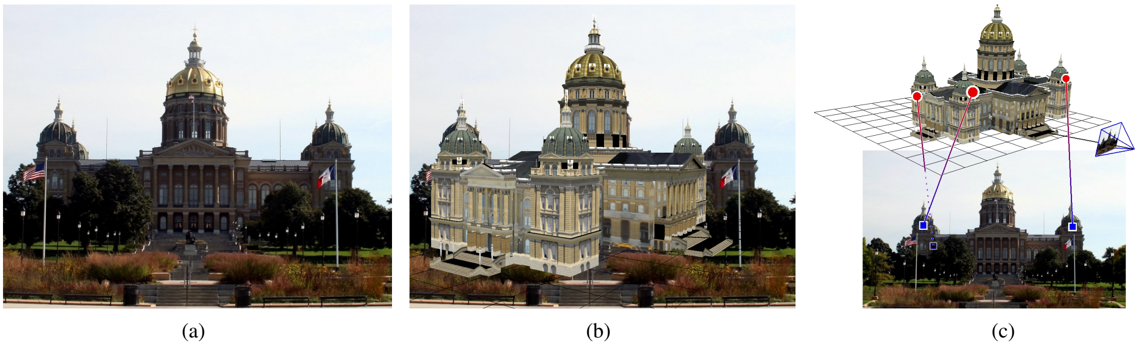
Joint Virtual Reality Conference of EGVE - EuroVR (2013)
B. Mohler, B. Raffin, H. Saito, and O. Staadt (Editors)

Exploring Distant Objects with Augmented Reality

Markus Tatzgern[1], Raphael Grasset[1], Eduardo Veas[1], Denis Kalkofen[1], Harmut Seichter[1], Dieter Schmalstieg[1]

[1]Graz University of Technology, Austria



(a) (b) (c)

Figure 1: *(a) How can I explore the Iowa State Capitol without physically moving? We present Augmented Reality interfaces using a virtual copy metaphor to access additional views. (b) A 3D copy is directly overlaid with its real world counterpart to access different views. (c) A variation of the interface spatially separates the 3D copy from the real object. To facilitate bridging the spatial offset between copy and real object, users can optionally create visual links to connect corresponding locations.*

Abstract

Augmented reality (AR) enables users to retrieve additional information about the real world objects and locations. Exploring such location-based information in AR requires physical movement to different viewpoints, which may be tiring and even infeasible when viewpoints are out of reach. In this paper, we present object-centric exploration techniques for handheld AR that allow users to access information freely using a virtual copy metaphor to explore large real world objects. We evaluated our interfaces in controlled conditions and collected first experiences in a real world pilot study. Based on our findings, we put forward design recommendations that should be considered by future generations of location-based AR browsers, 3D tourist guides, or in situated urban planning.

Categories and Subject Descriptors (according to ACM CCS):
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology

1. Introduction

Augmented Reality (AR) is a natural choice for exploring location-based information of real world objects, because AR overlays information directly into the user's surroundings. For instance, a user can easily access additional information about a building in an urban environment by pointing an AR-enabled mobile phone into its direction. However, the inherent egocentric reference frame of an AR interface becomes an obstacle once the user wants to explore objects that are out of reach. The user would need to physically move to a new position, which might be too cumbersome or even impossible.

3D maps allow exploring real world objects freely since they are not bound to the egocentric viewpoint of the user.

DOI: 10.2312/EGVE.JVRC13.049-056

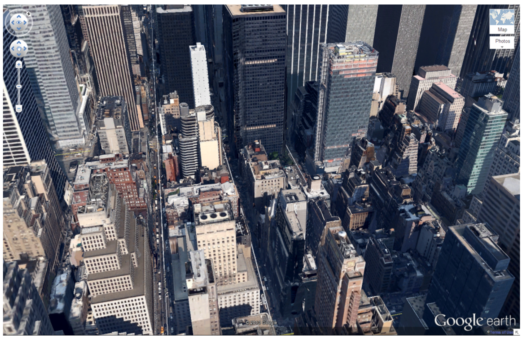


Figure 2: *3D map. A 3D map (here: Google Earth) allows users to explore surrounding real world objects. However, the user first has to identify the corresponding virtual object in the map and then relate it to his current position. Furthermore, in densely built-up areas, neighboring buildings will cause occlusions of the virtual viewpoint during exploration.*

However, 3D maps only provide limited capabilities to access the data and to relate it to the real world. For instance, users of 3D maps often try to align the virtual viewpoint of the map with their egocentric viewpoint for easier orientation, a strategy that is not well supported by the interface [OEN09]. The only alignment feature 3D maps offer is to align the exocentric top-down view with the general viewing direction of the user. Another issue of currently available 3D maps is that the camera view of the object is often occluded by nearby structures, which is especially problematic in densely built-up areas (Fig.2).

