Multiple Camera Augmented Viewport: An Investigation of Camera Position, Visualizations, and the Effects of Sensor Errors and Head Movement

Thuong N. Hoang* and Bruce H. Thomas†

Wearable Computer Lab – University of South Australia

ABSTRACT

This paper presents an extension to our augmented viewport technique for action at a distance for outdoor AR systems by employing the use of different physical camera positions. The original technique augments the user's view with video images from a physical zoom lens camera to provide advantageous viewing windows into the augmented environment, through which the user can perform image plane manipulation of virtual objects. Our extended augmented viewport technique utilizes a range of camera positions, including remotely located cameras, head mounted zoom lens cameras, and tripod mounted zoom lens cameras, to offer several benefits: closer views of the scene of interest, novel and complementary viewing angles with multiple viewports, stability against sensor errors, and view-dependent interaction to enhance precision. We introduce new visualizations to assist in the discovery of the cameras. We conducted a user study to evaluate the effects of different camera viewpoints, sensor error, head movement, and the multiple viewports visualization on the usability of the augmented viewport.

KEYWORDS: Augmented viewport, interaction technique, image plane, outdoor augmented reality.

INDEX TERMS: H.5.2 [Information interfaces and Presentation]: Graphical User interfaces - Interaction styles; I.3.6 [Computer Graphics]: Methodology and Techniques - Interaction Techniques

1 INTRODUCTION

This paper presents our continued work on the augmented viewport technique [1] for action at a distance (AAAD) with outdoor augmented reality (AR) systems. AAAD is the problem of interacting with virtual objects that are located out of arm's reach. In our original augmented viewport technique, we demonstrated a set of techniques that augment the user's view with video images from a physical zoom lens camera to provide advantageous viewing windows into the augmented environment, through which the user can perform image plane manipulation of virtual objects located at a distance. The main benefit of augmented viewports is their support for precise interaction with virtual objects at a distance in an AR environment.

The results from our previous investigation posed a number of interesting questions:

1. The augmented viewport can utilize a range of different physical camera locations in the environment, so what effect do different camera viewpoints have on the usability of the technique?
2. Considering that the types of physical cameras include sensor tracked cameras and head mounted cameras, what effect do sensor error and user's head movement have on the usability of the technique?
3. When there are a number of physical cameras in the environment, how does the technique support the user in the discovery and utilization of physical cameras?

To answer these questions, we investigated three types of camera location and their effects on precise manipulation, in terms of head movement and sensor error in the zoom lens camera, and oblique viewing angle using the remote camera. We developed virtual visualizations to assist the user in discovering and selecting suitable cameras for the desired manipulation tasks, based on the location and the viewing area of each camera. We conducted a user study to evaluate various error effects and the multiple viewport visualization (see Figure 1) and present the results with post-study discussions.

Figure 1. Multiple viewport visualization

1.1 Augmented Viewport Technique

The augmented viewport technique [1] enhances two common AAAD techniques for outdoor AR systems, the image plane [2] and AR working plane [3] techniques, which use the projection of the augmented environment as seen through a user's head mounted camera. Our technique leverages other cameras in the environment that can provide closer views of the distant location. There are two main types of cameras that can offer such an advantage: remotely located cameras and cameras with an optical zoom lens. In this paper, we investigate the use of remotely located cameras, and two variants of zoom lens cameras, namely head mounted and tripod mounted, for the augmented viewport technique. Remote cameras are mounted in a fixed remote location and orientation, while zoom lens cameras are located near the user and have adjustable orientation and position.

The augmented viewport shows a virtual window showing the video feed of a physical camera. The viewport window is overlaid with the view from a virtual counterpart of the physical camera, with the same intrinsic parameters, orientation, and location as the physical one. The combination of the video image and the overlay produces a windowed view into the AR environment through which the user can interact with virtual objects, using close body

*email: thuong.hoang@unisa.edu.au
†email: bruce.thomas@unisa.edu.au

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

interaction techniques. Figure 2 shows an augmented viewport with a blue virtual object overlaid on the physical background of a brick wall.

