Bimanual Handheld Mixed Reality Interfaces for Urban Planning

Markus Sareika
Graz University of Technology
Inffeldgasse 16
A-8010 Graz
+43 316 873 5076
markus@sareika.de

Dieter Schmalstieg
Graz University of Technology
Inffeldgasse 16
A-8010 Graz
+43 316 873 5070
schmalstieg@icg.tugraz.at

Figure 1. Urban planning with handheld mixed reality.

ABSTRACT
Tabletop models are common in architectural and urban planning tasks. We report here on an investigation for view navigation in and manipulation of tracked tabletop models using a handheld Mixed Reality interface targeted at a user group with varying professional background and skill level. Users were asked to complete three basic task types: searching, inserting and creating content in a mixed reality scene, each requiring the user to navigate in the scene while interacting. This study was designed to naturally progress on classic problems like travel, selection and manipulation in an applied scenario concerned with urban planning. The novel bimanual interface configurations utilize a handheld touch screen display for Mixed Reality, with the camera/viewpoint attached or handheld separately. Usability aspects and user satisfaction are scrutinized by a user study, aimed at optimizing usability and supporting the user's intentions in a natural way. We present the results from the user study showing significant differences in task completion times as well as user preferences and practical issues concerning both interface and view navigation design.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Artificial, augmented and virtual realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Graphical User Interfaces, Screen design;* I.3.6 [**Computer Graphics**] Methodology and Techniques – *Interaction Techniques;* I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism – *Virtual Reality;* J.5 Arts and Humanities - Architecture

General Terms
Design, Experimentation, Human Factors, Measurement, Performance

Keywords
Design, 3D Interaction, Bimanual Interaction, Mixed Reality, Augmented Reality, Urban Planning, Architecture

1. INTRODUCTION
Mixed Reality (MR) technology can enhance communication [3] and provide deepened understanding for urban planning activities, which are richer than usual, leading to an improved shared vision of the future urban environment [33]. Mixed groups of stakeholders can explore the complex societal and other implications of an urban planning project at early project stages and aim to avoid planning mistakes affecting investors, technical specialists and citizens. Environments for urban and architectural planning and education have repeatedly been the topic of human-computer interaction research. Tabletop interfaces are popular for this area of application, as they easily accommodate architectural scale maps and models commonly used in architectural communication, and facilitate tangible interfaces. Previous work has already explored numerous MR interface designs aimed at supporting various planning and negotiation stages with tools for collaborative working situations.

A central round table is an established real-world tool for communication, whereas the quest for the optimal display of an MR scene on the table is still ongoing. Tracked head mounted displays (HMDs) can augment individual view points of the scene [26], but restrict the free movement and eye contact, thereby imposing constraints on the communication process. In contrast, fixed MR displays can presents information simultaneously to all collaborators from the same point of view, establishing a common base for eye to eye discussions.

In this work we are exploring interface configurations using a semi-mobile handheld display. This display is movable, but unlike *e.g.* a mobile phone its screen is large enough for interacting with good quality MR images and for accommodating a small group of collaborators. It features a touch screen, which is used as an in-place input device (Figure 1).

A crucial factor in the overall application experience is the navigation in the MR scene, as it determines what is visible on the screen and therefore focuses the interaction or discussion on a particular area. A large body of work on 3D navigation is available; but most of this work focuses on egocentric, immersive Virtual Reality (VR) rather than MR conditions. An important motivation for our work was therefore to investigate the navigation using bimanual MR interfaces in context of a real-world setting.

Our work builds on Urban Sketcher [30], a conceptual design application capable of augmenting the urban reconstruction site with sketches, facades, buildings, green spaces or skylines. It utilizes an easily accessible 2.5D interface operated by screen input. Urban Sketcher is used for direct interaction, while real-time visual feedback is given to the user by video augmentation on the mobile screen. It allows sketching in the space of the video augmentation and virtually modifying the tabletop model.

The motivation of our work is to better understand usability issues when interacting with an urban planning application through MR. There are many prototypes for architectural application involving variants of MR. However, most of this work focuses on application specific experiences and on the basic qualities of the interface in a real-world scenario.

A bimanual operation for simultaneous view navigation and manipulation tasks is used in the experiments setting. We evaluate two possible bimanual interface configurations, one with the camera in hand while the display is stationary, and one with the camera mounted to the mobile display. Subjects performed three elementary tasks – searching, inserting and creating content. These are commonly found in, but not limited to, urban planning scenarios when working with tabletop models. In order to characterize both interface device configurations, we investigated task completion times, mental load and user ratings. The results allow natural optimizations in the bimanual MR user interface design for applications.

2. RELATED WORK

2.1 User interfaces for architectural design

Tables with architectural scale maps and models are established tools in planning discourse, enabling an observer to quickly grasp an overview of the design from an exocentric viewpoint. Interactive tabletop displays with MR capabilities and tangible user interface approaches have been developed to facilitate architectural education and also design negotiation [19][17][32][23]. In "Architectural Anatomy" [10] the structural skeleton of a building was augmented. Neumann *et al.* [27] describe Augmented Virtual Environments combining virtual models with live video textures, mainly for surveillance applications. Lee *et al.* [21] describe an MR environment for 3D modeling and texturing.

MR tabletop interfaces aim to combine the advantages of MR and collaborative interactions. They mostly use HMDs, showing an individual perspective of the scene to the users. However, HMDs limit collaboration to some extent. Interaction is based on hand gestures or physical objects. The systems support the creation of geometries, architectural 3D scenes or building forms [26].