To deal with these limitations, we introduce object-centric exploration (OCE) techniques for handheld AR, which use a virtual copy metaphor [PSP99] to gain access to distant viewpoints of a real world object in the user's AR view. In contrast to 3D maps, OCE techniques allow a user to focus on the one object he is interested in. OCE techniques also do not suffer from occlusions from neighboring structures, because a virtual copy of only a single object is presented. To present additional viewpoints of this real world object, our OCE interfaces separate the virtual copy (focus) from its real world counterpart and from its surroundings provided by the AR video (context). We consider spatial and temporal techniques for combining focus and context [CKB09]. While the former separate focus and context in space, the latter do so over time, thus removing the context from the interface. Fig. 1 shows spatial OCE techniques that preserve the context by either overlaying the copy on the context (Fig. 1(b)) or separating the copy from the context (Fig. 1(c)).

We explore different designs of OCE interfaces for the exploration of buildings in an urban setting. We perform a series of studies to evaluate our initial designs and the ability of the user to relate the virtual information to the real world. We perform studies under controlled conditions and collect real world experiences with our interfaces in a real world pilot study. We summarize our findings in design recommendations that should be considered when developing OCE interfaces for potential application areas such as future generations of location-based AR browsers, 3D tourist guides, or situated urban planning. Relevant real world objects could be annotated with additional information that can easily be explored using OCE interfaces.

2. Related Work

In line with Cockburn et al. [CKB09], we classify the related work into spatial and temporal techniques.

Spatial Techniques. A 3D world-in-miniature (WIM) [SCP95] can complement the egocentric view of the user. Bell et al. [BHF02] use a WIM in AR that shares annotations with the real world. Bane and Höllerer [BH04] present a WIM interface, in which users are able to seamlessly switch to a copy of an occluded room and interact with this copy. Similar to OCE, the WIM interfaces for AR provide copies of real world objects. However, unlike our interfaces, these interfaces were designed for head-mounted displays (HMD), not handheld devices. Unlike HMDs, handheld AR offers a permanent peripheral view of the real world around the display and an AR view on the display. Thus, even after completely removing the context from the interface, users can still relate the content of the display to the peripheral view.

Keil et al. [KZB*11] overlay a historical 3D representation of a building on a previously taken picture (context) on a mobile phone, but restrict its viewpoint to the egocentric viewpoint of the picture. Another mode allows users to freely explore the 3D model, but, unlike our interfaces, without providing the context and without seamless transitions between the real world and the virtual copy.

Spatial techniques can also be realized through the use of multi-perspective presentations such as mirrors [ANC11], panoramas [MDS10] or deformations [VGKS12]. While these techniques extend the egocentric viewpoint, they do not allow viewpoint changes for exploring distant objects.

In another work, Veas et al. [VMK*10] allow users to transition between live video feeds for exploring outdoor environments. Similarly, Sukan et al. [SFTE12] allow users to virtually return to previously captured viewpoints in a tabletop application scenario. Both approaches register images of these viewpoints in the real world, thus creating a multi-perspective rendering, which is similar to the ring of images used to navigate viewpoints in our 2D interface (Fig. 4(a)). However, their designs are not focused on exploring a single, real world object of interest, but aimed at communicating available viewpoints of the environment [VMK*10], or manipulating VR content in AR [SFTE12].

Temporal Techniques. Bowman et al. [BKH97] present instant teleportation techniques in VR environments and discover that the lack of continuous motion cues causes user to be disoriented. Kiyokawa et al. [KTY99] allow users to seamlessly teleport between a virtual and an augmented collaborative workspace. Seamless transitions are also provided by the MagicBook [BKP01], which allows users to switch between an exocentric AR view on a VR scene and an immersed egocentric VR view on the same scene. In contrast to this previous work, our interfaces switch from an egocentric user view to an elevated exocentric viewpoint and are designed for exploring real world objects.

Avery et al. [AST09] and Mulloni et al. [MDS10] switch to exocentric viewpoints to provide an overview of the surroundings. However, the overview is focused on the user's position and does not allow viewpoint changes around a focus object. Sukan et al. [SFTE12] and Veas et al. [VMK*10] switch to already established viewpoints and perform transitions to these viewpoints. In contrast to our 3D interfaces, they provide only access to a discrete number of views.

3. Interface Design

Our focus is the design of spatial interfaces for exploring large objects in an outdoor setting and evaluating these interfaces with respect to the spatial awareness of the user. Spatial techniques seem to be the most relevant choice, because they preserve the real world context. We expect that preserving the context creates an artificial bridge for mapping content discovered in the copy back to the real world. To investigate this aspect, we first developed a 3D interface and a 2D interface that preserve the context by spatially offseting focus and context (e.g. Fig. 1(c)).

In our designs, we only consider handheld devices as display device, because these devices are widely available and a major platform for AR applications. We assume that the mobile device has a large screen, to be able to experiment with screen-space demanding designs. Our interfaces are designed for portrait mode, which is the default mode of currently available AR browsers. Furthermore, we only use a common single-touch interface, and we highlight selectable real world objects with a simple frame.

