

Techniques for View Transition in Multi-Camera Outdoor Environments

Eduardo Veas¹

Alessandro Mulloni¹

Ernst Kruijff¹

Holger Regenbrecht²

Dieter Schmalstieg¹

¹ Institute for Computer Graphics and Vision, Graz University of Technology, Austria

² Department of Information Science, University of Otago, New Zealand

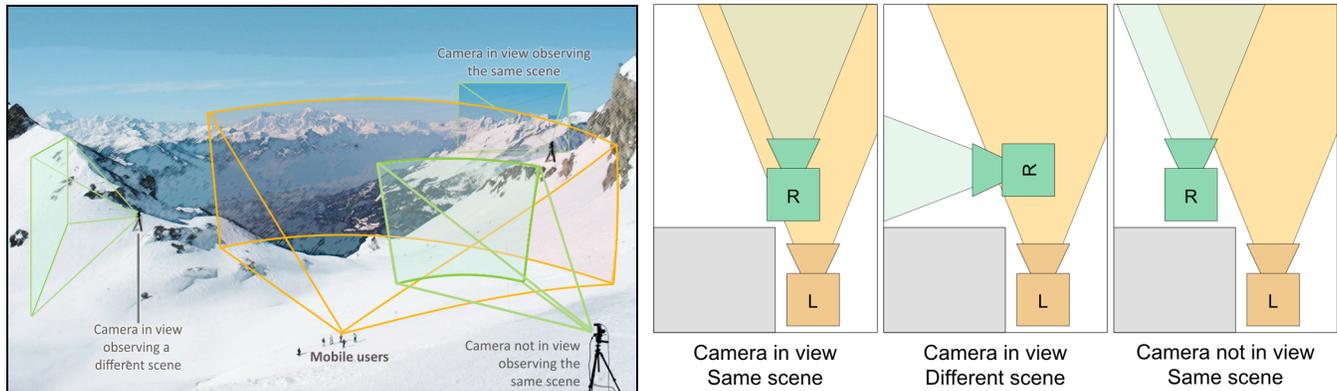


Figure 1. A complex outdoor configuration of cameras, illustrating the different relationships between remote cameras and mobile users. For a local camera (L) and a remote camera (R), we classify camera configurations based on the remote camera's location from the viewer's point of view, and on the scene the remote camera is looking at.

ABSTRACT

Environment monitoring using multiple observation cameras is increasingly popular. Different techniques exist to visualize the incoming video streams, but only few evaluations are available to find the best suitable one for a given task and context. This article compares three techniques for browsing video feeds from cameras that are located around the user in an unstructured manner. The techniques allow mobile users to gain extra information about the surroundings, the objects and the actors in the environment by observing a site from different perspectives. The techniques relate local and remote cameras topologically, via a tunnel, or via bird's eye viewpoint. Their common goal is to enhance spatial awareness of the viewer, without relying on a model or previous knowledge of the environment. We introduce several factors of spatial awareness inherent to multi-camera systems, and present a comparative evaluation of the proposed techniques with respect to spatial understanding and workload.

KEYWORDS: Navigation techniques, multi-camera systems, human factors, situation awareness, workload.

INDEX TERMS: H.5.1 [Information interfaces and presentation] Multimedia Information Systems - Artificial, augmented, and virtual realities; H.5.2 [Information interfaces and presentation]: User Interfaces - Ergonomics, Evaluation/methodology

1 INTRODUCTION

Mobile computers and wireless video transmission can enable a user in an outdoor environment to observe video feeds from multiple cameras deployed in the user's surroundings. Remote cameras can help users in tasks such as obtaining an overview of the situation, planning for movement or searching for anomalies. Our motivation scenario stems from the HYDROSYS¹ project, which focuses on environmental monitoring using mobile devices. In this project, teams of users monitor environmental changes on-site using mobile augmented reality (AR). A prototypical scenario for such project includes static cameras (deployed at particular points of interest), semi-mobile cameras (attached to pan-tilt units), user-attached cameras, and a drone, totaling up to 10 devices. The arrangement of cameras within the environment and in relation to each other, as well as the user's position within that environment, is variable. Cameras are placed in "wild" areas such as mountains, riverbanks, ridges, and are accessed by mobile devices through a sensor network. Users often might have only little knowledge about the surroundings.

This outdoor scenario differs significantly from a conventional surveillance application, a common application of multi-camera setups. Surveillance setups include large numbers of cameras deployed statically, and rely on user's knowledge of the location. Users stay indoors and are only confronted with remote camera information without the ability to physically observe and navigate the environment. In contrast, in an outdoor scenario information can be gathered directly from the environment, from the visualization on the mobile device or from a combination of both.

In the process of dealing with the camera feeds, an observer must deal with a view discrepancy: The presentation of a remote camera view on a mobile device is separated from what the observer sees with her own eyes and perceives with the rest of her sensory system. The sensory qualities of the remote camera feeds are necessarily reduced compared to direct perception, and users

¹ e-mail: veas | mulloni | kruijff | schmalstieg @icg.tugraz.at

² e-mail: holger@infoscience.otago.ac.nz

¹ HYDROSYS project: <http://www.hydrosysonline.eu/>

may lack information on how to get to the remote location or what lies in between camera positions. The available information is mostly egocentric; though the user's persistent representation – the mental map – is exocentric. The process of building mental maps is affected by the pre-knowledge of the site, the spatial ability of the user and the features of an environment that can be matched from the cameras' multiple perspectives.

Navigation techniques are designed to assist the process of building mental maps, by substituting missing sensory input in different forms. Previous work in this area concentrates on indoor surveillance and does not consider the characteristics of agile outdoor users with small mobile computers.

