

Controlling and Filtering Information Density with Spatial Interaction Techniques via Handheld Augmented Reality

Jens Keil¹, Michael Zoellner², Timo Engelke¹,
Folker Wientapper¹, and Michael Schmitt¹

¹ Fraunhofer IGD, Darmstadt, Germany

² Hof University, Germany

Abstract. In our paper we are proposing a method for contextual information filtering based on the user's movement and location in order to enable the intuitive usage of an "internet of things" via augmented reality (AR) without information overload. Similar to Ray & Charles Eames' "Power of Ten" and Jef Raskin's "Zooming Interface" we are displaying seamless information layers by simply moving around a Greek statue or a miniature model of an Ariane-5 space rocket. Therefore we are employing concepts of camera- and motion-based interaction techniques and use the metaphors of "investigation" and "exploration" to control the way augmented and visually superimposed elements are presented in order to mediate information in an enhanced and engaging manner with aspects of digital storytelling techniques.

Keywords: Adaptive and personalized interfaces, Human Centered Design, Information visualization, Interaction design, New Technology and its Usefulness.

1 Introduction

The idea of retrieving and viewing geo-located information with augmented reality (AR) additionally or compared to traditional maps has been eased with the advent of handheld augmented reality on recent smartphones. AR browsers like Junaio, Layar, Nokia World Lense and other apps superimpose geo-referenced information on top of the camera's video stream like annotations, which relate and stick to a point/target-of-interest in the environment. Usually, these annotations are grouped in special "channels" and are visually depicted by icons or textual labels and are in reference to a specific place or position. Although oftenly used, they aren't of course technically limited to geo-referenced information.

With this video-see-through effect on smartphones, it has more generally become a commodity to superimpose annotations in AR views: technically speaking, in contrast to creating rich 3D models and assets, annotations are quite informative, they do not need intense processing power of GPUs, and their creation is relatively simple and of low costs. However, adding such descriptive information obviously come with some visual drawbacks: those annotations and labels

may occlude real object or more seriously a point-of-interest in the video or, if not aligned accordingly, they may tend to create ambiguity, since it might not be clear, to which target the information belongs. Moreover, if not filtered, the amount of too many augmented information turns the idea of data/information exposition and clarification into a visual overload of the scene.

Inspired by these issues and with the question of how to present and ease data access in handheld augmented reality, the methods we are going to present in this paper take on the concepts of motion controlled interaction in AR, and of zooming interfaces, as e.g. proposed by Raskin [6], where users control the density of information presented by scaling the viewed area in order to see more or less details while browsing through data. With our approximation technique we transfer this concept from originally virtual 2D spaces into application in real 3D environments, where the density and shape of augmented and overlaid information changes according to the distance from device to tracking target. We believe that this approach will not only make the experience of contextual information more engaging and intuitive but also more lasting, because users may focus on perceiving relevant information instead of coping with the user interface (UI) structure of the application that presents it.

2 Related Work

Almost all handheld AR apps nowadays work like magic lenses. The term was introduced at Siggraph 1993 by [2]. They proposed a new see-through interface to "reveal hidden information, to enhance data of interest, or to suppress distracting information". Originally designed for display interfaces this concept is to be found in most augmented reality applications today. Up until now, there exist several methods to dynamically adapt the content richness of the augmented data, especially, when dealing with labels and annotations. In handheld AR apps representations appear usually rather static: targets such as magazine pictures or posters are extended with text, videos or simple 3D assets (cf. IKEAs catalogue app, [10]). Interaction, if any, are click targets that link to external content. In case of AR-browsers, which are mainly used while being on the walk, only the user's movement, i.e. position, device heading, and orientation influences changes of the visual representation of data.

The work of [8] presents concepts for automatised positioning of superimposed annotations to avoid cluttering and occlusion in an adaptive manner. In terms of information flow, however, the presented methods are quite global and aren't controlled by the user itself, though. Although they tend to structure augmented elements, they do not introduce an informed order or sequence of data elements.

With adaptive visual aids [9] presents a technique, where the guidance level of superimposed annotations, i.e. the strength of information presented, changes dynamically. This mainly results in different visualisations, starting from unobtrusive visual indicators up to complex animations. The presented methods are explicitly tailored to support maintenance tasks. Users may not influence directly the depth-level, or eventually need to click through a menu beforehand,

which would interrupt and interfere with the video-see-through experience. The technique also doesn't account the absolute or relative position from user and device to the target-of-interest.