We focus our application case on large-scale outdoor exploration. Furthermore, we assume that the user is located at an ideal distance (not too distant or too close) from the real world object so that all of the features relevant to our studies are clearly visible in the context. We also assume that we have access to a 3D model of the object to create the virtual copy. In the following, we refer to the initial view containing only the real world object as AR mode, and to the mode containing the copy of the object as VR mode.

In the **3D separation interface** (3DSEP) (Fig. 4(b)), a 3D copy is presented, which allows for the continuous exploration of different viewpoints of the object. The user interacts directly with the 3D copy through a virtual orbit metaphor. When entering the VR mode, the copy is viewed from a bird's eye perspective. We integrated common spatial cues into the interface to allow users to mentally link the viewpoint of the copy to the original viewpoint of the context. A grid shows the ground plane of the copy and a camera icon, located in the coordinate frame of the copy, indicates the original egocentric viewpoint relative to the object. A radar icon in the top right shows the same information in a more abstract visualization and from a top-down view. The copy is in the center of the radar, while a dot rotating around the center indicates the camera position relative to the object.

The **2D separation interface** (2DSEP) (Fig. 4(a)) uses images as copy. These images could be pictures taken from the real world object. To avoid visual disparity of the focus between both interfaces, we render them from the same 3D model used in 3DSEP, taken at equidistant positions (45°) on a horizontal circle around the object, with the camera pointing towards its center. The viewpoints are elevated to bird's eye views. The user can replace the zoomed image at the top of the interface by using an explicit one-finger tap, or by swiping over the set of images. The ground plane is rotated upwards around the x-axis so that the images do not occlude each other or the object. In contrast to 3DSEP, 2DSEP does not provide continuous viewpoint updates.

We included corresponding spatial cues from 3DSEP in 2DSEP. We did not include the cues in the rendered images, but only applied them to the image circle, so that we could investigate if the circle is sufficient for users to orient in the interface. We added a grid to visualize the ground plane on which the images are placed and removed its center to avoid occlusions of the real object in the video image. Each image in the circle received a camera icon representation. A radar-like cue is achieved by the relation of the currently selected highlighted image to the image showing the frontal view.

Aside from these spatial cues, both interfaces provide a smooth transition between AR mode and VR mode to connect these spaces [BKP01]. When entering the VR mode, the video image is scaled down and moved to the bottom of the screen, while the copy is moved to the top of the screen. Spatial separation fully preserves the context at the cost of introducing a spatial offset between focus and context. The offset is alleviated by seamlessly animating the transition of focus and context. To facilitate associating the copy with the real world object, we gradually fade the copy in and out.

To further alleviate the spatial offset and to facilitate mentally linking the different viewpoints of focus and context, we added a switchable spatial cue called visual links (as shown in Fig. 1(c)) to 3DSEP, thus creating interface 3DSEP+L(inks). Links provide a visual connection between the copy and the real world object. By tapping on a location on the 3D copy, a user can create a 2D line to the corresponding location in the video image. The line style is adapted to communicate occlusion with the focus object, and color coded to communicate the end points.

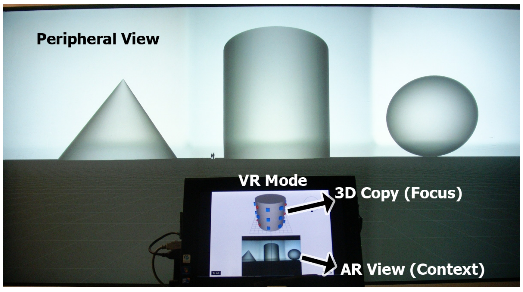


Figure 3: *Indoor apparatus. The view of an abstract scene from the participant's position showing the peripheral view in the background an the used tablet PC in the foreground. The tablet PC shows the VR mode of the 3DSEP interface.*

4. Laboratory Study: Abstract Scenario

We explored the usage and usability of OCE techniques in a series of user studies. We focused on how users interact with our techniques independent from the semantics and salient content of the real world. Therefore, we evaluated the interfaces using abstract scenes with basic geometric shapes.

4.1. Evaluation Testbed

To avoid confounding factors from the real world, we used a simulation testbed for AR. A testbed allows us to present artificial environments and structures with which the participants are not familiar. These scenes can represent real world environments, or can be purely abstract. Testbeds for simulating AR scenarios have already been used to control the registration error [LBHB10] or variable lighting conditions [LRM*13]. Testbeds were also used to overcome technical limitations of currently available hardware [BLT*12].

In our scenarios, a user has already found a real world object of interest and is looking towards it. We assume that the user remains stationary while exploring the object with our interfaces and thus does not require an immersive 360° view of the environment. Therefore, we simulate the peripheral view of the world with a back-projection wall (4×2m, 4000×2000 Pixel) used in daylight conditions (Fig. 3).