The main contribution of our work is a human-factors study of three multi-camera navigation techniques suitable for this kind of outdoor scenario. The three techniques are an image *mosaic* relating local and remote images topologically, a *tunnel* animation from local to remote and a *transitional* animation from local to remote via an intermediate bird's eye viewpoint. All techniques are simplistic in the sense that they require only minimal information to inform a user where the remote viewpoint is located with respect to the user. Thus, they assume no prior knowledge of the environment and can be deployed in an ad-hoc manner, for example by multiple roaming users with cameras. We conducted a comparative evaluation of the three techniques to assess how well they perform in a real outdoor scenario.

We first review related work in section 2 and discuss relevant human factors in section 3. The three techniques are detailed in section 4. The evaluation focuses on the performance of users at deriving spatial information with each of the technique, as described in section 5, followed by a discussion in section 6. Finally, an outlook and future work are presented in section 7.

2 RELATED WORK

Understanding multiple simultaneous camera feeds requires consideration of *situation awareness* and *mental workload*. The term situation awareness was coined in aviation, and has found its application in user interfaces [8][9]. The process of acquiring and maintaining situational awareness and its relation with mental workload are illustrated by [18]. Measuring of and designing for situation awareness, including the related mental workload, generally relates to single-view real environments. We adopt these methodologies for our multi-view, mixed-reality environment.

Our research is affected by techniques deployed in navigation systems, such as 3D virtual environments (VEs) and surveillance systems. These techniques focus on providing additional information to support the spatial awareness of the user, through methods that allow for egocentric or exocentric referencing. Navigation techniques for VEs in particular have been an active field of work. Bowman et al. [3] introduce a methodology for evaluating navigation techniques for VEs. Elmqvist et al. [7] consider occlusions and present a taxonomy for the acquisition of spatial information, such as object discovery and spatial relations. Bowman et al. [4] provide a good introduction to the topic of navigation in 3D virtual worlds, discussing the major processes that are affected in the user's cognition.

Multi-camera systems are often used in surveillance installations and much research has been carried out to assist the processes in the control room. Latest work in the area has proposed integrating VEs with multi-camera systems [16][15][21] to leverage existing VE navigation techniques for transitioning between camera views. In security systems, cameras are normally placed in a dense grid, resulting in partially overlapping view frustums. De Haan et al. [5] exploit this feature, creating a grid of transition paths between cameras and blending camera views with a 3D model of the infrastructure whenever there is a gap to keep

the transition smooth. DOTS [10] organizes the video feeds on the screen in a layout corresponding to the topology of the cameras. In combination with AR, multi-camera systems have been utilized to implement pseudo x-ray vision, among others to embed video imaging from occluded remote cameras [19][1].

While an overlap with previously developed techniques exists, our intended usage of multi-camera systems is quite different from traditional setups in either surveillance or VE-based systems. Novel issues include ways of perceptually combining information obtained from directly observed environmental cues and remote camera feeds, and the highly unstructured and dynamic camera setup. Instead of observing an environment from a control room or a VE, users are situated in the real environment. This can be both an advantage (improved perception) and a disadvantage (for example, loss of attention in demanding situations). We will further explain these issues in the following sections.

3 HUMAN FACTORS IN MULTI-CAMERA ENVIRONMENTS

The two main issues affecting the performance of tasks in a multi-camera setup are situation awareness and mental workload. In the past the term mental workload partially replaced situation awareness, but for our purposes they will be regarded as two separate constructs: Situation awareness as a cognitive construct, and mental workload as its "energetic" counterpart, mostly referring to the effort a user needs to invest [18]. This section aims to explain what affects the design of a user interface that imposes low workload but allows for high situation awareness. We also discuss the types of view discrepancy that can adversely affect situation awareness.

3.1 Situation awareness

We assume that deploying multiple cameras will likely provide a better overview to assess specific situations. This introduces the concept of *situation awareness*, a dynamic construct that results from a cognitive process entailing perception of cues in the environment, comprehension of the current situation and projection of future status [9]. Situation awareness encompasses a person's tasks and forms a basis for decision-making. Spatial awareness is a part of situation awareness that deals with the understanding of space. Spatial awareness includes a person's knowledge of self-location within the environment, of surrounding objects, of spatial relationships among objects and between objects and self, as well as the anticipation of the future spatial status of the environment.

Navigation involves gathering and applying spatial knowledge. While navigating, multi-sensory input is processed and stored in a mental map that represents spatial knowledge [4]. A person experiences the world from an egocentric perspective, where all perceptual input is relative to the personal frame of reference. However, spatial knowledge is presumably stored in exocentric form, representing elements relative to each other and to a spatial reference frame [6]. Each of these relations can introduce distortions to spatial information [20].

3.2 Mental workload

Reasoning about inter-object relationships causes *mental workload*, an energetic construct that refers to the supply and demand of attentional and processing resources [18]. Mental workload is affected by both the structure and the situation of a task (exogenous factors), and the abilities of a person (endogenous factors) such as spatial ability and experience with a system. A task can be characterized by its difficulty, priority and the related situational contingencies. Users may or may not have prior knowledge of the environment and the cameras located therein – this knowledge forms the connection to situation awareness.

Apart from the task's characteristics, both attention and processing of spatial information play an important role. Attention is a critical factor affecting the usage of handheld devices: Users get easily distracted by occurrences in the environment and may be further challenged by adverse conditions such as reflections on the screen or strong sunlight.

Attention is a scarce resource in the acquisition of situation awareness. The concentration required for switching between screen and environment likely affects the success of a navigation technique. When viewing a remote camera's feed, the user has to make a mental transformation of her own location to the remote camera to establish a mental relationship. Such transformations are prone to errors, and may be further challenged by the size and possible bad legibility of the screen. Research has indicated that the nature of the transformation affects the underlying neural implementation [22] and affects workload and accuracy [17].