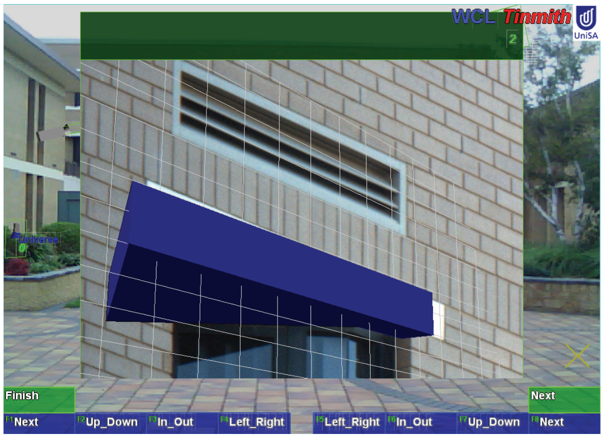

Figure 2. An augmented viewport

We previously investigated [1] three placements of the augmented viewport relative to the user's viewpoint based on a single tripod camera mounted next to the user. The placements of the viewports are defined as three different relative coordinate systems. *World relative* places the viewport at a fixed location in the world coordinate system (GPS), allowing the user to view the window from various angles by physically walking around the viewport. *Body relative* fixes the viewport in the coordinate system that takes the user's body as the origin, so that the viewport is always located at a fixed distance and orientation from the body; while *head relative* attaches the viewport to the user's head position and orientation, for a fixed and direct view of the viewport window. The three placements range in the flexibility of the viewing angle, with the *head relative* at the fixed viewing angle end, and the *world relative* at the flexible end of the scale. Our previous investigation found no difference in performance for each of the viewport placements.

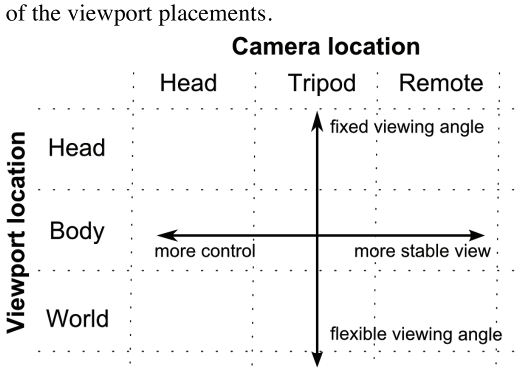

Figure 3. Variants of the augmented viewport

To better understand the relationship between the combinations of the three viewport placements and the three types of cameras locations, we place these concepts on a chart shown in Figure 3. The chart depicts nine combinations (3x3) between augmented viewport placement and physical camera position. The horizontal axis represents the camera location, and reflects the position of the camera relative to the user's view and the range of user control over the physical camera. A head mounted camera offers flexible control of the camera viewpoint using head movements. A sensor tracked tripod produces more stable viewpoints but requires a slower adjustment process. The remote camera offers the most stable view that is not affected by sensor errors, but its position is fixed and least flexible. As previously mentioned, our previous work [1] explored the effects of the viewing angle based on the viewport location (the vertical axis).

1.2 Contributions and Structure

This paper makes a number of contributions to AAAD manipulation techniques for outdoor AR:

1) A new set of the augmented viewport techniques and visualizations for the discovery and utilization of a range of physical cameras use for precise action at a distance manipulation tasks.
2) The results of a user study on the effects of different camera viewpoints, head movement and sensor errors on zoom lens cameras, and the multiple viewport visualization on the usability of the augmented viewport technique.

The paper starts with a description of the related research to AAAD and AR. Our extensions to the augmented viewpoint technique are then presented in detail. A description of the user study performed is given, followed by a discussion of the results. The paper finishes with a set of concluding remarks.