We seated participants in front of the wall and mounted the AR device on a tripod in front of them, to simulate holding a handheld device, while at the same time removing the associated physical fatigue. The AR device, a tablet PC (Motion Computing J3500, 12.1″), showed a static snapshot of the environment (1066×800 Pixels) that simulated the view through a video camera.

4.2. Experimental Design

The following studies are within-subject and share the same experimental design and apparatus (Section 4.1). They differ only in their interface conditions.

Table 1: *Questions asked in the studies. All questions except for Q11 use a 5-point Likert scale (1=strongly agree).*

Q1	It was easy to solve the task using the interface.
Q2	I did not feel confident using this interface.
Q3	The interface was intuitive to use.
Q4	I did not like the presentation of the interface.
Q5	The presentation of the images does not reflect the location relative to the object.
Q6	The camera icon helped me orienting.
Q7	The radar icon did not help me orienting.
Q8	I did not like the visual links.
Q9	The links helped me to solve the task.
Q10	The visual links are intuitive.
Q11	Rate how you liked the interface. (1 to 5; 1 = worst)

Scenario. We rendered a virtual scene consisting of only basic geometric shapes (cone, elliptic cylinder, sphere). The scale and position of these were chosen to resemble real buildings (e.g., elliptic cylinder, 35m in height, half-axes length x=17m and z=22m). The peripheral view was rendered using a virtual camera (60° FOV), placed 120m from the scene at eye level of the participants. A human scale icon was used as a reference. The AR view was taken with a camera (60° FOV) mounted on the tablet PC.

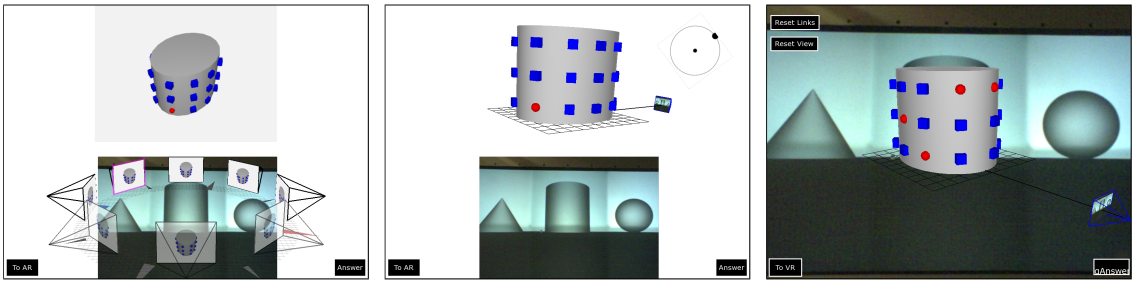
Tasks. The tasks are representative of interaction with real world 3D objects: (T1) a counting task, where users navigate the copy to find particular figures and count them; (T2) a mapping task, where users search the copy for a single object and point to its location in the peripheral view. For both tasks, the scene included distractors (blue cubes) and targets (red spheres), which were placed on the cylinder in a regular pattern (10 angles, 3 elevations). For T1, five to seven spheres were randomly distributed around the object. For T2, only one sphere was placed at a random location on the grid.

Procedure. For each task and interface, the participants had one practice trial without time constraints. T1 trials were completed by entering the number of counted spheres on an auxiliary keypad, T2 trials by point-and-click to the location of the sphere in the peripheral view with a laser pointer. Participants completed questionnaires (Table 1) between each interface and task, and after the experiment. We recorded task completion time for both T1 and T2, counting error for T1, and a pointing error for T2. The latter was estimated using a vision-based method, which provided the Euclidian distance for images with resolution 640 × 480.

4.3. First Study: Varying Copy and Cues

This explorative study compared our first interface designs (3DSEP, 2DSEP, 3DSEP+L) to evaluate 2D and 3D copy representations and the spatial cues. Fig. 4(a) shows 2DSEP and (b) 3DSEP as used in this study. 3DSEP+L is the same as 3DSEP with the option to create visual links.

Participants. A total of 24 people (12m/12f), 16−35 years



(a) (b) (c)

Figure 4: *Interface designs for studies in abstract conditions. Spatially separated interfaces using (a) images (2DSEP) and (b) a 3D copy (3DSEP), as used in the first study. (c) The 3D in-place interface (3DINP+L), as used in the second study.*

old (*mean*=25.9, *sd*=4.2), performed 5 repetitions (720 trials) for each task and interface. The presentation order of interfaces and tasks was counterbalanced.

Results. For each interface×task condition and participant, we calculated the mean completion time and error from the 5 repetitions (see Table 2). We performed non-parametric tests, because our sample violated normality. A significant effect of interface on time was only observed for T1 (Friedman, $X^2(2)=22.3, p<0.001$). A Post-hoc Wilcoxon signed-rank test with Bonferroni corrected $\alpha=0.0176$ showed that 3DSEP ($p<0.001$) and 3DSEP+L ($p<0.001$) were significantly faster than 2DSEP. Otherwise the performance data revealed no significant effects.