Minimizing complex mental transformations is a key requirement for achieving low mental workload and sustaining attention. This can involve conveying the (unknown) location of the camera and indicating how to arrive from the current point to the remote point. Moreover, conveying camera orientations helps in disambiguating common objects.

3.3 View discrepancy

Regular navigation techniques are based on the premise that they should provide enough sensory input (even though artificial) to give a sense of (or replace) the sensation of moving from one place to the other. This premise does not hold for handheld users and introduces problems such as view discrepancy.

While being in the field rather than in the office, the view shown on the display device is inconsistent with what the observer perceives directly. Perception of the immediate physical world is not completely discontinued while concentrating on the remote view. Thus, a perceptual conflict is introduced: users can observe multiple information sources that do not necessarily match in content or fidelity. Moreover, the premise of navigation in a VE does not hold: Instead of traversing between positions, the user remains at one position, but looks at the video feed from another.

The extent of the discrepancy is affected by the relationship between user and remote camera. We can classify the relationship between the camera held by the user and a single remote camera by considering the remote camera as seen from the user's point of view. The classification takes into account whether the viewpoint of the remote camera is visible or not, and whether the camera is observing the same or a different scene (see Figure 1):

camera in view – same scene (CS): the viewpoint of the remote camera is visible and is (partially) observing the same scene as the user. Mental transformations required to connect viewpoints can be derived by inspection.

camera in view – different scene (CnS): the viewpoint of the remote camera is visible, but due to occlusions in the environment, it is observing a scene that the user cannot see.

camera not in view – same scene (nCS): the viewpoint of the remote camera is not in view, but it is (partially) observing the same scene as the user. This is the case when the camera is occluded: If the user requires only rotating in order to see the remote camera, the camera is considered in view. To derive the relation between viewpoints the observer must match common objects in both views and perform mental transformations about those objects.

In a fourth category, **camera not in view – different scene**, the viewpoint of the remote camera is not in view and it is not observing the same scene as the user. In the present work we do not consider this case because the relation between viewpoints

cannot be derived without further knowledge. However, previous knowledge of the observer can aid in the process; the user may know the site and could thus derive the location of the camera.

4 VIEWPOINT TRANSITIONS IN MULTI-CAMERA ENVIRONMENTS

When browsing video feeds from local and remote cameras, different approaches can be taken to transition between a local and a remote view. Techniques can be either egocentric or exocentric, and may take an uninformed or an informed approach. The latter property refers to data about the environment provided to the user, such as maps or 3D models. Informed techniques require the user to interpret the extra data to make sense of the environment. Within this paper we deliberately use uninformed techniques.

The goals of our experiment are to establish whether uninformed techniques can help acquire correct spatial information and to evaluate the mental workload needed to obtain such information. Uninformed techniques cannot convey information about the spatial configuration of the environment: We assume that additional spatial knowledge is retrieved from observation of both the camera feeds and the environment itself.

The three techniques represent main research directions on navigation and multi-camera systems, adapted to mobile AR. They convey similar information in different ways, which allows for a proper comparison of workload and user acceptance. The Mosaic technique represents a typical surveillance system solution [10], displaying camera feeds in tiles organized topologically with respect to one another. The Tunnel technique is adapted from a recent technique used in AR to guide the user to an object [2]. The transitional technique is based on an interface that allows users to transition between contexts [11].

All of the techniques can be invoked through a single button. Each of them initially shows the local camera's video. A button press (Figure 3(A)) takes the user to a *transition view* presenting the spatial relationship between local and remote camera. A second button press transitions to the remote view. Pressing again changes back to *transition view* and further to the local view. The *transition view* and transition actions form the core of each technique (Figure 2).

A complete multi-camera system would require not only tools for transitioning to a remote camera, but also for browsing and selecting the desired remote camera. Since the focus of this work is on techniques for transitioning, our experiment scenario can be limited to two cameras. The techniques would need to be adapted (mosaic, transitional) or re-thought (tunnel) in order to support also tasks of browsing and selection.

4.1 Mosaic technique

The Mosaic technique allows users to transition between local and remote views using a mosaic of video thumbnails. The thumbnails resemble the visualization used in the control room of surveillance systems. We organize the thumbnails based on topology, as in DOTS [10]. The technique uses the angle between the viewing direction of the local camera and the position of the remote camera to position on the screen minimized versions of both videos. This conveys to mobile users how they should turn in order to "see" the camera, like a compass. As a user moves, the visualization updates accordingly. The mosaic technique does not show the 3D spatial relation and the distance between cameras. It is primarily a 2D technique, providing a directional cue towards the remote camera with respect to the user. The thumbnails show both videos, allowing users to get a minimized view of both cameras at the same time. Since the organization of the thumbnails does not depend on distance, this technique allows the visualization of several cameras simultaneously, as long as the cameras are not in the same direction.