2 Background

The augmented viewport technique is based on image plane technique [2] for virtual immersive systems. The image plane technique [2] collapses the virtual world along the depth dimension onto a planar surface, and simplifies the interaction to two dimensions. This approach poses an inherent limitation of the inability to perform interaction along the depth/normal axis of the current viewing image plane. The AR working plane [3] extends the image plane approach to enable action at a distance interaction in AR environments. Instead of defaulting to the user's viewing plane, the AR working plane technique supports the creation of virtual planes for the input cursor and virtual objects to be projected onto. Both techniques are based on the image plane from the first person perspective using the user's head mounted camera. Our augmented viewport technique extends the image plane approach to use the viewpoints of other physical cameras in the environment. Augmented viewports are based on virtual environment viewports. SEAM [4] is a method of employing virtual viewports to intertwine multiple virtual environments in concert. The viewports are attached at various locations in the group of virtual worlds, acting as viewing platforms between two distant locations. Through-the-lens techniques [5] implement similar metaphor for navigation and object manipulation in VR. Interaction from two separate distant locations could be enabled by using multiple viewports [6].

Precise interactions are a major issue faced by many immersive modeling systems. HoloSketch [7] is a virtual environment sketching tool that relies on a highly accurate tracking and display system to function properly. 3DARModeler [8] and ARpm [9] are two hybrid immersive modeling systems that use desktop-based CAD systems for precision inputs. There are many factors affecting precision in direct manipulations in immersive environments. One such factor is the user's inability to perform constant and precise movements with their arms and hands. PRISM [10] is a manipulation technique that addresses this issue by applying a scaled mapping ratio between hand movements and virtual object displacements, based on the speed at which the hand travels. Our augmented viewport technique reduces such effects over the distance by offering a closer view of the remote scene.

Sensor error is another factor impacting on precise manipulation. Holloway's error model [11] identifies the sensor as the main source of most registration errors in AR systems. Sensor fusion is a common approach that combines various types of

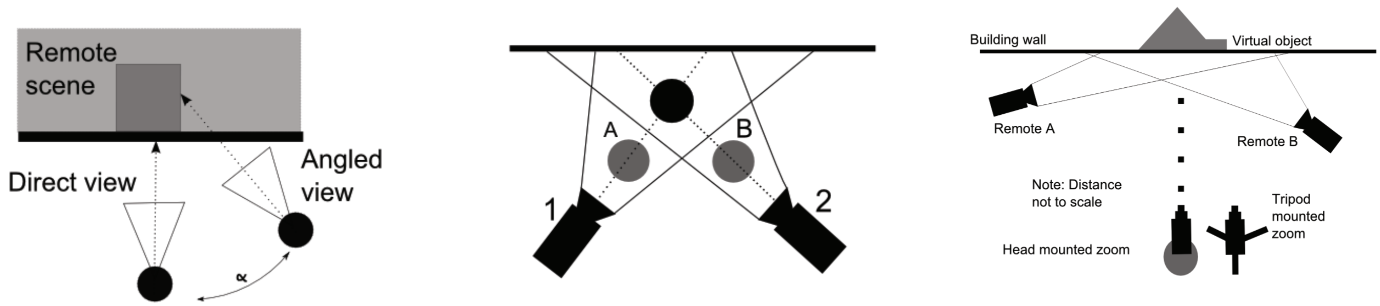

Figure 4. Left: View-dependent interaction Middle: Multiple viewport setup Right: User study setup

sensors to achieve better registration results. Effective combinations include inertial sensor and vision data [12, 13], with GPS tracker [14, 15], or fixed and head-mounted sensors [16]. The augmented viewport technique does not use sensor fusion for error correction; however, its usage with stationary remote cameras bypasses sensor errors for better precision.

3 Augmented viewports with different camera locations

In this section, we describe in detail the new extension to the augmented viewport technique to support multiple camera locations for the video image feed of the viewport, with regards to the benefits of the technique and the visualization used.

3.1 Benefits

The augmented viewport techniques with cameras in different physical locations have many potential benefits: novel viewpoints, the elimination of sensor errors, and view-dependent interaction.

3.1.1 Novel viewing angle

Depending on its physical location, remote cameras offer the advantage of a novel viewpoint at the scene of interest. Interaction techniques conventionally used for outdoor AR such as the image plane [2] and the AR working plane [3] are not effective along the normal axis of the head mounted display (HMD) camera. The augmented viewport enables remote image plane interaction using the imagery from the physical cameras, and has the same image plane limitation along the principal axis of the remote camera; however, this axis is not generally parallel to that of the user's head mounted display. Therefore, the use of the augmented viewport with remote cameras enables image plane interaction along the normal axis of the HMD by providing novel viewing angles of the scene of interest. The benefits of the novel viewpoints include: (1) a closer view to the scene, (2) a remote image plane interaction that is effective along the normal axis of the user's head mounted display to enable more precise manipulation along this depth axis, (3) a viewing angle that could not be obtained from the user's current location, due to physical constraints.