Questionnaires. The questionnaire data is summarized in Table 3, significant effects in Table 4.

Observations. A general strategy, which was applied to all of the interfaces, was panning around the object. Alternatively, in 3DSEP and 3DSEP+L, some participants moved to a top-down view and performed small viewpoint adjustments to look around the edges of the top and locate the spheres. In 2DSEP, participants also adopted the strategy of only selecting viewpoints, which were 180° or 90° offset.

During T2, participants either pointed directly at the periphery after finding the sphere, or rotated back to the frontal view. In 3DSEP+L, 54% of the participants used the visual links to highlight the sphere and find it in the spatially separated video image. In 2DSEP, 63% used the small images to quickly find the sphere in the small images. Participants either clicked directly on the corresponding viewpoint, or in some cases (16%) pointed directly at the periphery.

Discussion. In general, both 3D interfaces were preferred over the 2D interface (Q11). Questionnaire data and feedback collected from the participants support this result. For instance, participants perceived that solving the tasks was easier with the 3D interfaces (Q1). The interview also revealed that participants had difficulties with orientation using the discrete image switches in 2DSEP. Participants found this especially challenging in T1, where they had problems keeping track of multiple neighboring spheres. This issue is reflected by the significantly higher confidence (Q2) and intuitiveness (Q3) when using the 3D interfaces for T1. It may also be the reason why participants rated the presentation of 2DSEP significantly lower only for T1 (Q4), while for T2 there was no significant effect in presentation. Also the effect on intuitiveness (Q3) diminishes in T2, although 3DSEP was still perceived as more intuitive than 2DSEP.

Table 2: *Mean completion times in seconds and point errors in pixels, both with SD, for the first and second study.*

		T1	T2	
	Interfaces	Time	Time	Error
Study 1	2DSEP	22.8 (7.3)	18.3 (7.3)	41.5 (28.0)
	3DSEP	15.1 (4.5)	16.7 (7.6)	48.1 (26.3)
	3DSEP+L	16.9 (5.6)	20.5 (9.6)	39.4 (21.2)
Study2	3DSEP+L	19.7 (8.1)	25.5 (12.9)	22.5 (9.9)
	3DINP+L	20.8 (7.7)	25.7 (12.9)	23.4 (7.8)

For T1, only 3DSEP was significantly preferred (Q11) over 2DSEP. This is reasonable, given that participants did not require the visual links to solve this task. It is also reflected by Q9, where links were more helpful for T2 than T1. In general, visual links were well received (Q8) and found to be intuitive (Q10). Nevertheless, only 58% of the participants used the links, because, according to their feedback, the task could easily be solved without them. We did not find any significant difference in point error between 3DSEP and 3DSEP+L. However, when dividing the trials of 3DSEP+L and 3DSEP into those with ($n=68, mean=31.2, sd=14.4$) and those without ($n=172, mean=48.8, sd=40.58$) visual link usage, the results indicate that participants made less errors when they used links.

The interviews showed that the camera icon was a strong cue for communicating the starting point of rotation. Based on the interviews, we believe that participants were unsure when rating the radar cue, which is also reflected by the trend

Table 3: *Questionnaire data with mean and SD (rounded) for studies with abstract content.*

	Study 1						Study 2			
	T1			T2			T1		T2	
	2DSEP	3DSEP	3DSEP+L	2DSEP	3DSEP	3DSEP+L	3DINP+L	3DSEP+L	3DINP+L	3DSEP+L
Q1	2.8 (1.1)	1.4 (0.5)	1.3 (0.5)	2.2 (0.9)	1.7 (0.8)	1.4 (0.6)	1.7 (0.5)	1.6 (0.5)	1.5 (0.5)	1.8 (0.9)
Q2	3.4 (1.1)	4.3 (0.8)	4.6 (0.6)	3.8 (1.0)	4.0 (1.1)	4.2 (1.0)	4.0 (1.0)	4.1 (0.9)	4.3 (0.7)	4.3 (0.8)
Q3	2.3 (1.0)	1.6 (0.5)	1.5 (0.7)	2.0 (0.8)	1.6 (0.7)	1.6 (0.6)	2.1 (0.8)	2.2 (0.8)	1.6 (0.7)	1.6 (0.7)
Q4	3.1 (0.9)	4.0 (0.9)	4.2 (0.7)	3.8 (0.9)	3.8 (1.1)	4.0 (0.9)	4.1 (0.7)	3.5 (1.2)	4.4 (0.7)	4.1 (0.9)
Q5	3.7 (0.9)			3.8 (0.8)						
Q6		2.1 (1.4)			2.1 (1.3)		2.6 (1.1)	2.7 (1.0)	2.8 (1.0)	2.4 (1.0)
Q7		2.8 (1.2)			3.3 (1.2)					
Q8			3.6 (1.3)			3.9 (1.0)	3.6 (1.0)	4.0 (1.0)	4.4 (0.7)	4.3 (0.7)
Q9			3.0 (1.6)			2.2 (1.4)	2.7 (1.6)	2.3 (1.4)	1.3 (0.7)	1.5 (0.8)
Q10			1.8 (0.7)			1.9 (1.1)	2.7 (0.9)	2.1 (1.0)	1.6 (0.7)	1.8 (0.6)
Q11	3.0 (1.4)	4.5 (0.5)	3.8 (1.3)	3.3 (1.3)	4.1 (0.8)	4.3 (1.0)	4.1 (1.1)	4.0 (1.1)	4.5 (0.9)	3.7 (1.0)