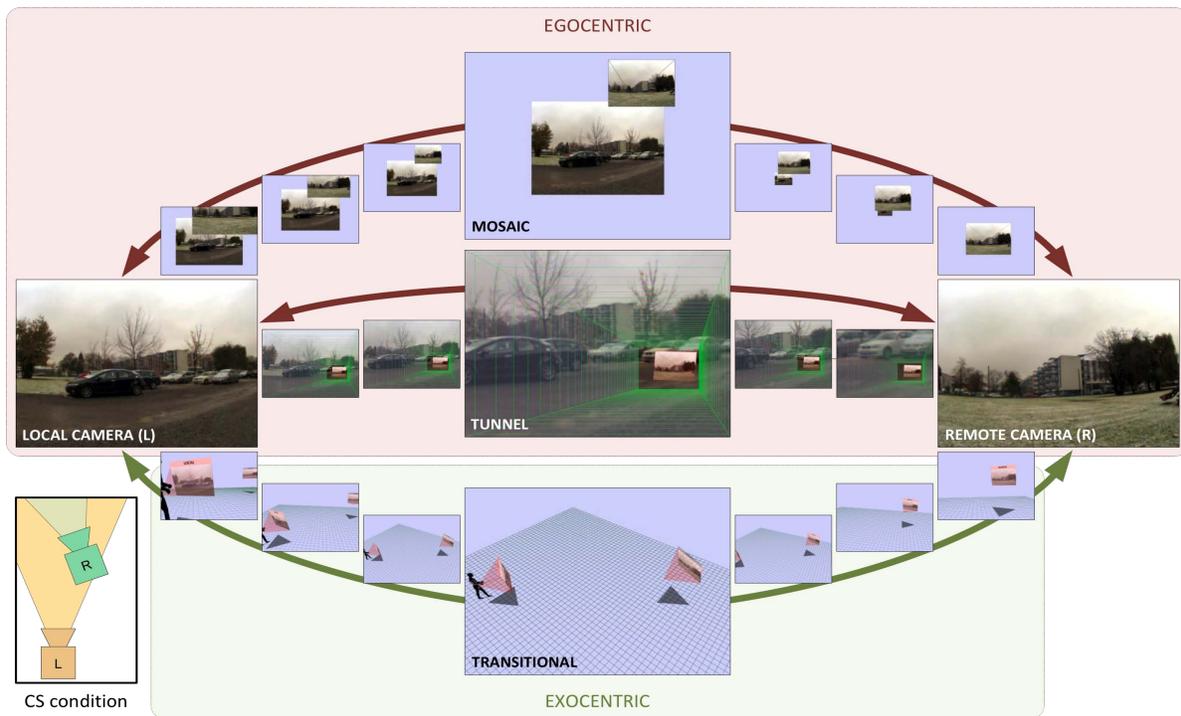


Figure 2. An example of the three proposed techniques applied to a CS condition. Using the techniques, users can browse the video stream from either the local or the remote camera, or they can smoothly move to a view where both videos are visible.

4.2 Tunnel technique

The Tunnel technique is a variation of the attention funnel, first introduced by Biocca et al. [2]. This is an egocentric technique to guide a user to an object of interest. The technique displays a tunnel oriented to the remote camera. Users can travel down the tunnel to the other camera. We blend the tunnel over the video background so that tunnel and video are both visible. When the remote camera is in view, the user can see its video feed at the end of the tunnel. If the remote camera is not in view, the tunnel indicates the turning direction.

Traveling down the tunnel brings the view to the other camera. The technique is expected to work best when a user first rotates until the remote camera is in view and then travels to it. In this case, the technique conveys a complete spatial relation, including rotation, translation and view direction. The distance to the camera is correctly shown in 3D from the perspective of the user.

4.3 Transitional technique

The Transitional technique implements the concept of transitional interface in the sense of [11]. In a transitional interface, users can transition between contexts, each possibly having a different space, scale and representation. In our case, users can move between an egocentric AR context, where a full-screen augmented video is visible, and an exocentric VE context, where users get a bird's eye overview on both cameras and their respective spatial position and orientation. In the exocentric view, an avatar is used to disambiguate the user's camera from the remote camera. We employ smooth animations to support coherent transitions.

5 EXPERIMENT

We conducted a user study to compare the techniques asking users to infer relationships between a local camera and a remote camera. For each type of camera configuration (see Section 3.3) we focused on the impact of the techniques on the users' spatial awareness and mental workload. The test setup used a handheld

device consisting of an ultra-mobile computer (Panasonic CF-U1 with a 5.6" screen), a Ublox GPS sensor and a uEye UI-2210 color camera (640x480 resolution) mounted with a 4.2MM Pentax wide-angle lens. The uEye camera was physically bound to the whole setup, and acted as the local camera in the user study. All three techniques were used in combination with pre-recorded video feeds from static cameras. During the experiment, a tripod was positioned in the field to represent the remote camera (see Figure 3).

The aim of the comparative test was to answer the following questions, for each type of camera configuration:

- Q1. Were there differences in spatial knowledge obtained from the different techniques?
- Q2. Which technique has less impact on the user's workload?
- Q3. What is the user's preference: Do users tend towards one technique when they are asked for subjective impressions on various parameters?

We conducted the study as a single experiment and employed a 3x3 factorial design. We treated the 3 camera configurations (CS, CnS, nCS) as a between-subject independent variable: Users were divided in three groups, and each group experienced a different type of camera configuration. The transition technique (mosaic, tunnel, transitional) was treated as a within-subject independent variable. Hence, every participant was assigned the same camera type over three locations and varied the technique per location (all camera configurations were available at every location).

We used a Latin square distribution to balance the order in which the techniques and the locations were assigned to each user. We also enforced that users' familiarity with the locations and gender were balanced between the camera types. Before the experiment, we collected demographic data, some information on the amount of time users spend with both paper and digital maps, and information on their spatial abilities, for which we used the SBSOD questionnaire [13].



Figure 3. Pictures of the experiment. From left to right: (A) the mobile setup used for the experiment, as a user presses the button to trigger a transition, (B) during the experiment, we always positioned one tripod as a landmark for the location of the remote camera, (C) one participant searches the remote camera in the environment, (D) throughout the experiment participants were asked to draw maps of the locations.