3.1.2 Stability against sensor noise

There are two main components to an augmented viewport: the physical and the virtual cameras. Physical remote cameras are often fixedly mounted and not tracked, thus an augmented viewport utilizing a remote camera is not affected by sensor errors and jittering.

The use of a remote stable camera for the viewport with *head or body relative* placement is also free from noise generated by the user's position and head orientation sensors (*head relative* only) in an outdoor AR system. A head relative augmented viewport using remote camera fixes a virtual window to the user's viewpoint, regardless of the current head orientation and location. *Body relative* placement offers similar stability against the location sensor noise, because the viewport window is attached to the user's body, regardless of their position. The alignment of the virtual and physical world inside the augmented viewport is also fixed, as mentioned earlier. Therefore, the user can perform manipulation tasks through the remote camera augmented viewport without being affected by the errors of the user's head orientation and location sensors.

The freeze-frame technique [17] is an example of eliminating sensor errors by capturing still snapshots of the environment, together with sensor data. The augmented viewport technique with a remote camera offers stability against sensor noise in a non-disruptive manner. The main video stream of the head mounted camera and all the sensors are kept running throughout the manipulation process. The augmented viewport is presented as a virtual window and does not cover the entire vision of the user. Stability of virtual objects inside the viewport is seamlessly achieved without the user's intervention. The freeze frame and similar techniques help reduce sensor jittering, but are still affected by sensor drifts. The accuracy of the virtual overlays at the time of freezing could be affected by accumulated drifts of the sensors running over time. The augmented viewport is not affected by such drifts, but only by the initial errors in the orientation and position measurements of the physical cameras.

3.1.3 View-dependent interaction

The implementation of the augmented viewport technique allows the user to achieve additional viewpoints of the virtual objects through the viewport, simply by adjusting the angle at which they interact with the viewport. The different placements of the viewport in head, body, and world relative coordinate systems enable adjustments of the viewport viewing angle.

The remote location is rendered in full 3D using OpenGL stencil buffers; thus, the user can gain extra viewing angles at the remote scene. When the viewport is placed in a head relative position, the user looks directly into the viewport (direct view in Figure 4 Left, where the shaded region represents the remote scene, and the dark box in the shaded region is a virtual object at the remote scene), gaining the view as if they were standing at the physical camera location. This viewport placement gives a constant direct view. In the body relative placement, the viewport window is fixed at a certain angle and distance from the body, enabling the user to look into the viewport constantly from an oblique α angle. The remote scene is then seen by the user as if the user was standing at a location that is rotated the same α angle about the remote scene from the physical camera location, as illustrated as the angled view in Figure 4 Left. This view allows the user to view different portions of the virtual object, such as the right side of the virtual box, as an example. The world relative placement supports both direct and angled views as it allows the user to walk around the augmented viewport.

This view-dependent viewport interaction allows the user to perform exploratory tasks to gain extra insights into the remote scene. For the best result, we suggest using image homography to generate the physical world view from camera images corresponding to the user's viewing angle of the viewport.

3.2 Visualizations

The augmented viewport can be used with many types of physical cameras available in the environment. Depending on the task requirements, one camera may offer more favorable viewing angle than the others. Therefore, we have introduced visualizations to support the discovery of physical cameras for the augmented viewport technique, as well as the utilization of multiple viewports setup.

3.2.1 Camera discovery

Upon the user arriving at an outdoor setting, information about the available cameras in the surroundings is downloaded to the wearable computer. There are many possible scenarios for camera positions: the camera itself can be (a) visible, or (b) not visible to the user; and the physical area the camera is looking at is either (a) visible, or (b) not visible to the user. Even if the cameras and its viewing areas are visible through the normal vision of the head mounted display, it is not clear to the user as to what the cameras are looking at.

