Towards Ad Hoc Mobile Multi-display Environments on Commodity Mobile Devices

Jens Grubert*
Coburg University
Germany

Matthias Kranz†
University of Passau
Germany

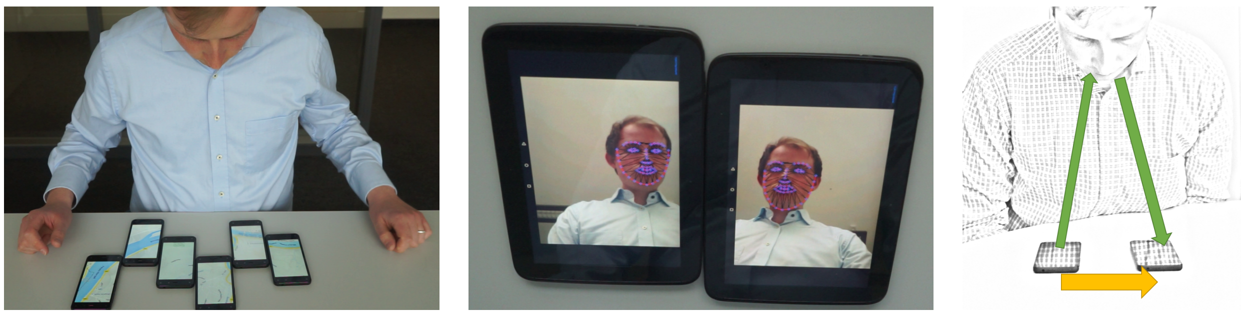

Figure 1: Left: HeadPhones allows dynamic registration between mobile devices using the human head as reference frame. Middle: Head pose estimation and concatenation is used to determine the relative positions of the devices to each other. Right: Transformation between two devices is computed by chaining the (inverse) transformations between each device and the user's head.

Abstract

We present a demonstration of HeadPhones (Headtracking + smartPhones), a novel approach for the spatial registration of multiple mobile devices into an ad hoc multi-display environment. We propose to employ the user's head as external reference frame for the registration of multiple mobile devices into a common coordinate system. Our approach allows for dynamic repositioning of devices during runtime without the need for external infrastructure such as separate cameras or fiducials. Specifically, our only requirements are local network connections and mobile devices with built-in front facing cameras. This way, HeadPhones enables spatially-aware multi-display applications in mobile contexts.

Index Terms: H.5.1 [Information Interfaces and Presentation (e.g. HCI)]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1 Introduction

Large displays are compelling for many work tasks, but are typically non-mobile. In contrast, many mobile users own at least one smartphone, along with a tablet and potentially a smartwatch. When people meet, the number of concurrently available mobile screens easily reaches a dozen or more. Collectively, these individual small displays could be used to create large-scale displays in an ad hoc manner.

Still, creating these ad hoc multi-display environments is typically cumbersome. Besides initially *binding* or associating devices into a group (i.e. each device is aware of the existence of nearby devices), devices generally lack the knowledge about their *spatial position* relative to the devices around them. While a number approaches for ad hoc device binding already allows for easy mobile use [1], determining the spatial position of devices in an ad hoc mobile multi-device environment remains challenging [3]. So far, approaches for the *spatial registration* of multiple mobile devices have either required external infrastructure such as cameras [12] or fiducials [8], required an instrumentation of the user [**?**], resulted in low spatial registration accuracy [5, 6] or allowed for only static device configurations (i.e. devices could not be freely moved after initial registration [4, 10]).

Our system, HeadPhones[1], instead, makes use of the user's head as external reference frame, allowing to overcome the limitations of previous approaches [2]. The system solely assumes that the devices have a front facing camera, which is used for head tracking. To the best of our knowledge, HeadPhones is the first system for the spatial registration of mobile devices in an ad hoc multi-display environment, which does not rely on external infrastructure such as cameras, allows for dynamic repositioning of devices after initial registration and achieves comparable accuracy of existing camera-based and gesture-based approaches.

2 Related Work

Binding and spatially registering multiple mobile devices into multi-display environments has been of interest to the research community for a while. Chong and Gellersen present a recent overview of binding or device association techniques [1].

HeadPhones primarily builts on previous work in the domain of spatial registration in mobile multi-display environments. A popular approach for spatial registration is to use finger or pen gestures across display boundaries [4]. However, those approaches typically do not allow for repositioning of individual devices after the initial registration step. IMU-based approaches are often employed for dynamic peephole navigation on mobile devices [11], but might be combined with gesture-based approaches to enable dynamic repositioning. Still, those approaches are prone to sensor drift.

There are various camera-based approaches, e.g., [12]. However, they require an external camera or depth-sensor to be available and, hence, might constrain the mobility of device registration.

*e-mail: jens.grubert@hs-coburg.de
†e-mail: matthias.kranz@uni-passau.de

[1] source code: https://gitlab.com/mixedrealitylab/HeadPhones

Figure 2: Map applications multiple map layers can be browsed by changing the height of individual devices.

The closest approach to our work was presented by Li and Kobbelt [8]. They propose to use a fiducial marker, which is tracked using the front camera of mobile phones. However, the need for a printed fiducial marker limits the mobility of their approach. In contrast, we propose to use the user's head to enable spatial registration of multiple mobile devices.