The experiment started with an outdoor introductory session, where users could get familiar with the handheld device and the techniques. A dummy remote camera was provided for practicing. After the introduction, we blindfolded and walked the users to three different locations at our University campus. The three locations had varying levels of features, including different types of buildings and varying density of trees. Upon arrival, users were provided again with one of the techniques and they were asked to identify the position and orientation of the remote camera in the environment, by making use of both the device and the physical environment. To prevent user biasing, the techniques were named with neutral names (A, B and C). Users were allowed to look around but not to move from the designated location.

Once the users felt confident about drawing a map representing the main objects in the scene, the camera locations and their orientation, we gave them paper and pencil and we took the device from them. Users could therefore not use the technique while drawing. Users also filled in a short questionnaire to further investigate on their spatial awareness and their workload, using parts of the NASA TLX questionnaire [12] and an RSME (Rating Scale for Mental Effort, scale 0-150, 150 for maximum effort [23]). Finally, we collected the level of user confidence at each location. After the three techniques were used, we asked the subjects to state their preferred technique for a variety of factors related to spatial awareness and workload. In total, participants had to fill out 12 pages of questions (31 spatial ability questions and 41 ratings based on Likert scale) and draw three maps.

27 users (16 male, 11 female, aged between 22 and 48) participated in the study. All (but one) participants had normal or corrected to normal vision. To partially compensate for the effects of prior knowledge of the environment, we invited 17 users that regularly visited the campus and 10 users that had hardly visited the campus before. Of the 17 regularly visiting users, 9 users did hardly know at least one of the locations. Results were collected from the 27 users x 3 locations, totaling 81 trials. The duration of the experiment was about 1.15h per participant.

5.1 Results on situation awareness

As a result of the SBSOD questionnaire analysis, we retrieved a median score of 65.56%. Based on the median and differentiation of the results, we separated the users in three groups:

G1: below average (< 55%), 2 male / 7 female

G2: average (55–75%), 9 male / 3 female

G3: above average (> 75%), 5 male / 1 female

Female users were more prevalent in the G1 and G2 spatial ability groups, which is in line with other studies [14] (Pearson

correlation, $p < .01$). It is important to note that, though we applied a Latin square distribution of users and we enforced balance of gender and familiarity with the environment, we cannot make exact statements on user group effects per camera type: no high spatial ability users fell within the nCS condition.

We started with interpreting the maps drawn by the users. Due to the diversity and quality of drawing, we only made rough estimations on errors: we used a voting mechanism among researchers to check for errors in the overall spatial configuration (VS), and separately in the position (VP) and orientation (VO) of the remote camera. The results of this analysis provide very interesting indications.

Table 1. Total number and types of errors in the map drawings, for each camera type and each technique.

	Mosaic			Tunnel			Transitional		
	VS	VP	VO	VS	VP	VO	VS	VP	VO
CS	0	1	1	3	2	3	2	1	3
CnS	2	1	1	3	1	2	3	1	0
nCS	1	3	0	2	3	3	1	1	1
TOTAL	10			22			13		

Mosaic caused the least errors in the drawings (Table 1). Of particular interest is that users made surprisingly few errors when drawing the remote camera position, even if the technique itself does not provide any distance information. Transitional performed very well in the nCS condition, especially when one considers that no high spatial ability users fell within this condition. The technique seems to provide quite accurate information on the remote camera's placement and orientation when the remote camera is not visible by the user. High-ability participants made significantly fewer errors (Pearson correlation: $p < .01$). Previous knowledge of the environment only had a significant main effect on errors produced by the Tunnel technique, but not on the other techniques (one-way ANOVA, $F_{1,25} = 9.04$, $p < .01$): Users with previous knowledge performed better with Tunnel than users with no previous knowledge.

Table 2. Averages and standard deviation (in italic) for self-assessed success ratings (for each spatial ability group and for gender). 7 point Likert scale, lower scores are better.

	G1	G2	G3	Male	Female	AVG
Mosaic	3.78 <i>1.09</i>	2.67 <i>1.07</i>	2.17 <i>0.75</i>	2.50 <i>1.10</i>	3.55 <i>1.04</i>	2.93 <i>1.17</i>
Tunnel	3.11 <i>1.27</i>	2.58 <i>1.51</i>	2.50 <i>1.05</i>	2.31 <i>1.20</i>	3.36 <i>1.29</i>	2.74 <i>1.32</i>
Trans	3.44 <i>1.24</i>	2.33 <i>1.37</i>	1.67 <i>0.52</i>	1.88 <i>0.96</i>	3.55 <i>1.21</i>	2.56 <i>1.34</i>

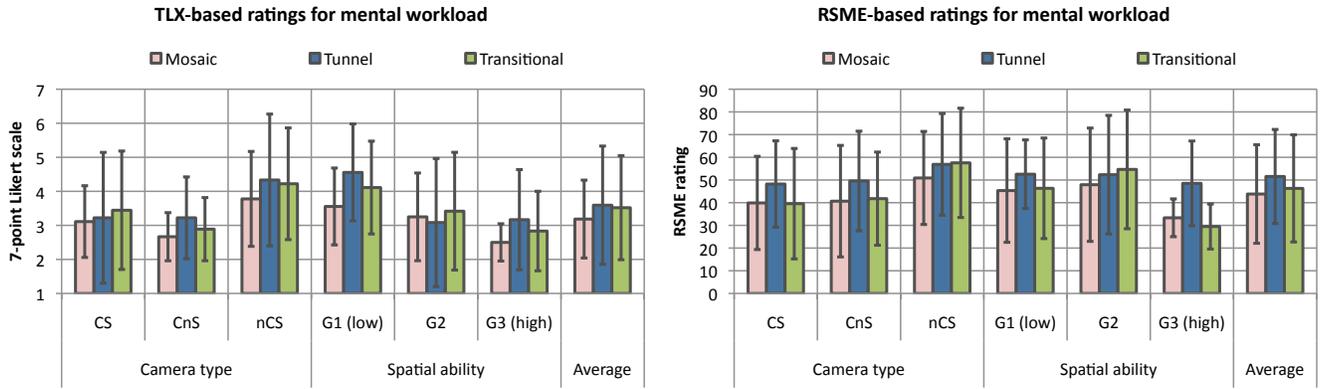


Figure 4. Average and standard deviation of subjective ratings of mental load (7 point Likert scale, lower is better) and RSME (including standard deviation in italic) for each camera type and each spatial ability group, and in average.