Therefore, for each physical camera, we render a virtual overlay to highlight its position, orientation, identification, and viewing area. The overlay consists of a virtual model of the camera placed directly over the physical camera, a virtual frustum extending from the camera's position to the viewing area, and an identification number uniquely assigned for each camera in the surroundings. This visualization can be viewed in two different modes, namely immersive and orbital view.

In immersive mode, the virtual cameras are rendered in the first person perspective. This mode is mostly effective when the area of interest for the task is known, because the user can immediately identify if there are any cameras pointing at the required area and if their viewing angles are suitable for the manipulation tasks.

Orbital view [18], on the other hand, is a pure virtual viewpoint that gives an overview of the environment from a higher vantage point, allowing the user to explore the broader surroundings to discover more cameras. With the purely virtual nature of the view, the user can freely navigate around the environment. Wireframe models of physical buildings and landscapes, if available, could be rendered as reference for the relative positions of the cameras. If such models are not available, the orbital view could be taken from the viewpoint of a virtual camera that is fixed to the user's head orientation, but flew backwards and upwards to reach a higher perspective, so that the yaw orientations of the orbital and the immersive view are still aligned. Such alignment enables the user to switch between the two views without being disoriented about the locations of the physical cameras relative to the user.

3.2.2 Multiple viewports

The augmented viewport suffers from the same limitation as other image plane interaction techniques: ineffectiveness along the normal axis of the plane. We investigate and implement the usage of multiple augmented viewports to tackle this limitation.

Figure 4 Middle depicts an example scenario for the benefits of multiple augmented viewports. There are two cameras, 1 and 2, viewing the scene with a virtual sphere (dark circle) in its correct position, from different angles. The shaded circles, marked A and B, represent the possible erroneous positions of the virtual sphere that would potentially go undetected when using a single viewport only. At location A, the virtual object would appear as almost unchanged from the perspective of camera 1; in a similar manner that the object at location B would be mistaken as the correct position in camera 2. Both locations represent object displacement along the normal axes of the respective cameras. However, when both augmented viewports are visible, the user can detect such an anomaly and perform correction operations to put the virtual object in the correct position. The multiple viewports allow the users to build up a 3D model of the position of the virtual object, as single camera may not supply enough depth information for the user to understand the object's relative depth position.

4 User study

Our previous study [1] evaluated the concept of the augmented viewport and showed an improvement in precision, time, and effort in manipulation tasks. In this paper, we extend the augmented viewport to support a wider range of physical cameras, which introduces several factors potentially affecting the usability of the technique. We were motivated to conduct a user study to evaluate the performance of different camera positions and determine the effects of different viewpoints, head movement and sensor noise, as well as the multiple viewport visualization.

4.1 Design

In order to separately examine the above-mentioned factors, we designed multiple task conditions in which the participants used the augmented viewport to perform common manipulation tasks. The aim of the base task is precise manipulation, by scaling or moving virtual objects to match with the size or position of a physical artifact, located at a distance, called the *distant scene*.

We implemented the following four different camera placement augmented viewports (head relative view), each characterized by a single or compounded evaluation factors (see Figure 4 Right):

1. **Remote** camera: This condition uses a single remote camera (remote A) looking at the distant scene from an oblique angle, which is different from the first person perspective viewing angle of the distant scene. The single factor of a different camera viewpoint is embedded in this condition.
2. **Head** mounted zoom lens camera: This condition uses a single zoom lens camera, controlled by the participant's head orientation. The participant's location and head orientation are tracked; therefore, this condition is compounded with the head movement and sensor error factors.
3. **Tripod** mounted zoom lens sensor tracked camera: This condition uses a single zoom lens camera, mounted on a tripod next to the user. This tripod is tracked with orientation and location (GPS) sensors, similar to that on the participant's HMD. The participant controls the orientation of the tripod. This condition is affected by a single factor of sensor errors, of both orientation and location sensors.
4. **Multiple** remote camera: This condition uses two remote cameras: remote A, and the second camera is another remote camera (remote B), mounted on the opposite side about the participant, pointing at the distant scene at a different oblique angle. This condition is compounded with the multiple viewport and different camera viewpoint factors.