Other types of tabletop interfaces use projections and multiple screens to visualize the scenes that are created [23]. The Luminous table [17] is an augmented reality workbench integrating multiple forms of physical and digital representations, such as 2D drawings, 3D physical models and digital simulations, which are all on the same table surface. More specific architectural topics are addressed by Urp [32], a physically based workbench that allows users to study light properties and flows of an architectural scene, and by Illuminating Clay [28] a system for altering the topography of a clay model in order to design and analyze landscapes. The results of these modifications are constantly projected back into the workspace. The Envisionment and Discovery Collaboratory [9] uses computer simulations and tangible objects to represent elements of the domain, such as a simulated bus route.

The MR-Tent [23] combines multiple MR interfaces, thereby bringing collaborative MR from the laboratory to the field. The MR-Tent facilitates collaboration on a round tangible table and an augmented wall projection operated by laser pointer input all integrated by Urban Sketchers interfaces. The viewport of the projected scene can be altered interactively. Egocentric as well as exocentric perspectives are equally visible to all collaborators in the tent. For the work reported in this paper, we transpose the idea of manipulating an architectural scene seen through an interactive MR interface from an on-site situation to a tabletop architectural model. Our intention was to specifically investigate view navigation in combination with classic tasks, which was only informally evaluated in our previous experiments, focused on ethnographic issues and not on factorial analysis.

2.2 Travel and navigation

For video based MR, it is usually assumed that physical camera and display are either stationary or move together as one rigid combination. In contrast, there is a large body of work on travel and navigation needed in the field of VR.

Travel and navigation in immersive VR have been studied by Bowman *et al.* [6][5] who identified that reducing the disorientation of the user in a pure egocentric setting is challenging. The disorientation issue is also present in desktop VR setups were constraints alleviate navigation [16][12][14]. Tools like Navidget [13] [11] aim at reducing the mental load on the user. Multiscale 3D Navigation [24] puts an emphasis on seamless navigation between egocentric and exocentric views on desktop VR setups, building on previous work of HoverCam [20] used for 3D object inspection, just like StyleCam [7] or ShowMotion [8], which resort to predefined motion paths to control the observing camera. Mackinlay *et al.* [21] also compute the camera animation path from a user-selected point of interest.

In contrast to all these approaches we refrain from reducing the degrees of freedom (DOF) and thereby the immediacy of interaction. We rather aim at supporting the user by adding real-time information relevant for the perceptual-motor loop and by combining naturally occurring 2D and 3D interaction.

Early work by Ware and Osborne [34] introduced the "eyeball in hand" metaphor in VR. This approach required a mental model of the scene, because it did not provide any direct visual feedback. We adopt this metaphor, but provide a video augmented scene [27] on a mobile display, as suggested by McKenna [25]. In addition the real model on the table in our experimental setting resembles a WIM [31] with depth cues and supports tangible navigation [4] with a 3D map (as suggested by Haik *et al.* [14]). This empowers the user to intuitively change 6 DOF of the virtual camera by a spatially registered web cam as tangible input device, while independently interacting with the other hand.

The proposed interface configuration adds natural and intuitive qualities loosely inspired by the Rockin' Mouse [1] and trackball-mice [18]. It has implicit "safe 3D navigation" [11], because of the direct reference of the input device to the interaction space.

Similar to our configurations, bimanual interaction, using only mouse and keyboard for desktop 3D environments, has been studied by Balakrishnan and Kurtenbach [2], where the non-dominant hand controlled the virtual camera while the dominant hand was used for manipulation tasks. The survey of Hinckley *et al.* [15] concentrates on spatial design issues from a large body of work and proved to be helpful for our design choices. However, to our knowledge, none of the systems described in literature use bimanual interface configurations for navigation and interaction in handheld MR as described in this application driven paper.

3. EXPERIMENTAL DESIGN

The experiment reported in this paper draws its motivation from utilizing bimanual MR interface design for an urban planning scenario addressing three classic tasks. In particular, we were interested in applied viewpoint navigation combined with real-world challenges, which turned out to be essential in previous informal experiences with architectural and urban MR. We concentrate on a specifically designed, imaginary planning scenario addressing users with a wide range of backgrounds and varying computer experience as in real-world situations.

Rationale for our design choices was led by the application area of urban planning. Two camera navigation techniques were designed for comparison in the scenario, one similar to the viewfinder of a photo camera and the other similar to an "eyeball in hand", often used in practice by MR experts but hardly mentioned explicitly. Guided by related work, especially by Balakrishnan and Kurtenbach [2] who found that "operating camera control in the non-dominant hand is beneficial" we made interface decisions. Also the preferences of creative people for interfaces that "feel right" were taken into account, as well as our previous workshop experiences, observations and discussions. We chose the display size to address a mobile setting, so a small group of collaborators can have the same point of view into the MR scene with good quality images, a tradeoff between heavy large displays and too small, but light mobile phone sized displays.

The two specific techniques we created for this experiment isolate interesting factors in the context of a real-world application, with relevance also beyond this specific scenario. The planning scenario on the table was scaled based on informally quantified measures and accounted for enough space for free camera movements as well as for architectural design space.