The idea of mapping gestures and well-known interaction tasks (such as selection, scrolling, navigation and object manipulation) to motion has been closely discussed in literature ([12] - [15]). The work presented in [5] and [4] uses camera and motion-based interaction in a way to control application and content paradigms that account camera-acquired targets in reality and device motion to turn the handheld device into a versatile interface. However, the view of interaction for both is technology-driven: [5] illustrates scenarios where users may "select" functions depending on their position to the tracking target by moving the device from far-to-near (fly-through) in order to e.g. de-/re-activate clipping masks that let users look inside an superimposed 3D model. Although their concept works with the camera/device position, the concept mainly maps rather arbitrary functions to movements in order to trigger application-modes, which might possibly work better with traditional menus. [4] proposed a conceptual framework of movement patterns, which are connected to a printed marker-map that allows to retrieve rotation or distance from device to target with computer vision. In order to use them accordingly, users need training in order to learn the basic gesture/movement-set and moreover the more complex ones.

3 User-Centric Spatial and Motion Interactions in AR

With computer vision based 2D and 3D target-tracking as a core enabling technology, the concepts that are going to be presented in this paper underlie paradigms and interaction-methods that are user-centric, and thereby in contrast to the typical understanding of gestures and patterns in computer science.

Instead of having pattern-sets, our idea is to link interaction to the superimposed visuals in a way that feels natural. We are using movements and mental models that are familiar to the user and thus almost instinctively applied. By doing so, we connect superimposed annotations and other content to motion based interaction and to camera-acquired targets, where the paradigms do not necessarily need to be trained beforehand. This connection should be perceived as joyful, easy-to-understand and passively, since it influences, filters and re-arranges superimposed material, but doesn't act as a change between system modes.

Being also based on the core paradigm of handheld augmented reality, where devices behave like a magic lens [2] (deployed through the video-see-through effect) our "gestures" make use of the natural movements that users already do, when working with handled augmented reality. Thereby the gestures we present work like AR-gazing, which is pointing extendedly at the same area, as well as observing, which correlates with walking around a target-of-interest, and inspecting and magnifying, which uses the distance/proximity to a target to influence the level-of-depth of information presented.

3.1 Conceptual Outline of Interaction Principles

While working with AR on mobiles, we have explored that users intuitively tend to align a focus or point of interest in the middle/center of the screen. Hence, as a core metaphor, we employ this as a gazing technique for usual selection tasks, where an extended pointing over time acts as a trigger in order to select virtual superimposed elements or ones from reality. We can think of two ways to implement this: while in the first one the trigger is represented visually by a small crosshair and thereby close to the representation of a mouse pointer (cf. fig.1), the second one lacks such an obvious representation and is displayed much more passively. Nevertheless, both are solely controlled by device movement.

In an extended and more elaborated way, the system intentionally adapts content information density (cf. [1]) of visual presentations to the user's location and distance to the object of interest: the greater the distance to the object the lesser the information to be displayed, based on a priority sequence. When moving closer to the object new information entities are appearing while other eventually irrelevant ones will disappear. This concept works similar to Eames' 1968 documentary short film "Powers of Ten" [3] where magnitudes are illustrated by a flight from a picnic on earth out to the outer edges of the universe and back into a proton of a carbon atom in a blood cell with the speed of ten times magnification every ten seconds. It also continues Jef Raskin's idea of zooming interfaces [6] by replacing the mouse with a smartphone held by a moving user and adding the dimension of viewing angles.



Fig. 1. AR-Gazing: an embodied pointer acts as a visual proxy, where in reality selected components are virtually highlighted



Fig. 2. Image sequence showing enhanced interaction coupled to approximation, where information density increases or decreases accordingly

Additionally, a second parameter for locative information adaption is the user's angle to the object: certain digital information is connected to a point on or surround the three dimensional object, and only visible from the right angle; one, where it makes sense to these information chunks to be consumable, while all other information is faded out.

3.2 Adaptivity

Information presentation and filtering is not considered as a binary process. A parametric model of thresholds allows to define hotspot-zones for each superimposed item. It reacts to the user's and the device's angle and distance alike and controls, whether the items appear or vanish from the viewpoint.

The result is a highly dynamic information space that encourages the user to explore it. Especially in museums and education, exploration is an important didactic principle, where mobile devices and AR can have strong impact on museum visitors: [16] could show that AR interfaces were perceived being intuitive with strong impact on user's memory and learning curve. Via AR, participants recalled information afterwards more reliable and seamlessly shifted their focus from digital to physical information and vice versa without any major problems.