We then add an additional augmented viewport as a baseline comparison condition:

5. **Fixed** tripod camera viewports: This condition uses a single zoom lens camera, mounted on a tripod whose position and orientation are fixed by calibration and not tracked, at the *same position* as tripod camera with sensors. This condition is not affected by any of the evaluation factors above.

In the experimental design, we ensure that there are separate conditions that use one of the three different camera positions, namely head, tripod, and remote mounted cameras, in order to compare overall effects of different positions of the cameras.

Table 1. Error in moving tasks (in meters) and scaling task (in unit) for five camera conditions

	Moving task (in m)						**Scaling task (in unit)**					
	Depth		*Side*		*Up*		*Depth*		*Side*		*Up*	
Condition	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Remote	0.41486	0.03694	0.58346	0.09957	0.21667	0.01284	0.0802	0.0020	0.1426	0.0035	0.0241	0.0001
Head	1.49689	0.71231	2.04042	1.01558	0.81769	0.19663	0.2480	0.0070	0.2459	0.0137	0.0372	0.0005
Tripod	1.20025	0.42868	1.16354	0.43478	0.39892	0.02749	0.2296	0.0246	0.2121	0.0110	0.0348	0.0002
Multiple	0.41545	0.09051	0.37281	0.03606	0.17453	0.00444	0.1082	0.0027	0.1436	0.0125	0.0232	0.0001
Fixed	0.30255	0.03173	0.03087	0.00056	0.06639	0.00356	0.0460	0.0012	0.1286	0.0044	0.0191	0.0006

Performance was measured with three different quantitative methods: *task error, completion time, and the number of mouse clicks required*. Task error was measured as the difference in the virtual object's final position and size from the actual position and size of the matching physical object, physically measured during calibration as the ground truth. The time to complete the task in microseconds was determined by the participant upon satisfaction of the task's result. Our hypotheses are as follows:

H1: There is a measurable reduction in the performance of the augmented viewport when affected by the different camera positions.

H2: There is a measurable reduction in the performance of the augmented viewport when affected by head movement.

H3: There is a measurable reduction in the performance of the augmented viewport when affected by sensor errors.

H4: There is a measurable reduction in the performance of the augmented viewport when affected by multiple viewport visualization.

4.2 Experiment

We had 16 participants (15 males and 1 female), aged 18 – 44 (mean: 25.37, SD: 7.71). Nine participants had never used an AR system or a wearable computer before. The participants were asked to wear the Tinmith wearable computer system to scale and move a virtual window lentil to match the size and position of a physical window lentil, using the augmented viewport technique. There were five different camera conditions, as described in the previous section and illustrated in Figure 4 Right (condition 3 and 5 were co-located at the Tripod mounted zoom location). In total there were ten tasks (2x5) to perform for each iteration. Each participant completed two iterations with randomized task orders, after one training session. There were breaks in between iterations.

The remote cameras were mounted with identical lenses with a focal length of 25 mm, while the head mounted camera and the tripod mounted camera used identical 75 mm fixed focal length lenses. Traditional variable focal length zoom lens was not used to reduce calibration errors. All cameras were set to capture at 640x480 resolution. Both remote A and B cameras were mounted on a fixed tripod, while the zoom lens camera located near the user was on an adjustable tripod. The user and the tripod were approximately 50 meters away from the building, while the remote cameras were mounted within 10 meters, so that each camera covers the same viewing area of the physical environment.

The user performed the tasks using a trackball mouse to control the onscreen cursor for direct manipulation, and a Bluetooth button box for command control. For each of the tasks, either scaling or moving, the participant could individually manipulate the object in the X, Y, or Z axes of the object's coordinate system, by clicking to select the object, moving the cursor to scale/move the object along the selected axis, and clicking again to release the object. At the start of each moving task, the virtual window sill object was misplaced at random positions, all equidistant from the correct position. For scaling tasks, the starting size of the window sill was randomly either smaller or larger than the correct size, all by an equivalent ratio. The randomization was done so that through the three iterations including training, the participant would not see the same starting position or size of the virtual object using the same camera, to reduce learning effects.