Good stimulus-response is assumed meaning high affordances distinguish and characterize our two settings each unique for itself. We identified user preference to be an important factor since technology should be adapted to aid the human. In addition task completion time, mental and physical load reflect the achieved performance with a specific technique. Accuracy and error-rate were not considered to play an important role in this application scenario and were therefore not measured explicitly, but are reflected by the user performance question.

The goal of the evaluation is to clarify our research questions and hypotheses. And furthermore, provide insights concerning the efficiency of the proposed interface and device configurations. The results should inform interface designers and assist them with natural design decisions concerning bimanual MR and also related types of interfaces. We evaluate in a quantitative and qualitative manner using measurements, questionnaires and video observation to find out which type of mixed reality view navigation is suitable for specific types of tasks when working with tracked tabletop models.

Figure 2. The fixed camera gives more exocentric viewpoints.

3.1 Hardware Setup

The hardware setup adopted in the experiment consists of a 2.6GHz quad core PC and a semi-mobile pen touch screen with a resolution of 1280x800 and a weight of 1.75kg. A Logitech camera with a weight of 0.1kg provides a video stream at a resolution of 640x480 with 30Hz. The video is displayed on the screen and also used for natural feature based tracking of the camera without obstructing the view with another sensor or fiducial targets. The video augmentation overlays a digital model registered in 3D to the real model in real-time, *i. e.*, at the camera frame rate. The software used in the experiment is based on the Urban Sketcher application and the mixed reality framework Studierstube (www.studierstube.org).

Figures 1, 2 and 3 show the architectural model. It is 1.08m x 0.80m and has a maximum height of 0.15m. The model is represented by phantoms [29] in the virtual space, so occlusions of virtual objects intersecting with the real model are handled correctly in the resulting augmented view. Model size, number and density of objects were designed to fit a natural interaction space giving some freedom for the movement of the camera.

In previous informal experiments with Urban Sketcher, we had used both handheld MR and stationary display settings. The two most promising bimanual interface configurations were chosen as the main conditions:

1.) A free camera (with a small tripod attached for convenience) which can be moved around the mixed reality model with one hand, while the display is stationary (Figure 2).

2.) A fixed camera rigidly attached to the display, which can be moved together with the display in order to adjust the viewpoint into the mixed reality scene (Figure 3).

In each case, operation is bimanual: One hand manipulates the viewpoint, while the other hand interacts with the touch screen using the pen.

3.2 Software Setup

On the software side, the user interface consists of an overlay menu and a set of tools which allow manipulating objects in the three dimensional MR scene following, a 2.5D interaction metaphor: The working environment is three-dimensional, but the simultaneous change of object parameters is limited to two dimensions. For instance, changing the position of an object is constrained to moving on the ground plane with additional controls for height where appropriate. The interface was deliberately designed in a constrained way, so that users with little or no experience can learn its operation quickly and with minor effort. This is important to remove barriers in a collaborative working situation with experts.

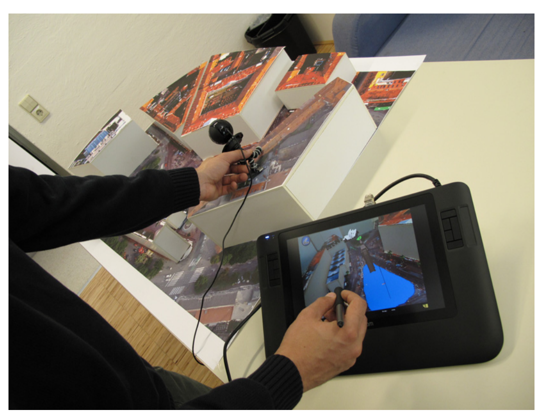

Figure 3. The free camera allows more egocentric viewpoints.

3.3. Evaluation procedure

In order to obtain meaningful observations and measurements, we designed the experimental scenario to comprise three characteristical elementary tasks which had to be completed in both of the two view navigation configurations. All tasks were evaluated by the user's perception as reflected in NASA's Task Load Index and the measurement of the task completion time. A post hoc questionnaire was created to summarize the user impressions, followed by a brief interview. All together the average evaluation time per subject took 40 minutes and was considered sufficient for sustained concentration, avoiding tiring effects.

After filling in a questionnaire on demographic user information, an introduction to the procedure of the experiment followed. The test subjects were asked to work at normal pace. Before each task, they were instructed specifically how to accomplish it. We deliberately refrained from any explicit training as this would have distorted the closeness to a real-world setting.

3.4 Task description

The tasks procedures are explained in full detail in the following. There were three tasks:

(T1) Seven cars have to be found in the MR scene. This is a pure browsing task and requires no user input on the mobile screen apart from the view navigation. Once all the car locations are reported and sketched on an overview paper map by the user (using her dominant hand), the elapsed time is noted. The task was chosen because it is essentially needed to find objects in larger models and scenes.

(T2) This task requires the user to insert and position three trees at marked locations in the scene. This task represents the adding and placing of content in the scene, which is part of a common workflow, but is more complex in terms of interaction than pure browsing. It requires user input and demands bimanual interaction for working with the content. In the fixed camera configuration, the user initially needs to learn moving the screen with one hand for navigation while using the pen in the other hand.

(T3) Similar to task T2, two hands are needed to accomplish the goal to generate 3D content. For this task, the user needs to construct a fence with the 3D construction tool around the region in the MR scene marked in blue. This is the most complex task. It was chosen because it represents interactive content creation, which is essential for planning processes.

