Advanced Interaction Techniques for Handheld Mobile Augmented Reality

Matthias Baldauf
FTW - Telecommunications
Research Center Vienna
Donau-City-Strasse 1
Vienna, 1220, Austria
baldauf@ftw.at

Peter Fröhlich
FTW - Telecommunications
Research Center Vienna
Donau-City-Strasse 1
Vienna, 1220, Austria
froehlich@ftw.at

Peter Reichl
Université Européenne de
Bretagne
5 Boulevard Laënnec
Rennes, 35000, France
peter.reichl@irisa.fr

MobileHCI 2011, Aug 30–Sept 2, 2011, Stockholm, Sweden.
ACM 978-1-4503-0541-9/11/08-09.

Abstract
Research in the field of handheld augmented reality has so far often focused on technical tracking and visualization challenges, while the investigation of interaction possibilities beyond the classical 'magic lens' paradigm has been neglected. We argue for a closer investigation of advanced interaction techniques that could be used in combination with handheld augmented reality visualizations. We introduce several such promising 'interactive magic lens' techniques enabled on handhelds and finally briefly describe our current efforts towards the recognition and usage of finger gestures within interactive magic lenses.

Keywords
Mobile interaction, mobile augmented reality, finger detection, handheld computer vision

ACM Classification Keywords
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - *Input devices and strategies, Interaction styles*

General Terms
Human Factors, Design

Figure 1. Icon overlay in the 'real-world browser' *Wikitude* [9]

Figure 2. Targeting with a magic lens in an AR shooter game [3]

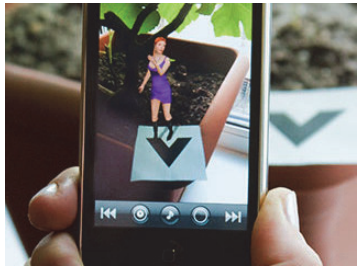

Figure 3. Animated virtual character in a mobile AR app [2]

Introduction

Due to steady advances in mobile computing technology, handheld devices have become a promising lightweight platform for mass market augmented reality (AR) applications. While most related research activities have strongly focused on technical challenges, such as improved camera pose estimation and more robust tracking methods (e.g. [11]), research towards complementing augmented reality visualizations by advanced interaction techniques is underrepresented. Existing applications such as the GPS-based 'real-world browser' *Wikitude* [9] often are restricted to touch-sensitive overlay icons (Figure 1) or use the camera viewfinder for targeting purposes only, such as the AR game *ARhrrrr* [3] (Figure 2).

In this paper, we introduce a number of promising novel interaction techniques that could be combined with classic handheld augmented reality. To clearly separate our efforts from portable AR systems (often with special glasses and a computer in a backpack) and mobile approaches utilizing projectors for 'augmenting' physical objects, our notion of 'mobile augmented reality' refers to the traditional *magic lens* metaphor [5] realized on smartphones, that is, the live camera view of the device is overlaid with precisely referenced virtual objects, usually by means of computer vision. Additionally, we give a brief glimpse on our current practical research work on such 'interactive magic lens' techniques and we introduce our recently developed prototype for the visual detection of a user's fingertips in a mobile AR scenario.

Novel Interactive Magic Lens Techniques

This section introduces several input possibilities trying to explore the full interaction potential in mobile AR. In contrast to former work (such as [7]) considering the 'freezing' of AR views for interaction we focus on real-time interaction with virtual objects.

Touch-based live-video interaction

Researchers have recently started to explore the interaction with mobile live-video. Boring et al. [6] presented a system for moving photos on a remote screen via drag'n'drop operations performed on the camera viewfinder of a smartphone in real-time. Baldauf et al. [4] introduced a related prototype for controlling a remote screen with arbitrary content by touching the screen content's control elements on the mobile video.

In a similar vein, one could envisage touch-based live-video interaction beyond selecting overlay icons that enable the manipulation of virtual objects. Examples include drag'n'drop actions on virtual objects for repositioning them in the real world and transferring virtual objects from the mobile device to the real-world scene to be integrated there, or vice versa. In addition, aforementioned screen interaction approaches could be combined with displayed virtual objects, e.g. touching projected content through a mobile device's live-video might both affect the projected information and overlaid virtual objects.

Gestural Free-Hand Interaction

The vision-based detection of hand and finger gestures is a well-investigated task in the field of machine vision.

Figure 4. Hand analysis on the threshold image

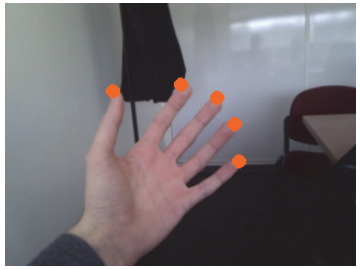

Figure 5. Live-video overlay with the calculated fingertip positions

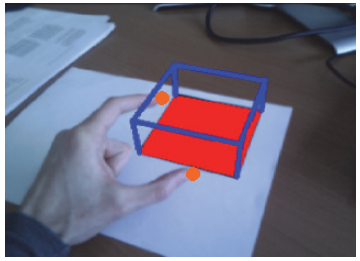

Figure 6. Combining fingertip detection with a virtual object

However, respective research work on implementation and design issues for mobile magic lens scenarios is scarce, despite its manifold use cases.