For each task, the time and number of mouse clicks required to complete the task was recorded. For the Tripod condition, this included the time the participant spent adjusting the tripod in order to complete the task. For the Head condition, the time to locate the physical window sill using the head mounted zoom lens camera was counted. The time to finish each task was decided by the participant when he/she was content with the correct position or size of the virtual object. The final position and size of the virtual object were recorded after each task. A questionnaire was completed at the end of three iterations for a qualitative evaluation of the participant's preferences among different camera positions.

5 RESULTS AND DISCUSSION

Based on the GPS data recorded, we detected an outlier where the GPS position of one of the participants was displaced by a considerable amount from the actual position. Therefore, we discarded the data for this participant and performed analysis on the remaining 15 data sets. We performed ANOVA analysis on the measurement errors in position and size of the virtual object, as well as time to complete the tasks and the number of clicks required. ANOVA error analysis was done separately on error measurements in the X, Y, and Z axes, for both scaling and moving tasks. Based on the first person perspective, the X axis was the depth axes along the normal axis of the user's head mounted display image plane; the Y axis was the horizontal image plane axis, and Z was the vertical image plane axis. Therefore, scaling and moving errors along the X, Y, and Z axes will be referred to as depth, side, and up axes errors, respectively, in the results presented below.

5.1 Error analysis

For the error analysis, there was a significant effect ($p < 0.05$) over the five camera conditions, for all axes in both scaling and moving tasks, see Table 1. A post-hoc analysis on the error measurement was performed with a pairwise t-Test on six pairs of conditions, with a Bonferroni correction ($\alpha < 0.008$). We selected the following pairs to explore different error effects:

1. *Viewing angle:* Comparing the Remote camera and the Fixed tripod camera examines the effects of an oblique viewing angle, because the remote A camera looked at the scene from a 30 degree angle, while the fixed tripod camera shared a direct 90 degree angle view of the virtual object as the user's first person perspective.
2. *Multiple viewport*: Comparing the Remote camera with the Multiple remote cameras to examine the effects of multiple viewpoint against a single view. Both conditions shared the similar oblique viewing angles.

Table 2. Time (in s) and number of clicks (units) for moving and scaling tasks across five conditions

	Time (s)				Number of clicks (unit)			
	Moving task		*Scaling task*		*Moving task*		*Scaling task*	
Condition	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Remote	67.59	25.02	42.75	14.45	9.06	4.03	4.75	2.38
Head	98.38	41.78	82.29	35.34	6.88	2.95	4.03	1.84
Tripod	102.19	36.19	83.56	34.37	7.97	2.95	5.22	2.67
Multiple	110.00	54.00	63.53	23.85	13.53	10.45	5.34	2.83
Fixed	51.15	18.58	45.32	17.90	5.53	2.05	4.41	2.30

3. *Head movement*: Comparing the Head mounted zoom lens camera and Tripod adjustable camera examines the effect of head movement. In both conditions, the cameras were tracked by GPS and orientation sensors; the head mounted camera was additionally affected by head movement.
4. *Sensor noise:* Comparing the fixed tripod camera and the Tripod adjustable camera, in both conditions, the cameras had the first person perspective view, but the Fixed tripod was not affected by any sensor.
5. *Ranking Camera Position:* Remote camera and Head mounted cameras.
6. *Ranking Camera Position:* Remote camera and Tripod adjustable camera.

For the first pairwise t-Test to show the effects of oblique viewing angle, there was a significant effect ($p < 0.008$) to support the hypothesis H1 that the viewing angle adversely affected the precision in moving task in the side and up axes, and scaling task in the depth axis (Fixed tripod performed better than Remote).

For the second pairwise t-Test comparing the effect of multiple viewports, there was no effect ($p > 0.008$) that the multiple viewports adversely affected precision for the moving and scaling tasks in any axis. Hypothesis H4 was rejected.

For the third pairwise t-Test comparing the effects of head movement, there was a significant effect ($p < 0.008$) to support the hypothesis H2 that head movement reduced precision in the moving task *only* in the vertical image plane axis. However, it was noticed that there are no significant affects in any other axes for moving and scaling tasks.