For each technique, users assessed their success after drawing the map (Table 2). Users with higher spatial ability felt more confident. The previously stated correlation between spatial ability and errors in drawing the maps supports the personal assessment. Regarding the techniques, groups G2 and G3 felt most confident with the transitional technique, whereas group G1 preferred the tunnel. Hence, it is important to note that, though Mosaic caused the least errors, users did not report the highest confidence in this technique for drawing the map.

We noticed an interaction effect between spatial ability groups and the success rating ($F_{2,24}=5.17, p = .01$). A post-hoc test showed a main effect of spatial ability on success ratings for Mosaic ($p = .02$) and Transitional ($p = .03$): Higher-ability users felt more confident using these two techniques than users with lower ability. Spatial ability didn't show a significant impact on the self-assessed success for Tunnel. One-way ANOVA also showed that camera types did not have a significant effect on the subjective success rating. The only noticeable result is that nCS is rated slightly higher, thus showing a tendency that users were less confident in that condition. There was no significant effect between the techniques.

In general, users estimated that with either technique they needed to retrieve as much information from the screen as from the environment itself. This partially confirms our expectations: Since the techniques provide only limited information, users would have to observe the environment to fulfill their task. We expected that the Mosaic technique would force the users to observe the environment more, whereas the Tunnel technique would require more attention to the screen. However, a repeated measures ANOVA provided no significant difference among techniques, suggesting that for all of them users had to pay as much attention to the screen as to the environment. There was also no significant effect of camera type on the focusing of attention on either screen or environment.

Q1. Were there differences in spatial knowledge obtained from the different techniques?

Mosaic tends to perform better, producing the least errors but not causing the highest confidence among users. Transitional tends to give higher confidence (for drawing maps), and seems to perform better when the remote camera is hidden from the user. Users with higher spatial ability apparently prefer either Mosaic or Transitional to Tunnel. Users in the lower spatial ability group had a slight preference for Tunnel, although it seemed to produce more errors.

5.2 Results on mental workload

To analyze workload, we considered both the workload-related questions (derived from TLX) and the RSME scale. We found a direct correlation between mental workload and RSME per-technique ratings (Pearson correlation, $p < .01$ for all techniques), thus forming a rather reliable base to judge the user's workload (Figure 4). There is a tendency of Mosaic to require less workload and of Tunnel to require more, but a one-way ANOVA did not show any significant effect between technique and mental workload, nor between spatial ability and technique.

Additionally, no significant effect could be found between group, technique and camera conditions after multivariate analysis. Finally, a one-way ANOVA did not show any effects on the order and progress of the test on the mental workload.

Q2. Which technique has less impact on the user's workload?

There was a tendency of the Mosaic technique to require less workload. Although not significant, Transitional received a better rating than Tunnel. No significant effects of the camera types on the workload could be found.

5.3 Technique preference

Analyzing the general preferences of the techniques (3-point Likert scale), most users liked Mosaic best, followed by Transitional. A repeated measures ANOVA showed an effect on user preference: a post-hoc t-test showed that Tunnel was significantly less preferred than Mosaic ($p < .01$) and Transitional ($p = .03$). We noticed no significant effect of spatial ability on the technique preference. However, if we look into the details of the ratings, several differences can be noticed (Table 3).

Table 3. Average preference ratings (3 point Likert scale, higher is better) and standard deviations (in italic).

	<i>Ease of navigation</i>	<i>Ease of use</i>	<i>Drawing map</i>	<i>Effort</i>	<i>Helpfulness</i>	<i>Attention</i>	<i>Confidence</i>	<i>Like</i>
Mosaic	2.67 <i>0.62</i>	2.70 <i>0.67</i>	2.70 <i>0.54</i>	2.33 <i>0.83</i>	2.89 <i>0.32</i>	2.37 <i>0.84</i>	2.74 <i>0.53</i>	2.70 <i>0.67</i>
Tunnel	2.37 <i>0.79</i>	2.41 <i>0.64</i>	2.63 <i>0.63</i>	2.37 <i>0.69</i>	2.59 <i>0.57</i>	2.19 <i>0.79</i>	2.26 <i>0.76</i>	2.11 <i>0.75</i>
Trans	2.74 <i>0.53</i>	2.67 <i>0.55</i>	2.70 <i>0.47</i>	2.44 <i>0.70</i>	2.85 <i>0.46</i>	2.48 <i>0.70</i>	2.63 <i>0.56</i>	2.56 <i>0.64</i>

Users in the G2, G3 groups liked the Tunnel technique less, consistently for all camera conditions. Mosaic was liked most in all camera conditions, whereas Transitional is not considerably disliked, especially in the CnS and nCS conditions.

There are no significant differences between the ratings of the techniques for attention, effort, general navigation preferences, as well as the usage of the techniques for drawing a map. We noticed that Mosaic required more attention than the other techniques in the CS condition, performing worse than the Transitional method. Tunnel performed worst in all conditions. Subjective effort ratings support the findings from the workload section.