For example, the recognition of a user's finger gestures within the range of a magic lens could enable pointing at virtual objects or indicating a target position for a moving virtual object. Other exemplary use cases are drag'n'drop operations for items through pinch gestures performed with thumb and index finger. For example, animated virtual characters might react on a variety of different gestures. Thus, games might benefit from this novel interaction technique in particular. In addition, hand gestures might not only be performed 'inside' the magic lens but also be captured by a front camera facing the user. Such gestures do not allow for a 'direct' interaction with virtual objects but may rather act as commands, e.g. indicating a direction by a pointing gesture. Beyond finger gestures, the recognition of faces or even of facial expressions may influence the reaction of a virtual object.

Acceleration-based gestural interaction
Gesture recognition based on accelerometers is an active research field (compare [8]). Respective solutions have already found their way onto mass-market mobiles such as gesture-sensitive music applications with random play triggered by shaking. Such acceleration-based gestures may also affect virtual objects in mobile augmented reality applications. A virtual item, once captured through the magic lens, might react on shaking by changing its appearance. Or a jerky movement with the device might indicate a direction for a moving virtual object.

Auditory Interaction
Another increasingly relevant interaction technique for mobile devices is auditory input. Applications for the recognition of human speech are available on several mobile platforms while some even support the integration into third-party applications (cf. [1]). Thus, speech is another obvious possibility to interact with virtual objects or characters. Predefined commands might control virtual objects, speaking out names might select respective objects. In addition, also non-speech auditory input seems worth a closer investigation. E.g. virtual objects could be controlled through whistling or flicking or they might react on music sensed via the built-in microphone or other background noises. A related existing fun application is *ARGirl* displaying a marker-based animated dancer moving to the rhythm of a song played by the mobile.

Current Research Focus

We are currently building interaction prototypes to investigate the aforementioned gestural free-hand interaction with virtual objects within a mobile magic lens. To enable in-depth experiences with this interaction technique, we recently developed a real-time fingertip detection engine especially targeted at mobile devices. Our lightweight implementation approach applies a pixel-based skin detection based on RGB thresholds according to [10]. We then search for contours in the resulting binary and assume the largest contour to be the hand. By calculating the largest enclosing circle and the largest incircle for the contour, we infer information about the hand's size and orientation. The positions of the actual fingertips are then calculated as local distance maxima from the

largest incircle supposed to be the palm, as shown in Figure 4. Figure 5 shows the calculated fingertip positions on top of the live-video. As an exemplary use case, Figure 6 illustrates the usage of a pinch gesture to select or grab a virtual object, in this case a simple wireframe model placed on a fiducial marker for demonstration purposes.

Conclusions and Outlook

The interactive magic lens approaches presented in this paper widen the design space for mobile augmented reality research. In upcoming research, we will investigate this design space, by high-fidelity prototyping and comparative user testing. An interesting aspect will be to support the user's co-ordination between gestural free-hand interaction and the related presentation on the magic lens display.

Acknowledgements

This work has been carried out within the project PRIAMUS financed by FFG and A1. FTW is supported by the Austrian Government and by the City of Vienna within the competence center program *COMET*. Additional funding from the SISCom International Research Chair *Future Telecommunication Ecosystems* at Université Européenne de Bretagne Rennes and from INRIA Rennes Bretagne-Atlantique is gratefully acknowledged.

References

[1] Android Developers. Speech Input. http://developer.android.com/resources/articles/speech-input.html

[2] APetrus. ARGirl 2.0. http://apetrus.com/en/argirl/

[3] Augmented Environment Labs. ARhrrrr. http://www.augmentedenvironments.org/lab/research/handheld-ar/arhrrrr/

[4] Baldauf, M., Fröhlich, P., Reichl, P. Touching the untouchables: vision-based real-time interaction with public displays through mobile touchscreen devices. In Adj. Proc. of Pervasive 2010.

[5] Bier, E.A., Stone, M.C., Pier, K., Buxton, W., and DeRose, T.D. 1993. Toolglass and magic lenses: the see-through interface. In Proc. of SIGGRAPH '93, 73-80

[6] Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch projector: mobile interaction through video. In Proc. of CHI '10, 2287-2296.

[7] Güven, S., Feiner, S., Oda, O. Mobile augmented reality interaction techniques for authoring situated media on-site. In Proc. of ISMAR 2006, 235-236.

[8] Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. 2009. uWave: Accelerometer-based personalized gesture recognition and its applications. Pervasive Mob. Comput. 5, 6, 657-675.

[9] Mobilizy. Wikitude World Browser. http://www.wikitude.org

[10] Peer, P., Kovac, J., and Solina, F. Human skin colour clustering for face detection. In Proc. of EUROCON 2003, Int. Conf. on Computer as a Tool, 2003.

[11] Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D. Real-time detection and tracking for augmented reality on Mobile Phones. IEEE Transactions on Visualization and Computer Graphics, vol.16, no.3, 355-368.