The interaction procedures for all the tasks are now described in detail. For task T1, the user simply took the device either the camera or the camera attached to the display and hovered through the physical model, while changing viewing directions in order to find and report all the seven hidden cars.

The actions of the application for inserting and constructing content for the MR scene are shown when the user touches the round tool icon, in the top left corner of the screen, revealing an overlaid interface. This popup menu (see Figure 4) gives access to common actions.

For task T2 a file dialog is shown, after selecting the "load 3D object" menu item. Once a three dimensional object, such as the tree, is selected for placement in the scene, it is loaded into the centre of the MR scene. In order to move the object, the user needs to select the moving tool in the overlay menu. Once activated, an arrow icon is shown for feedback. If any object in the scene is selected, it will be enclosed by a thick bounding box for user feedback. Now this object can be moved by dragging its bounding box on the ground plane of

Figure 4. The 2D overlay menu is operated with the pen.

the scene to a new location, such as one of the T2 destinations marked in green.

For the construction task T3, the user can activate the construction mode by clicking the appropriate icon in the menu. This tool allows creating a polygonal outline in the ground plane, which can be extruded with a separately adjusted height for every polygon vertex. Three extra buttons as well as a yellow arrow on the ground plane on the MR scene appear for building the three dimensional geometry (see Figure 5). When indicating a position on the ground plane, the arrow moves correspondingly. The tip of the arrow indicates the position on the ground and can be used to adjust the height of a segment. With the "add point" button, the segment is added to the geometrical structure of the new object. Once the user has added all points and confirmed the completion, a textured object is generated. The objective of task T3 is accomplished by surrounding the blue area on the ground plane.

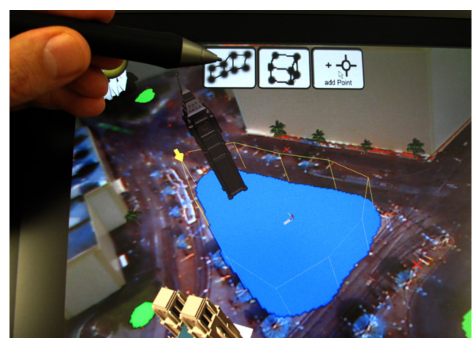

Figure 5. Completing the construction of a polygon extrusion.

3.5 Research questions

Aiming at optimizing the natural interface performance we formulated our research questions and hypotheses.

(R1) Which viewport navigation will be preferred for each of the three different tasks?

(R2) Does the type of viewport navigation speed up the task completion time for the tasks?

(R3) How do the viewport configurations affect mental and physical load?

In addition to the more general questions, we formulated assumptions in the following hypotheses:

(H1) For the fixed camera configuration, task completion time for all tasks will be faster and the mental load lower.

(H2) For the browsing task (T1), users will prefer working with the moving display and the fixed camera.

(H3) For adding and moving content (T2), users will prefer working with the moving display and the fixed camera.

(H4) For constructing content (T3), users will prefer working with the static display and the free camera.

(H5) For the free camera configuration the physical load will be lower.

4. USER STUDY

In order to find answers, we had all the subjects perform three different tasks for each of the two view navigation configurations using the stated evaluation setup. We have selected a user group of 31 people (19f/12m) aged from 15 to 47 (*Mean*=28.97, *Standard Deviation*=6.12), including urban planning professionals and ordinary citizens with varying background and expertise. All participants had normal or corrected to normal vision. The order of the three tasks and their two configurations followed a balanced Latin square distribution to reduce carry-over and learning effects among all tested subjects.

4.1 Empirical Results

Concerning the application area of urban planning, the subjects have varying experience, which was recorded with 5 variables on a 7 point Likert scale. Although not originally anticipated, we observed strong differences among subjects with little or much expertise during the execution of the experiment and therefore performed a regression analysis for the collected data on task completion time to test for the applicability of covariates in the statistical model.

One person who gave a strange combination of answers to the expertise questions was removed as an outlier, so 30 subjects remained for the analysis. The result with the predictors computer experience (β=-0.73), 2D software experience (β=0.30), 3D software experience (β=-0.26), 3D interface experience (β=0.07) and virtual reality experience (β=0.05) proved significant with ANOVA ($p<0.05$) and α=0.05 and reduced variance (R^2=0.402) by 40.2%. We now analyzed the effects on time with a 3 (Task) x 2 (Camera) repeated measures ANOVA with α=0.05 including the covariates.

With the covariates, the entire main effects were significant. A weak interaction between them was detected, as the lines in Figure 6 converge slightly. Looking at the camera configuration ($F_{1,24}$=5.61, $p<0.05$), it was especially interesting to see that the free camera viewport configuration (*M*=1.76, *SE*=0.10) took more time in general than the fixed camera viewport configuration (*M*=1.52, *SE*=0.09). The interaction Task x Camera ($F_{2,23}$=2.91, p=0.08) is not significant.