Users gave high ratings to the usefulness of all techniques in helping them to draw a map. A repeated measures ANOVA showed a significance in the confidence rating ($p = .02$): a post-hoc t-test showed that users were significantly more confident in Mosaic than Tunnel ($p = .03$), but not significantly more confident in Mosaic than in Transitional. Spatial ability did not have a significant effect on the preference ratings. In general, lower spatial ability resulted in slightly lower preference rates, and worse attention rates. Also, effort was slightly increasing with decreasing spatial ability. It is interesting that increased effort did not significantly reduce the preference for a technique.

Q3. What is the user preference: Do users tend towards one technique when they are asked for subjective impressions on various parameters?

There is an overall preference for the Mosaic technique. Though it does not always perform significantly better than the Transitional technique, for most camera types and spatial ability groups it seemed to receive higher preference ratings.

6 DISCUSSION

The way in which the relationship between user viewpoint and remote camera is communicated varies widely between techniques. Tunnel and Mosaic work in an egocentric mode – Tunnel uses AR methods to overlay all information on the video stream, whereas Mosaic shows only the relative location of the remote camera, hence just giving a directional cue. Transitional, on the other hand, works in exocentric mode, and reveals the full spatial relationship, including distance. We found that with all techniques users had to gather information from the surroundings and from the video feeds equally, to infer the spatial relationship between cameras and the environment. To our surprise, the fact that some techniques provide more information does not seem to have a strong impact on the behavior to observe the environment.

The preference evaluation shows that all users found the techniques equally helpful in drawing the maps. It is interesting to contrast this result with the self-assessment on spatial awareness, where users of group G3 thought they were most successful with any technique but the tunnel, while users in G1 believed they were mostly successful with the tunnel. In contrast, the preference evaluation shows that most users considered the Mosaic and Transitional techniques most helpful, while the Tunnel was rated lower. Similarly, users deemed Mosaic and Transitional easier to use and easier to navigate with compared to Tunnel. An apparent result is that users felt more confident performing with Mosaic, just a bit less with Transitional, but significantly less confident with Tunnel. This actually contrasts with the confidence users reported after drawing the map: the success rate for Mosaic was lowest in that case. This is quite surprising and might infer either irregularity during the ratings of the users, or a difference between how users interpret confidence and success. It should be noted, at this point, that during the experiment we observed that the users found it difficult to assess the success rating.

An unexpected result from the preference survey is that users found that they needed as much effort for any of the techniques,

while attention scores were similar, with tunnel scoring a bit lower. However, when comparing these results with the ones obtained for workload, the results show a difference especially in favor of Mosaic, and partially of Transitional. This may indicate that users prefer a technique that requires less workload and provides more confidence.

Table 4. Overview of best-performing techniques (highest averages) on main factors in situation awareness, workload and user preference. Mosaic (M), Tunnel (Tu) and Transitional (Tr).

	CS	CnS	nCS	G1	G2	G3	AVG
<i>SA errors</i>	M	M Tr	Tr	M	Tr	M	M
<i>SA success</i>	Tu Tr	Tr Tu	Tr	Tu	Tr	Tr	Tr
<i>Mental load</i>	M	M	M	M	Tu	M	M
<i>Like</i>	M	M	M	M	M	M Tr	M
<i>Attention</i>	Tr	M Tr	M	Tr	M	Tr	Tr
<i>Confidence</i>	M	Tr	M	M	M Tr	M	M

Regarding camera types, we can conclude that all techniques help users to infer where the cameras are, and produce similar results even for the nCS condition. However, what surprised us is that camera types mostly did not have a significant effect on the results. We only noticed a main effect when analyzing the quality of maps in the nCS condition using the transitional technique: even though no high spatial ability users were in this condition, it performed remarkably well (Table 4). Hence, for sites with “hidden” cameras, the transitional technique might be the option to prefer instead of the mosaic technique.

We could not show any significant effects of the varying difficulty of the chosen locations. This does not imply that the techniques work well with more difficult scenarios. True outdoor scenarios, such as the mountainous environments considered in HYDROSYS generally present fewer cues that can be used for matching remote views with one’s current view. Results may vary in such challenging scenarios.

It is instructive to compare the results on workload with those on spatial awareness. When observing the self-assessment on spatial awareness, we noted that users in groups G2 and G3 felt they were most successful with Transitional, followed by Mosaic and then Tunnel, although there is no significant difference. In comparison, the number of errors was lowest using Mosaic followed by Transitional and Tunnel. This implies that users are not fully aware of their overall performance. Compared to workload, we can infer that users prefer a technique that lets them perform reasonably well while imposing lower workload.

While observing the users, we noticed minor technical difficulties that may have an effect on the practical usage in real-world scenarios. The techniques vary in their resilience to accuracy and registration error (error in the alignment between virtual and real objects): Mosaic can cover better for errors than the other two techniques. We noticed that some users were disappointed when registration errors occurred (the remote camera appeared to be off in the transition view) with either technique, and were forced to observe the environment more closely, even though this is not directly confirmed by the ratings. Time allocated to screen and environment was about equal throughout all techniques. In the comments and discussion section after the experiments, users did not raise the issue. However, some users complained about the view distortion introduced by the use of wide-angle lenses, and the sometimes limited legibility of the screen. Illumination and weather conditions were roughly the same for all users, and should therefore not have had a significant influence on our findings.

7 CONCLUSION

In this paper, we presented human factors influencing the perception and usage of multi-camera systems on mobile devices. We discussed the cognitive background relevant for the specific problem domain, which significantly differs from traditional surveillance applications, and we provided a camera classification for the different types of camera setups that can be deployed. Based on the issues identified in the problem space, we designed, developed and evaluated three camera transition techniques: Mosaic, Tunnel and Transitional. These techniques are designed to help users in gathering knowledge on the spatial configuration of the site being observed, as well as on the utility of the remote camera. We particularly focused on the location and orientation aspects of remote cameras, as well as on the relationship between local and remote cameras and the observed scenery.