Table 4: *Significant effects in questionnaire data. Study 1 was tested with Friedman (not reported) and Wilcoxon signed-rank tests with Bonferroni corrected* α*=0.0167; Study 2 with Wilcoxon signed-rank tests with* α*=0.05.*

		Study 1			Study 2
	Task	2DSEP& 3DSEP	2DSEP& 3DSEP+L	3DSEP& 3DSEP+L	3DINP+L& 3DSEP+L
Q1	T1	**p<0.001**	**p<0.001**	p=0.317	p=0.564
	T2	**p=0.011**	**p=0.001**	p=0.07	p=0.257
Q2	T1	**p=0.002**	**p=0.001**	p=0.059	p=0.783
	T2	p=0.285	p=0.026	p=0.096	p=0.564
Q3	T1	**p=0.004**	**p=0.003**	p=0.527	p=0.317
	T2	**p=0.012**	p=0.032	p=0.705	p=1.0
Q4	T1	**p=0.002**	**p<0.001**	p=0.234	p=0.109
	T2	p=0.666	p=0.119	p=0.238	p=0.102
Q11	T1	**p=0.001**	p=0.067	p=0.035	p=0.713
	T2	**p=0.006**	**p=0.008**	p=0.512	**p=0.029**

of neutral answers for the radar cue (Q7). The arrangement of images in 2DSEP was well perceived (Q6). Participants also stated that it provided a good overview of the object in T2, because the single red sphere was very salient.

4.4. Second Study: Varying Spatial Separation

Since 3DSEP and 3DSEP+L were the preferred interfaces and both performed better during exploration task (T1), we focused on investigating 3D interfaces further. We kept 3DSEP+L as representative 3D mode, because the visual links showed value as spatial cue in the mapping task (T2). Based on our observations, we introduced a reset button, which automatically realigns copy and context viewpoint. We also removed the radar cue from the interface. Aside from these changes 3DSEP+L corresponded to the same interface as used in the first study (Fig. 4(b)).

In this study, we explored two variations of spatial separation. We created an in-place interface (3DINP+L) that is similar to 3DSEP+L, but which has the copy overlaying the real-world object (Fig. 4(c)). We included the visual links in 3DINP+L, even though their end points are occluded by the 3D copy. Our assumption was that participants would need to switch between AR and VR modes to remove the occlusion and to mentally connect focus and context.

Participants. Twelve participants (6m/6f), aged between 19 and 30 ($\mathit{mean}=24,7,\mathit{sd}=3.3$), performed 5 repetitions (240 trials) of each task and interface. The presentation order of interfaces and tasks was counterbalanced.

Results. In the analysis, we used the same methods and statistical tests as in the previous study. Time and error measurements are summarized in Table 2. We performed non-parametric tests, because our sample violated normality. Statistical analysis did not reveal any significant effects.

Questionnaires. The questionnaire data is summarized in Table 3, significant effects in Table 4.

Observations. As in previous studies, participants either panned around the object or used a top-down view to solve the tasks. For T2, 92% of the participants used visual links for both interfaces. In 3DSEP+L, 42% switched back to AR to increase the size of the video image. As expected, in 3DINP+L the majority of participants (67%) switched back to AR to resolve occlusions of the link endpoints.

Discussion. Participants generally found the tasks easy to solve (Q1), felt confident with the interfaces (Q2) and found them to be intuitive (Q3). For T2, participants significantly preferred 3DINP+L over 3DSEP+L (Q11). The lack of significance for T1 can be explained by the comments of participants who stated that they only focused on the 3D copy, and did not consider the video background for this task. In the interview, participants stated that they preferred 3DINP+L because it was a more natural and intuitive approach to not separate focus from context. They also mentioned the increased size of the object in 3DINP+L. This is reflected by the higher values in presentation for 3DINP+L (Q4).

Interestingly, the visual links still served as orientation cue in 3DINP+L, even though they penetrated the copy and the endpoints were occluded. Participants noted that visual links showed the misalignment between the copy and the context. As before, trials in which links were used showed smaller point errors (n=97, $mean$=19.6, sd=10.6) than trials without visual links (n=23, $mean$=36.9, sd=26.4), which underlines their value as spatial cue.