4 Paradigms in Practice

We tested and implemented the paradigms in context of several prototype AR apps on iPads. We used our own instantAR framework [11], which runs on an iPad 4. The system tracks 3-dimensional tracking targets at roundabout 20 to 30 Hz, which makes the experience fast and stable enough in order to work with intensive device and user movement. Tracking is done with computer vision using an advanced version of [17] which is based on KLT and SLAM.

We applied our methods in several scenarios. At first, we experimented with pointing on a pump machine. Here the user could point and focus on a mechanical component, that was then visually highlighted (cf. figure 1). Once selected, additional information popped up on the screen. We implemented this scenario with and without a visual proxy.

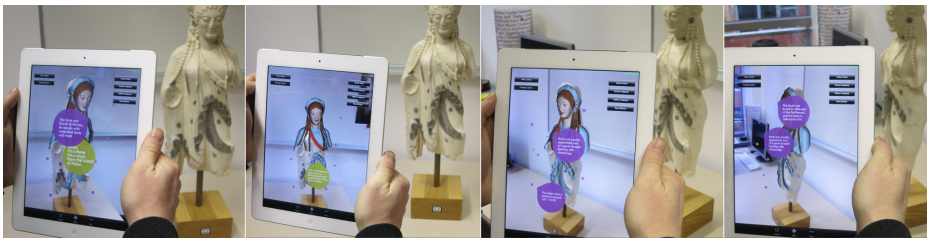


Fig. 3. Image sequence showing locative information adaption depending to the user's angle to the tracking target

Within the second case we used the approximation principle in order to observe a physical miniature model of an Ariane-5 space rocket in an exhibition at Cit de l'Espace in Toulouse. While the user holds an iPad at a fair distance, the first level of superimposed overlays presents a lineup of different types of former or foreign rockets in order compare them in size and shape. The closer the user gets, the more detailed the augmentation becomes, where on a second level labels appear and inform about the structure, and enable to even look virtually into the interior on a third level; literally turning the mobile into an interactive x-ray device.

In a third case, we use device orientation and motion to control the amount of annotated information superimposed on exhibits at a time: digital audio and visual elements, which are spread around a real physical object, are shown or hidden depending on the angle of user/device to the real object in focus (cf. figure 3). The technique turns AR's "what's around me" into "whats in front of me", where the device is not only a virtual lens but also a pointer: just like gazing, a user may center or focus in order to select digital content on through devices camera. Therefore we used statues from the archaic gallery of the Acropolis Museum. Besides influencing the displayed elements, the fashion of soundness and speech changes accordingly: while the user is literally face-to-face to the statue, information is told in first-person, whereas being aside shifts narration to third-person.

Technically, each tracking target has a predefined front or ground axis. The distance of each superimposed item is calculated relatively to the core target. In relation to it, we calculate an entrance region for all items, which controls their disappearance and reappearance once the user approximates. By using these transitional zones, we are able to fade in and out each label depending on how close it is to the user's current view. It is important to mention, that each virtual item is view aligned and always fronting towards the user's view, which is done for readability reasons. With these zones we are not only able to control the general fading while the user walks around the target, we also may observe and control the relative distance from user to virtual items alike.

In order to be visually convincing without cluttering the scene we took care that the render process is able to deal with occlusion (meaning that the real physical object masks superimposed items) and that it handles back-face culling, even when an object is rendered transparently. By doing so, superimposed objects seamlessly fit into the video image and appear in the right depth order (and are this way perceived as being in front or behind the physical object). We thereby may also highlight regions of the real three-dimensional objects.

5 Findings and Informal Evaluation

In order to get an initial user feedback and to test the proclaimed paradigms, we conducted an informal user testing. A group of 15 people, all familiar with handheld augmented reality applications and computer vision tracking, have been asked to use the system with no more information than the introduction of

each tracking target. It was of interest to us to see, how each participant would start to use the system, how they would work with the paradigms, and if they would start understanding the principles just by using them.

To start, the biggest finding was, that all participants in general liked the idea of working with motion controlled interaction and that they understood the principles without major issues. In fact, the majority almost expected a movement-coupled interaction, which seems to be a key appeal of augmented reality, since it is the motion that reveals that virtual annotations stick to reality.

The gazing as well as the approximation principle were generally inconspicuous. Although we wanted to work without the visual pointer at first, it turned out that people found it rather confusing without that kind of feedback, which frankly worked as an indicator. Without it, people didn't understand exactly, why or how the system reacted and why it started to highlight mechanical components and giving extra annotations. However, with the proxy, some people were then trying to touch and move it with a finger, too.