After each task, the users filled out a NASA standard TLX questionnaire reporting on their task related impressions and

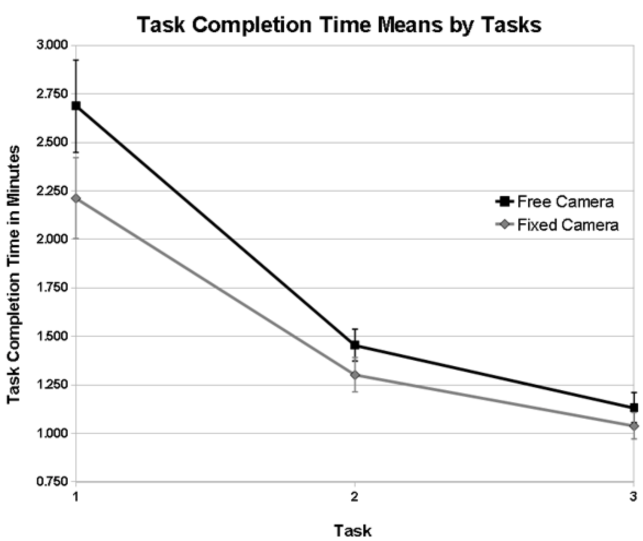

Figure 6. Task completion times (error bars +/- SE).

experiences on 21 point scale. A 2 (Camera) x 3 (Task) x 6 (TLX) repeated measures ANOVA with α=0.05 showed main effects for Task ($F_{2,28}$=7.99, $p<0.05$) and TLX ($F_{5,25}$=3.89, $p<0.05$) as well as an interesting interaction of Camera x TLX ($F_{5,25}$=4.47, $p<0.05$) (see Figure 7).

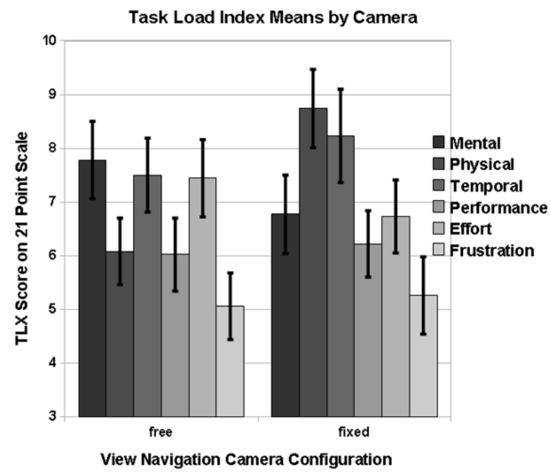

Figure 7. TLX experiences by camera (error bars +/- SE).

Closer analysis of Camera x TLX showed that the mental demand ($F_{1,29}$=4.09, p=0.05) lies on the borderline of significance, suggesting that the free camera viewport configuration (*M*=7.78, *SE*=0.72) has a higher mental demand on the user than the fixed camera viewport configuration (*M*=6.77, *SE*=0.73).

Another interesting effect of physical demand ($F_{1,29}$=15.97, $p<0.05$) on the user proved to be higher for the fixed configuration (*M*=8.74, *SE*=0.73) than for the free configuration (*M*=6.08, *SE*=0.62). The potentially interesting interaction Task x TLX did not prove to show any significant relations (Figure 8). All post-hoc comparisons included Bonferroni adjustments.

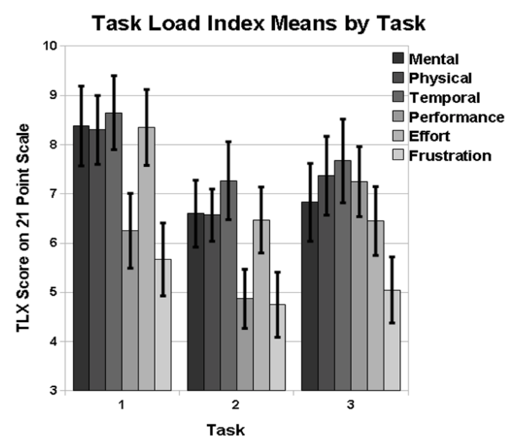

Figure 8. TLX experiences by task (error bars +/- SE).

4.2 User Questionnaire

The questionnaire was filled out after all the tasks had been completed and therefore summarizes the individual insights on the experiment. The answers were reported on a 7 point psychometric Likert scale (1=disagree and 7=agree).

The questions about the tracking and system performance were stated to get an impression on how the responsiveness of the application was perceived.

Q1: Do you think the tracking quality was good?
Q2: Do you think the tracking quality should be improved?

Questions three to six were asked in order to verify some of our previously stated hypotheses.

Q3: When just browsing (T1), do you prefer working with the attached camera?
Q4: When adding and moving content (T2), do you prefer working with the free camera?
Q5: When constructing (T3) do you prefer working with the free camera?
Q6: When interacting in general, do you prefer working with the free camera?

We intended to get an impression how general system parameters such as performance and screen size were perceived:

Q7: Do you think the system performance is sufficient?
Q8: Do you think the screen size is sufficient?

The last two questions were also open questions intended to give the opportunity to formulate wishes and alternative design choices for future hardware interfaces.

Q9: Would you like to have different input devices?
Q10: Would you like to have different output devices?

The results of Q1-Q10 were each analyzed using a two tailed t-test α=0.05 and are summarized in Table 1.

Table 1. T-test results of the questions.

	Mean	SD	t(29)	p (2-tailed)
Q1	4.60	1.38	2.38	<.025
Q2	5.40	1.48	5.19	<.025
Q3	4.57	2.21	1.41	.170
Q4	3.07	2.07	-2.47	<.025
Q5	3.50	2.15	-1.28	.212
Q6	3.47	1.96	-1.49	.147
Q7	5.47	1.01	7.97	<.025
Q8	6.10	1.19	9.71	<.025
Q9	3.00	2.15	-2.55	<.025
Q10	3.10	2.20	-2.24	.033

4.3 Interviews and Video Observations

The information gained from the interviews and the observation of the subjects is concentrated in this section.