5. Pilot study: Real-world Setting

In the previous studies, we focused on general properties of our interfaces and avoided confounding factors from real world scenes by using only abstract scenarios. To collect qualitative feedback and identify issues with the interfaces, we introduced the real world into our interface design and performed a pilot study in a real world setting We conducted a study in a popular urban area of our city center with the 3DINP+L and 3DSEP+L interfaces. Fig. 5 shows the 3DSEP+L interface with one of the target buildings.

Task and Methodology. Participants had to find and point to the real world location of a sphere located on the copy of the focus object. Participants were bound to a fixed location, but could rotate with the mobile device (InertiaCube3 sensor). The task was repeated with three visible distinctive cultural buildings located in varying distance around the participant: an art gallery (40m), a building floating on the river (200m), and a tower (370m). Pointing was estimated roughly by visual and verbal assessment. After the experiment, participants completed a questionnaire.

Participants. Ten people (7m/3f) aged between 16 and 32 ($mean$=24.2, sd=4.3) participated. They were recruited among local pedestrians and familiar with the surroundings.

Discussion. All participants were able to solve the task easily and generally gave positive feedback. All if them could imagine to use such an interface as a tourist, for exploring unknown landmarks and sights (5-point Likert, 1=strongly agree: $mean$=1.3, sd=0.48). Visual links were regarded useful as orientation cue.

In contrast to the previous study, we did not find any significant difference in preference between 3DINP+L ($mean$=4.0, sd=0.82) and 3DSEP+L ($mean$=3.8, sd=1.1). Participants who preferred 3DINP+L again stated that it was more intuitive and natural; the ones who preferred 3DSEP+L stated that it provided a better overview and that the copy was clearly visible due to the spatial separation from the video context. Hence, a main issue seems to be the visual interference of the copy with the real world.

6. Design Recommendations and Future Work

In the following, we put forward design recommendations for OCE techniques and outline future research directions.

We did not find performance differences between the in-place and the spatially separated interface. However, under controlled laboratory conditions, participants significantly preferred the in-place interface (3DINP+L), because the arrangement of focus and context was more natural. Occlusion of the context by the copy did not seem to be an issue.

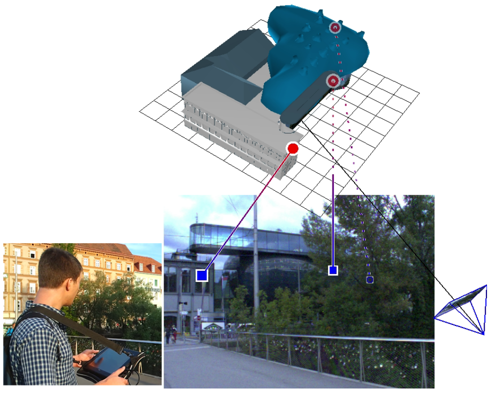


Figure 5: *Pilot Study. A spatial technique (3DSEP+L) applied in a real urban environment. The small inset shows a participant using our system.*

In the real world setting, we did not find a significant preference, because participants also preferred 3DSEP+L, because of the higher contrast between the copy and the white background. Therefore, an in-place interface may have to adapt the *context* to always achieve a good contrast to the overlaid focus object (e.g., desaturation of video background).

Generally, participants used the interfaces to get a quick *overview* of the focus object. Using 3D interfaces, participants switched to a top-down view to quickly look at the different sides of the object. In 2DSEP, participants used the small images as multi-perspective visualization to quickly identify the view containing the queried red sphere in T2. Therefore, OCE techniques should offer modes to explicitly get an overview over an object. When using images as overview, the relevant items on the object should be emphasized in an authoring step beforehand (e.g., by labels), due to the small size of the images, especially on mobile phones.

An overview can easily provide *shortcut navigation* to quickly access viewpoints. In 2DSEP, participants used the small images to quickly navigate between viewpoints in non-sequential order by accessing 90° and 180° offset views. This is also a main motivation for similar interfaces, such as SnapAR [SFTE12] or the one of Veas et al. [VMK*10].

A simplified *camera navigation* model with few degrees of freedom (e.g., orbit metaphor) was sufficient for the investigated structures. However, future designs should also consider zooming and the exploration of more complex structures, which require more sophisticated navigation metaphors. For instance, the HoverCam [KKS*05] allows to explore complex objects with few degrees of freedom.

Our findings are in line with Bowman et al. [BKH97], who found out that instant teleportation causes disorientation. In our studies, continuous 3D viewpoint changes outperformed discrete 2D switches. Therefore, a 3D interface is the most sensible choice for presenting viewpoint changes.

Participants could easily solve the given tasks in both the laboratory and the real world pilot study using a 3D interface. Hence, the *transition* between AR and VR mode and the *camera icon* seem to be sufficiently strong cues to connect separated views. The radar icon on the other hand was not considered helpful. This indicates that cues should be connected directly to the spatial reference frame of the copy.

