M. and Sheridan, J.G. The smart phone: a ubiquitous input device. In *IEEE Pervasive Computing* 5, 1, 2006.
2. Baur, D., Boring, S. and Feiner, S. Virtual projection: ecploring optical projection as a metaphor for multi-device interaction. In *Proceedings of CHI'12*, ACM, 2012.
3. Bier, E.A., Stone, M.C., Pier, K., Buxton, W. and De-Rose, T.D. Toolglass and magic lenses: the seethrough interface. *Proc. SIGGRAPH'93* ACM, 1993.
4. Boring, S., Baur, D., Butz, A., Gustafson, S. and Baudisch, P. Touch projector: mobile interaction through video. In *Proceedings of CHI'10*, ACM, 2010.
5. Boring, S., Gehring, S., Wiethoff, A., Blöckner, M., Schöning, J. and Butz, A. Multi-user interaction om media façades through live video on mobile devices. In *Proceedings of CHI'11*, ACM, 2011.
6. Canny, J., A Computational Approach To Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, 1986.
7. Kaltenbrunner, M.; Bovermann, T.; Bencina, R. & Costanza, E. TUIO: A protocol for table-top tangible user interfaces. In *Proceedings of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation*', 2005.
8. Peltonen, P.; Kurvinen, E.; Salovaara, A.; Jacucci, G.; Ilmonen, T.; Evans, J.; Oulasvirta, A. & Saarikko, P. It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In *Proceedings of 'CHI '08*, ACM, 2008.
9. Serra, J. and Luc, V., An overview of morphological filtering Journal Circuits, Systems, and Signal Processing - Special issue: median and morphological filters Volume 11 Issue 1, 1992 .
10. Tani, M., Yamaashi, K., Tanikoshi, K., Futakawa, M. and Tanifuji, S. Object-oriented video: interaction with real-world objects through live video. In *Proceedings of CHI'92*, ACM, 1992.