The angle-dependent technique turned out not being as "intuitive" as expected: almost all participants tried to tab/click presented elements in order to "activate" content. Even more, people seemed inclined to be lazy in movement: 10 of the 15 were moving the device but did neither reposition themselves much nor really walk around the statue. Instead, at first they pointed (gazed) the iPad towards the augmented annotations, and eventually tried to tab them. Only, when explained, they started to move around and experiment with the technique. This seems, as if people are eventually not familiar enough or simply not used to the technique.

Finally, the approximation controlled depth-level filtering wasn't as "unexpected": people more or less instantly started to come closer or go away, but again weren't much moving around sideways. This might be coupled also to the overall situation and the way objects are exposed: the space should invite people and encourage to move around an object, too. This could also imply further design considerations, e.g. to make this interaction more recognisable through the UI in general. We also tend to recommend to design such interactions redundantly and combine gazing and tabbing additionally to angle-dependant control or use visual hints and animations that tease this kind of principles a bit.

6 Conclusion and Future Work

In this paper we presented a method for contextual information filtering based on the user's movement and location for augmented reality on mobile devices. Based on the magic lens paradigm of handheld AR, we introduced gestures that employ users natural movements, such as gazing (pointing extendedly at the same area), observing (walking around a target-of-interest) and inspecting and magnifying (distance/proximity to a target) to influence the level-of-depth and overall appearance of contextual and spatial presented information. The results of the informal evaluation were surprising and promising. We think there is still potential for a deeper quantitative and qualitative evaluation combined with

usability testings. Especially when including users without AR knowledge which takes large parts of museum's target audience and mass users alike. These tests should be carried out together with museum partners and their expertise.

Acknowledgements. This work is part of the EU-funded project CHESSE. The project aims to investigate personalised digital storytelling aspects in heritage applications that are tailored to interests of visitors by a user-centered and personalised design in order to enrich a visit of museums or science centers. Being close to rich narrations like presented in [7], the project explores not only location- and object-centered story telling techniques but uses also mixed and augmented reality.

References

1. Tufte, E.: Beautiful Evidence. Graphics Press, Cheshire (2006)
2. Bier, E., Stone, M., Pier, K., Buxton, W., Derose, T.: Toolglass and magiclenses: The see-through interface. In: Proc. Siggraph 1993, Computer Graphics Annual Conference Series 1993, pp. 73–80 (1993)
3. Eames, C., Eames, R.: Powers of Ten (1968), <http://www.powersof10.com/film>
4. Rohs, M., Zweifel, P.: A conceptual framework for camera phone-based interaction techniques. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 171–189. Springer, Heidelberg (2005)
5. Harviainen, T., Korkalo, O., Woodward, C.: Camera-based interactions for augmented reality. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology 2009 (ACE 2009) (2009)
6. Raskin, J.: The humane interface: new directions for designing interactive systems. ACM Press/ Addison-Wesley Publishing Co., New York (2000)
7. Wither, J., Allen, R., Samanta, V., Hemanus, J., Tsai, Y., Azuma, R.: The Westwood Experience: Connecting Story to Locations via Mixed Reality. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality (2010)
8. Grasset, R., Langlotz, T., Kalkofen, D., Tatzgern, M., Schmalstieg, D.: Image-driven View Management for Augmented Reality Browsers. In: Proceedings of 11th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2012 (2012)
9. Webel, S.: Multimodal Training of Maintenance and Assembly Skills Based on Augmented Reality
10. Inter IKEA Systems B.B.: IKEA Katalog App (2012 /2013), <https://itunes.apple.com/de/app/ikea-katalog/id386592716>
11. Engelke, T., Becker, M., Wuest, H., Keil, J., Kuijper, A.: MobileAR Browser A generic architecture for rapid AR-multi-level development. Journal Expert Systems with Applications 40 (2013)
12. Rekimoto, J.: Tilting operations for smal screen interfaces. In: Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (1994)
13. Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C., Want, R.: Squeeze me, hold me, tilt me! An Exploration of Manipulative User Interfaces In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (1998)

14. Hinckley, K., Pierce, J., Sinclair, M., Horwitz, E.: Sensing Techniques for Mobile Interaction. In: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (2000)
15. Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., Conway, M.: The Video Mouse: A camera-based multi-degree-of-freedom Input Device. In: Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (1999)
16. Damala, A., Cuband, P., Bationo, A., Houlier, P., Marchal, I.: Bridging the Gap between the Digital and the Physical: Design and Evaluation of a Mobile Augmented Reality Guide for the Museum Visit. In: Proceedings of DIMEA The International Conference on Digital Media in Entertainment and Arts (2008)
17. Wientapper, F., Wuest, H., Kuijper, A.: Reconstruction and Accurate Alignment of Feature Maps for Augmented Reality. In: IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (2011)