Almost 80% of the subjects reported that they were annoyed by the cables on camera and display, which restricted their movement to some extent. Emphasis was especially put on the camera cable limiting the free movement of the observing camera when adjusting the viewport. A wireless camera may be more suitable.

The lost tracking when rapidly moving the camera or directing it to towards mainly untextured space was another undesirable issue reported by subjects. It was obvious in the observation that all subjects had to adapt their view navigation behavior to some extent in order to get a continuous and smoothly displayed MR view into the scene.

In task T2, the positioning of trees, a more fluent way for activating the moving tool in order to work more efficiently was alluded 21 times. Also a bug of disappearing objects was reported.

Some users with low expertise reported handling the free camera in one hand and using the pen in the other makes their view unstable, because their hand is not completely still. The resulting jitter was found annoying and sometimes even resulted in unwontedly offsetting the MR view. These subjects argued that the simultaneous coordination of both hands is mentally demanding, but they still liked this interface configuration and adapted fast. In contrast, users with more expertise instantly liked this navigation method and found it intuitive.

The observation of the subjects also revealed that for the searching task, it was easier for them to navigate around the occluded objects in the scene when using the free camera in their hand since it allows easier movement at low (near horizontal) angles and in between buildings. This observation was also backed up by several statements of subjects addressing this issue.

Especially for the searching task, subjects favored holding the display in their hands with the camera attached to it. They described this configuration as easy and intuitive to use in this particular interaction situation. In this context, it was suggested to mount a strap to the display so the weight is released from the hand holding it when interacting for a longer period of time. Another proposal was to optimize the display size and weight by removing the border around actual screen.

Most professional subjects from the field of urban planning enquired about having some sort of top projection onto the table giving feedback from the MR scene. They also suggested an additional wall projection of the tablet view, so this setup can be better used for collaborative work.

5. DISSCUSSION

We will look at the results of this study which was designed to answer three specific research questions and review our hypotheses which state our assumptions. Similar to [2], we think that the subjective preference data is in some ways more valuable than quantitative data. First we will summarize and discuss the mainly quantitative data followed by the examination of the qualitative data.

The first half of H1 is answered by the empirical result of the task completion time analysis, which showed that the fixed camera configuration was faster. Moreover the task load index analysis suggests that this configuration implies a lower mental load. Although this result is on the borderline of significance, we think that the hypothesis H1 is supported because of the strong verbal feedback of the subjects.

The hypotheses H2-H4, concerning the preferred condition for each of the tasks, are directly addressed by questions Q3-Q5. However, only Q4 had a significant result, expressing a slight tendency for the fixed camera configuration. Therefore, to our surprise there is not clear answer to research question R1, which configuration do subjects prefer. This is also evident from the lack of an overall preference in Q6.

In terms of the physical demand, the data is clear and proves with statistical significance that the free camera configuration is less physically demanding. This supports H5, and was also stated by some of the subjects during the evaluation.

Analogue to H1, we can answer R2 stating that the fixed view navigation leads to faster task completion times than the free view navigation. The question R3 is answered by looking at the TLX analyses, indicating a low mental load and a high physical load for the fixed viewport navigation configuration, which is obvious if one considers the extra weight of the display tablet. Exactly the opposite – high mental, but low physical load – occurs for the free camera navigation configuration.

Similar to the result of the study in [2] on bimanually operated desktop 3D graphics interfaces, using the non-dominant hand for camera control was received well by the users and seems to be intuitive in both camera configurations. The advantage of the free camera is the low weight and the higher flexibility for spatial movements needed for typical egocentric perspectives of the model, realized by navigating on street level. In general, the free camera configuration has initially a higher mental load and restricts the interaction space due to the length of the arms of the user working with a stationary display. The strength of the fixed camera setting is the low mental load and fact that the attached display is always at a convenient distance to the user even when working with large models. On the downside, the weight of the display and the spatial flexibility are not optimal.

Users had positive impression of the tracking quality (Q1), but still thought that it should be improved (Q2) for optimal operation in interactive settings. The overall system performance was found quite sufficient (Q7). In summary, the responsiveness of the application was perceived positive with a frame rate always well above 30fps. Screen size of the mobile display was found adequate (Q8), and the optional free form comment asking for the desired size was almost never filled in, so we conclude that the provided size of 12.1 inch is a good choice. The question concerning the need for alternative input devices (Q9) did not prompt many demands, although some non-professional users suggested finger touch input on the mobile display and directly on the table. Professionals liked the current state with the pen. Asking about different output devices (Q10) did not provide a clear answer. But many comments about future interface designs were received, suggesting hybrid display configurations using the mobile display in combination with projections. The suggestion for a wall projection of the scene is technically easy to realize and was already used in a previous experimental configuration, but considered out of scope for this paper. The enquiry for projected feedback of information onto the table was also realized in previous work, but will be technically more challenging in combination with the natural feature tracking, which is sensitive to texture and strong lighting changes.

Ishi *et al.* [17] found that a hybrid TUI/GUI approach can avoid clutter with tangible objects on a table. Using the proposed handheld interface, a tangible map table setup or a 3D model with low density could benefit from a 2.5D user interface in close proximity to the tangible augmented table in a collaborative working situation.