3 Display Registration Through Head Tracking

The core idea of HeadPhones is to enable the spatial registration of multiple mobile devices with front cameras through head tracking (HeadPhones: Headtracking + smartPhones). After an initial display binding step using existing approaches such as entering a server IP or photographing a QR code [1], each device estimates the pose of the user's head relative to its own coordinate origin. To determine the pose of two devices relative to each other the following transformation takes place in a right-handed coordinate system with the device coordinate system in the screen center (in portrait mode: x-axis pointing right, y-axis pointing to the top of the phone, z-axis pointing out of the phone), see also Figure **??**, left.

For our demo, we utilize Amazon Fire Phones, which have access to a robust head tracking API used in commercial products and resulting in satisfying user experiences [9]. To work on common mobile devices with a single front camera, we combine a 2D deformable face tracker [13] with a solver for the perspective-n-point problem [7]. In a first step, 2D image points of facial landmarks are estimated using deformable model fitting. For the second step, we use a rigid 3D model which is mapped to selected image points of the 2D model (eyes, nostrils, temples).

We demonstrate the applicability of our approach with a layered map application. Browsing maps on mobile devices typically requires frequent zooming between levels of detail. Using multiple mobile screens, the extend of a current map level is extended mitigating the need for zooming (to a certain level). Furthermore, the mobile devices can be dynamically repositioned to explore even further map areas, see Figure 1, left. Additionally, the user can pick up individual devices and switch between rendering modes of maps, such as standard view, satellite or traffic views (see Figure 2). This potentially allows viewing different rendering modes of the same map area simultaneously (e.g., if the user looks from the side).

We implemented HeadPhones as client/server application. The clients can join a multi-display group by entering an IP. The server can run on any of the mobile devices resulting in a infrastructure-free, completely mobile solution. One of the devices determines the coordinate origin of the virtual display (typically the device to join first, changeable at any time later).

For rendering we employ HTML5 + JavaScript with 3D rendering done in WebGL and three.js. The head tracking data is injected into JavaScript. The individual smartphones communicate via WebSocket.

On an Amazon Fire Phone, the demo, including head tracking, run at 30 frames per second (fps) using the Amazon API and at 20 fps using our monoscopic face tracker.

4 Demo Experience

We will demonstrate HeadPhones using a set of Amazon Fire Phones that can be dynamically reaaranged on a table surface, picked up for browsing of map layers and touched for map panning. A video of the system can be found here: https://youtu.be/X-jciek-Sg8

5 Conclusion

We present a demonstration of HeadPhones (Headtracking + smartPhones), a novel approach for the spatial registration of multiple mobile devices into an ad hoc multi-display environment.

References

[1] M. K. Chong and H. Gellersen. Usability classification for spontaneous device association. *Personal Ubiquitous Comput.*, 16(1):77–89, Jan. 2012.

[2] J. Grubert and M. Kranz. Headphones: Ad hoc mobile multi-display environments through head tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '17, New York, NY, USA, 2017. ACM.

[3] J. Grubert, M. Kranz, and A. Quigley. Challenges in mobile multi-device ecosystems. *mUX: The Journal of Mobile User Experience*, 5(1):5, 2016.

[4] K. Hinckley, G. Ramos, F. Guimbretiere, P. Baudisch, and M. Smith. Stitching: Pen gestures that span multiple displays. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 23–31, New York, NY, USA, 2004. ACM.

[5] H. Jin, C. Holz, and K. Hornbæk. Tracko: Ad-hoc mobile 3d tracking using bluetooth low energy and inaudible signals for cross-device interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 147–156, New York, NY, USA, 2015. ACM.

[6] H. Jin, C. Xu, and K. Lyons. Corona: Positioning adjacent device with asymmetric bluetooth low energy rssi distributions. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 175–179, New York, NY, USA, 2015. ACM.

[7] V. Lepetit, F. Moreno-Noguer, and P. Fua. Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2):155–166, 2009.

[8] M. Li and L. Kobbelt. Dynamic tiling display: Building an interactive display surface using multiple mobile devices. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, pages 24:1–24:4, New York, NY, USA, 2012. ACM.

[9] A. Mulloni, J. Grubert, H. Seichter, T. Langlotz, R. Grasset, G. Reitmayr, and D. Schmalstieg. Experiences with the impact of tracking technology in mobile augmented reality evaluations. In *MobileHCI 2012 Workshop MobiVis*, volume 2, 2012.

[10] H. S. Nielsen, M. P. Olsen, M. B. Skov, and J. Kjeldskov. Juxtapinch: An application for collocated multi-device photo sharing. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, pages 417–420, New York, NY, USA, 2014. ACM.

[11] M. Pahud, K. Hinckley, S. Iqbal, A. Sellen, and B. Buxton. Toward compound navigation tasks on mobiles via spatial manipulation. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 113–122, New York, NY, USA, 2013. ACM.

[12] R. Rädle, H.-C. Jetter, N. Marquardt, H. Reiterer, and Y. Rogers. Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, pages 45–54, New York, NY, USA, 2014. ACM.

[13] J. M. Saragih, S. Lucey, and J. F. Cohn. Face alignment through subspace constrained mean-shifts. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1034–1041, Sept 2009.