Based on the collected data and participants' feedback, we can say that *visual links* are a valuable spatial cue. We designed visual links with spatial separation between focus and context in mind (3DSEP+L). However, participants considered spatial separation as inferior to a more natural in-place interface due to the smaller size of the zoomed focus object. Hence, the links could be redesigned to better support in-place interfaces. On the other hand, we observed that participants mainly used links during the mapping task (T2). An intermediate mode could be introduced that switches from in-place to a spatially separated presentation, when users want to map information back to the real world.

Although we tested our initial designs in a laboratory setting, the experiences gathered from the real world pilot study make us confident that OCE interfaces are also feasible and practical in real world conditions. Future work will investigate the interfaces in real world conditions in more depth. Furthermore, we will compare spatially separated interfaces to temporal interfaces that do not preserve the context to investigate the impact of preserving the context in the interface. We will also compare our OCE interfaces to a 3D map interface, which also allows exploring real world objects. Our current designs only considered the ideal *distance* to the real world object. Future designs will investigate situations, where the interface needs to zoom the focus object, because it is either too close or too distant.

7. Conclusion

In this paper, we presented interfaces that allow users to explore distant real world objects. Relevant application areas for the presented techniques are tourism and urban planning. They can also provide guidance for supporting the exploration of real objects in the next generation of AR browsers. Based on our findings we provide a set of design recommendations, which we hope will inspire other researchers to explore the design of OCE techniques further.

Acknowledgements

This work was funded by the Christian Doppler Laboratory for Handheld Augmented Reality and the Austrian Science Fund (FWF) under contract P-24021.

References

[ANC11] Au C. E., Ng V., Clark J. J.: Mirrormap: augmenting 2d mobile maps with virtual mirrors. MobileHCI '11, pp. 255–264. 2

[AST09] Avery B., Sandor C., Thomas B. H.: Improving spatial perception for augmented reality x-ray vision. VR '09, pp. 79–82. 3

[BH04] Bane R., Höllerer T.: Interactive tools for virtual x-ray vision in mobile augmented reality. ISMAR '04, pp. 231–239. 2

[BHF02] Bell B., Höllerer T., Feiner S.: An annotated situation-awareness aid for augmented reality. UIST, p. 213. 2

[BKH97] Bowman D. A., Koller D., Hodges L. F.: Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. VRAIS '97, pp. 45–52. 2, 7

[BKP01] Billinghurst M., Kato H., Poupyrev I.: The MagicBook - moving seamlessly between reality and virtuality. *CG&A 21*, 3 (2001), 6– 8. 3

[BLT*12] Baricevic D., Lee C., Turk M., Tobias H., Bowman D.: Hand-held ar magic lenses with user-perspective rendering. ISMAR '12. 4

[CKB09] Cockburn A., Karlson A., Bederson B. B.: A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv. 41*, 1 (2009), 2:1–2:31. 2

[KKS*05] Khan A., Komalo B., Stam J., Fitzmaurice G., Kurtenbach G.: Hovercam: interactive 3d navigation for proximal object inspection. I3D '05, pp. 73–80. 7

[KTY99] Kiyokawa K., Takemura H., Yokoya N.: A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. SMC '99, pp. 48–53. 3

[KZB*11] Keil J., Zöllner M., Becker M., Wientapper F., Engelke T., Wuest H.: The house of olbrich - an augmented reality tour through architectural history. ISMAR AMH '11, pp. 15–18. 2

[LBHB10] Lee C., Bonebrake S., Hollerer T., Bowman D.: The role of latency in the validity of ar simulation. VR, pp. 11–18. 4

[LRM*13] Lee C., Rincon G. A., Meyer G., Tobias H., Bowman D.: The effects of visual realism on search tasks in mixed reality simulation. VR '13. 4

[MDS10] Mulloni A., Duenser A., Schmalstieg D.: Zooming interfaces for augmented reality browsers. MobileHCI '10, pp. 161–169. 2, 3

[OEN09] Oulasvirta A., Estlander S., Nurminen A.: Embodied interaction with a 3d versus 2d mobile map. *Personal Ubiquitous Comput. 13*, 4 (2009), 303–320. 2

[PSP99] Pierce J. S., Stearns B. C., Pausch R.: Voodoo dolls: seamless interaction at multiple scales in virtual environments. I3D, pp. 141–145. 2

[SCP95] Stoakley R., Conway M. J., Pausch R.: Virtual reality on a WIM: interactive worlds in miniature. CHI, pp. 265–272. 2

[SFTE12] Sukan M., Feiner S., Tversky B., Energin S.: Quick viewpoint switching for manipulating virtual objects in hand-held augmented reality using stored snapshots. ISMAR '12, pp. 217–226. 2, 3, 7

[VGKS12] Veas E., Grasset R., Kruijff E., Schmalstieg D.: Extended overview techniques for outdoor augmented reality. *TVCG 18*, 4 (Apr. 2012), 565–572. 2

[VMK*10] Veas E., Mulloni A., Kruijff E., Regenbrecht H., Schmalstieg D.: Techniques for view transition in multi-camera outdoor environments. GI '10, pp. 193–200. 2, 3, 7