6. CONCLUSION

All the user feedback concerning the setup was positive, confirming that experiencing and expressing is done naturally and with enjoyment when using our bimanual MR interface. Independent of the users' expertise, all tasks were solved after a brief introduction and intentionally without any additional training. Input using the bimanual interface combined with real-time visual feedback seems to be easily learned. We conclude that overall the user interface supports efficient navigation and manipulation in 3D, which was necessary to complete the tasks in either of the two configurations.

In general the factors influencing the experience are numerous and cannot all be quantified in a single statistical model. That's why we are in favor of the insights gained by triangulating methods and the qualitative user feedback containing rich information on the system in general. The analysis of the collected data answered most of our research questions in the discussion and clarified some of our assumptions.

When working with users of varying professional backgrounds and skill levels, giving options for individually optimizing the user interface in order to address a wide range of needs sounds intriguing. However, when an interaction artifact such as our handheld MR device is frequently passed from user to user, reconfiguration is cumbersome. For example, the handheld MR device allows removing and re-attaching the camera quickly, but for user groups working on real problems, it is still not really feasible. In previous work we found that workflow and natural communication are too much disrupted when the interface itself needs attention. However, when one device per user can be deployed, a certain amount of startup customization (such as taking on or off the camera based on personal preferences) may be acceptable.

If the interface configuration cannot be deferred to the users, the designer must pick the right type of interface. This can depend on external factors such as the level of detail, the elevation and size of the physical models, or the number and agility of the involved users.

Our findings indicate an advantage of the interface with the camera attached to the display in terms of task completion time and mental load. However, users did not express a clear preference for either interface.

For getting more statistically significant answers we are convinced that simple questions need to be asked in the context of an even more limited experimental setting in order to reduce noise. This can be cumbersome when aiming at settings for real-world applications usually involving a high amount of influential factors. Finding efficient methods to address this problem, so a fruitful development of natural interaction techniques is guaranteed, is a challenge.

Another field of application for the suggested interface configurations might be the bimanual 3D object inspection already mentioned in the related work section. We can imagine handheld MR interfaces for applications such as product presentation or 3D industrial design.

The experience we gained will go into the design of future interfaces since the goal to give easy access to a wide range of expertise without neglecting anyone is still a challenge. Future work will focus on further improving the tracking quality to strengthen the natural character of the interface relieving the user from having to adapt her behavior to fit the interface. Better robustness to lighting changes would also be nice to have. Finally, a future interface design should aim at achieving wireless input and output devices with reduced weight for this scenario.

7. ACKNOWLEDGMENTS

The authors would like to thank VRVis Research Center, the Studierstube team and the other members of the IPCity project (EU Grant FP-2004-IST-4-27571), in particular the participants of the user study, Manuela Waldner and Daniel Wagner.

8. REFERENCES

[1] R. Balakrishnan, T. Baudel, G. Kurtenbach, G. Fitzmaurice, The Rockin'Mouse: Integral 3D manipulation on a plane. Proc. of CHI'97, 311-318, ACM, 1997

[2] R. Balakrishnan, G. Kurtenbach, Exploring bimanual camera control and object manipulation in 3D graphics interfaces. In Proc. of CHI '99. ACM, New York, NY, 56-62, 1999.

[3] M. Billinghurst, I. Poupyrev, H. Kato, R. May, Mixing realities in shared space: An augmented reality interface for collaborative computing, in Proc. IEEE Int. Conf. Multimedia and Expo., New York, Jul. 2000.

[4] M. Billinghurst, H. Kato, I. Poupyrev, Tangible augmented reality, In ACM SIGGRAPH ASIA 2008 Courses, SIGGRAPH Asia '08, ACM, New York, NY, 1-10, 2008.

[5] D. Bowman, D. Koller, L. Hodges, Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control

Techniques. Proceedings of the Virtual Reality Annual International Symposium, 45-52, 1997

[6] D. A. Bowman, D. Koller, L. F. Hodges, A methodology for the evaluation of travel techniques for immersive virtual environments, Virtual Reality: Journal of the Virtual Reality Society, 3, 120-131, 1998.

[7] N. Burtnyk, A. Khan, G. Fitzmaurice, R. Balakrishnan, G. Kurtenbach, Stylecam: interactive stylized 3d navigation using integrated spatial & temporal controls. In UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology, ACM, New York, NY, USA, 101–110, 2002.

[8] N. Burtnyk, A. Khan, G. Fitzmaurice, G. Kurtenbach, ShowMotion: camera motion based 3D design review, In Proc. of the 2006 Symposium on interactive 3D Graphics and Games, I3D '06, ACM, New York, NY, 167-174, 2006.

[9] H. Eden, E. Scharff, E. Hornecker, Multilevel Design and Role Play: Experiences in Assessing Support for Neighborhood Participation in Design. Proc. of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '02, ACM Press, New York, NY, 387-392, 2002.

[10] S. Feiner, A. Webster, T. Krueger, B. MacIntyre, E. Keller, Architectural anatomy, In Presence, 4(3), Summer, 318-325, 1995.

[11] G. Fitzmaurice, J. Matejka, I. Mordatch, A. Khan, G. Kurtenbach, Safe 3D navigation. In Proc. of the 2008 Symposium on interactive 3D Graphics and Games, I3D '08, ACM, New York, NY, 7-15, 2008.

[12] M. Gleicher, A. Witkin, Through-the-lens camera control. In Proc. of ACM SIGGRAPH '92, New York, NY, 331-340, 1992.