The evaluation showed that navigation techniques act as a kind of equalizer between camera configurations, bringing them to a similar level of difficulty. It also hinted that users prefer a technique that imposes lower workload, if it allows them to perform reasonably well. The evaluation showed that Mosaic was preferred by most users, and in general tends to work best when analyzing the spatial awareness and workload of the users (Table 4). A dependency on the user's direct view was found – users gathered information from both sources to make decisions quite equally with all techniques. However, we could not determine which knowledge was actually gathered from which source: a separate investigation is left for future work.

In the future we want to extend the evaluation to informed techniques. This will also involve including maps, 3D models and other sources of information. For the users, the challenge lies in the interpretation of the extra information, particularly when its sources are not accurate or outdated (old maps or 3D models without new buildings). Once the cognitive impact of different techniques has been identified, a study will evaluate a more complex scenario including multiple static and mobile cameras. A technical challenge for future work will be providing more accurate registration for advanced navigation techniques.

ACKNOWLEDGEMENTS

We would like to thank all users participating in the experiment. This work was partially supported through the EU funded projects HYDROSYS (EU-FP7-224416), IPCity (EU-FP6-IST-27571) and MARCUS (PIRSSES-GA-2008-230831), through the Christian Doppler Laboratory for Handheld Augmented Reality and through the Austrian Science Fund FWF (W1209).

REFERENCES

- [1] B. Avery. X-Ray Vision for Mobile Outdoor Augmented Reality. Dissertation, School of Computer and Information Science, University of South Australia, 2009.
- [2] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention funnel: omnidirectional 3D cursor for mobile augmented reality platforms. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22-27, 2006). CHI 2006. ACM.
- [3] D. A. Bowman, D. Koller, and L. F. Hodges. A methodology for the evaluation of travel techniques for immersive virtual environments. *Virtual Reality: Research, Development, and Applications*, 3(2):120–131, 1998.
- [4] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.
- [5] G. de Haan, J. Scheuer, R. de Vries, and F. H. Post. Egocentric navigation for video surveillance in 3d virtual environments. In *3D User Interfaces*, pages 111–118, Los Alamitos, CA, USA, 2009.
- [6] R. D. Easton and M. J. Sholl. Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2):483–500, 1995.
- [7] N. Elmqvist and P. Tsigas. A taxonomy of 3d occlusion management for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14:1095–1109, 2008.
- [8] M. R. Endsley. Situation awareness global assessment technique (sagat). In *National Aerospace and Electronic Conference (NAECON)*, Dayton, OH, USA, 1988.
- [9] M. R. Endsley. Toward a theory of situation awareness in dynamic systems: Situation awareness. *Human factors*, 37(1):32–64, 1995.
- [10] A. Girgensohn, D. Kimber, J. Vaughan, T. Yang, F. Shipman, T. Turner, E. Rieffel, L. Wilcox, F. Chen, and T. Dunnigan. Dots: support for effective video surveillance. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 423–432, New York, NY, USA, 2007. ACM.
- [11] R. Grasset, J. Looser, and M. Billingham. Transitional interface: concept, issues and framework. In *ISMAR '06: Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 231–232, Washington, DC, USA, 2006. IEEE Computer Society.
- [12] S. Hart and L. Stavenland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Elsevier, 1988, ch. 7, pp. 139-183.
- [13] M. Hegarty, A. Richardson, D. Montello, K. Lovelace, & I. Subbiah. (2002). Development of a Self-Report Measure of Environmental Spatial Ability. *Intelligence*, 30, 425-447.
- [14] Lawton, C.A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles* 30(11/12), 765-779.
- [15] H. S. Sawhney, A. Arpa, R. Kumar, S. Samarasekera, M. Aggarwal, S. Hsu, D. Nister, and K. Hanna. Video flashlights: real time rendering of multiple videos for immersive model visualization. In *Proceedings of the 13th Eurographics Workshop on Rendering (Pisa, Italy, June 26 - 28, 2002)*, pages 157-168, vol. 28. Eurographics Association, Aire-la-Ville, Switzerland, 2002. ACM.
- [16] I. O. Sebe, J. Hu, S. You, and U. Neumann. 3d video surveillance with augmented virtual environments. In *IWVS '03: First ACM SIGMM international workshop on Video surveillance*, pages 107–112, New York, NY, USA, 2003. ACM.
- [17] R. N. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 171(972):701–703, February 1971.
- [18] P. S. Tsang and M. A. Vidulich. Mental workload and situation awareness, chapter *Handbook of human factors and ergonomics*. John Wiley & Sons, 2006. ed. G. Salvendy.
- [19] T. Tsuda, H. Yamamoto, Y. Kameda, and Y. Ohta. Visualization methods for outdoor see-through vision. In *ICAT '05: Proceedings of the 2005 international conference on Augmented tele-existence*, pages 62–69, New York, NY, USA, 2005. ACM.
- [20] B. Tversky. Cognitive mapping: past, present, and future, chapter *Levels and Structure of Spatial Knowledge*. Routledge Frontiers of Cognitive Science No.4. Routledge, 2000. ed. R. Kitchin and S. Freundschuh.
- [21] Y. Wang, D. M. Krum, E. M. Coelho, D. A. Bowman, Contextualized Videos: Combining Videos with Environment Models to Support Situational Understanding. In *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1568-1575, November/December, 2007.
- [22] J. M. Zacks, J. Mires, B. Tversky, and E. Hazeltine. Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation*, 2(4):315–332, 2001.
- [23] Zijlstra, F.R.H., (1993). *Efficiency in Work Behaviour; A Design Approach for Modern Tools*. Delft: Delft University Press.