[13] M. Hachet, F. Decle, S. Knodel, P. Guitton, Navidget for easy 3d camera positioning from 2d inputs, In 3DUI '08: Proc. of the 2008 IEEE Symposium on 3D User Interfaces, IEEE Computer Society, Washington, DC, USA, 83–89, 2008.

[14] E. Haik, T. Barker, J. Sapsford, S. Trainis, Investigation into effective navigation in desktop virtual interfaces. In Proceedings of the Seventh international Conference on 3D Web Technology (Tempe, Arizona, USA, February 24 - 28, 2002). Web3D '02. ACM, New York, NY, 59-66, 2002.

[15] K. Hinckley, R. Pausch, J. C. Goble, N. F. Kassell, A survey of design issues in spatial input. In Proc. of UIST '94, Ed. ACM Press, New York, NY, 213–222, 1994.

[16] A. J. Hanson, E. A. Wernert, Constrained 3D navigation with 2D controllers. In Proc. of the 8th Conference on Visualization '9, IEEE Computer Society Press, Los Alamitos, CA, 175-ff, 1997.

[17] H. Ishii, J. Underkoffler, D. Chak, B. Piper, E. Ben-Joseph, L. Yeung, Z. Kanji, Augmented Urban Planning Workbench: Overlaying Drawings, Physical Models and Digital Simulation, In Proc. of ISMAR 02, 203- 211, 2002.

[18] P. Isokoski, R. Raisamo, B. Martin, G. Evreinov, User performance with trackball-mice, Interact. Comput. 19, 3 (May. 2007), 407-427, 2007

[19] H. Kato, M. Billinghurst, I. Poupyrev, K. Inamoto, K. Tachibana, Virtual object manipulation on a table-top ar environment, in Proc. Int. Symp. Augmented Reality, Munich, Germany, Oct. 2000.

[20] A. Khan, B. Komalo, J. Stam, G. Fitzmaurice, G. Kurtenbach, HoverCam: interactive 3d navigation for proximal object inspection. In I3D '05: Proceedings of the 2005 symposium on Interactive 3D graphics and games, ACM, New York, NY, USA, 73–80, 2005.

[21] J. Lee, G. Hirota, A. State, Modeling real objects using video see-through augmented reality, Presence - Teleoperators and Virtual Environments 11(2), 144–157, 2001.

[22] J. D. Mackinlay, S. K. Card, G. G. Robertson, Rapid controlled movement through a virtual 3D workspace. In Proc. SIGGRAPH '90, volume 24, 171-176, 1990.

[23] V. Maquil, M. Sareika, D. Schmalstieg, I. Wagner, MR Tent: A Place for Co-Constructing Mixed Realities in Urban Planning, In Proc. of Graphics Interface 2009, ACM, 211-214, 2009.

[24] J. McCrae, I. Mordatch, M. Glueck, A. Khan,. Multiscale 3D navigation. In Proc. of the 2009 Symposium on interactive 3D Graphics and Games, Ed. I3D '09. ACM, New York, NY, 7-14, 2009.

[25] M. McKenna, Interactive viewpoint control and three-dimensional operations. In Proc. of SI3D '92, ACM, New York, NY, 53-56, 1992.

[26] T. B. Moeslund, M. Störring, W. Broll, F. Aish, Y. Liu, E. Granum, The ARTHUR System: An Augmented Round Table, Journal of Virtual Reality and Broadcasting, 34, 2003.

[27] U. Neumann, S. You, J. Hu, B. Jiang, I. O. Sebe, Visualizing reality in an augmented virtual environment, Presence: Teleoper. Virtual Environ, 13, 2, 222-233, MIT Press, 2004.

[28] B. Piper, C. Ratti, H. Ishii, Illuminating Clay: a 3D Tangible Interface for Landscape Analysis, In Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 355-362, 2002.

[29] M. Sareika, D. Schmalstieg, Urban Sketcher: Mixed Reality on Site for Urban Planning and Architecture, In Proc. of 6th International Symposium on Mixed and Augmented Reality, Nara, Japan, November 2007 pp. 27-30., IEEE, ACM, 2007

[30] M. Sareika, D. Schmalstieg, Urban Sketcher: Mixing Realities in the Urban Planning and Design Process, In Proc. of the 26th Annual CHI Conference Workshop on Urban Mixed Realities: Technologies, Theories and Frontiers, ACM, 2008.

[31] R. Stoakley, M. J. Conway, R. Pausch, Virtual reality on a WIM: interactive worlds in miniature. In Proc. of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press/Addison-Wesley Publishing Co., New York, NY, 265-272, 1995.

[32] J. Underkoffler, H. Ishii, Urp: A Luminous-Tangible Workbench for Urban Planning and Design, In Proc. of CHI, 386-393, 1999.

[33] Ina Wagner, Wolfgang Broll, Giulio Jacucci, Kari Kuutii, Rod McCall, Ann Morrison, Dieter Schmalstieg, Jean-Jacques Terrin, On the Role of Presence in Mixed Reality. PRESENCE - Teleoperators and Virtual Environments, Vol. 18, No. 4, 249-276, MIT Press, 2009.

[34] C. Ware, S. Osborne, Exploration and virtual camera control in virtual three dimensional environments. In SI3D '90: Proc. of the 1990 Symposium on Interactive 3D graphics, ACM, New York, NY, USA, 175–183, 1990.