For the pairwise t-Test to show the effects of sensor noise, there was a significant effect ($p < 0.008$) to support the hypothesis H3 that sensor noise degraded precision in the moving and scaling task across all three axes, except for scaling along the side axis.

For the ranking of camera positions, the remote camera showed a significant improvement in precision ($p < 0.008$) over the head mounted camera across all axes in both tasks, except for scaling in the up direction. The remote camera also showed a significant improvement in precision ($p < 0.008$) over the tripod camera across all axes in both tasks, except for scaling along the side axis.

5.2 Time analysis

For the time to complete the task, we performed an ANOVA analysis and found a significant effect ($p < 0.05$) among the five camera conditions for both scaling and moving tasks, see Table 2. A post-hoc analysis of the completion time was performed using a pairwise t-Test on seven pairs of conditions, with a Bonferroni correction ($\alpha < 0.0071$). We performed the analysis on the total time to complete the task, on the same six pairs as examined for error analysis, with an additional pair of Fixed camera and Head camera. The additional pair was added to evaluate the extra time taken to do the task due to the head movement as compared to the baseline Fixed camera condition.

There was a significant effect ($p < 0.0071$) to support the hypotheses H2, H3, and H4 for both scaling and moving tasks regarding the effects of head movement, error sensors using tripod adjustment, and the multiple viewport visualization. Hypothesis H1 was rejected for task time.

For the ranking pairs, there were significant effects that the remote camera improved in time to complete the task ($p < 0.0071$) over the head and tripod cameras for both scaling and moving tasks. There was no significant difference between the tripod and head mounted cameras.

5.3 Number of clicks

For the number of mouse clicks to complete the task, there was a significant effect ($p < 0.05$) over the five camera conditions, only in the moving tasks, see Table 2. A post-hoc analysis on the number of clicks in moving tasks was performed with a pairwise t-Test on seven pairs of condition, with a Bonferroni correction ($\alpha < 0.0071$). We selected the same seven pairs as in the time analysis tasks, to evaluate if the same factors caused the participants to perform more mouse clicks to complete the tasks.

There was a significant effect ($p < 0.0071$) that the participants were required to perform more mouse clicks in the moving task caused by sensor errors using tripod adjustment, and by the remote camera with oblique angle, as compared to a fixed person perspective view. In other words, hypotheses H1 and H3 were supported, while H2 and H4 were rejected. There was no significant difference between the three alternate pairwise tests among the remote, head, and tripod mounted cameras.

5.4 Questionnaire

The participants were asked to rank the three camera positions, head mounted, tripod mounted, and remote mounted (1 point for the most preferred, and 3 points for the lease preferred). Remote camera scored 23 points, tripod 25, and head 42 (the lower points the more preferred). This ranking mostly agrees with the error and task time analysis as presented above. The opinions fluctuated between the remote and the tripod condition. Most explanations for the higher rank of the remote camera were that the remote cameras were employed in a multiple viewport setting and assisted the user in completing the tasks. Among the participants who preferred the tripod camera, there were complaints about the confusion of the oblique angle presented by the remote camera.

5.5 Discussion

From the results of the study, we draw several conclusions regarding the types of errors, the different camera sources, and visualizations, specifically head movement, remote cameras, multiple viewports, and tripod cameras The following list summarizes the conclusions:

1. Head movement error is negligible in comparison to sensor noise.
2. Head movement error does not affect the estimation of size for manipulation tasks.

3. Head movement error on zoom lens cameras does not render the video image too blurry or too unstable for manipulation tasks.
4. Head movement error causes time delay, but does not complicate the manipulation tasks.
5. When the axes of manipulation of the virtual objects are not parallel to the horizontal and vertical axes of the image plane, precision in manipulation tasks is reduced.
6. Because of the discussion point listed above, remote cameras with oblique viewing angles require extra visualization cues to improve precision.
7. Simply adding another image plane from a different viewing angle does not increase precision (discussion point 5). Similarly, multiple viewport visualization on its own does not increase precision (discussion point 6).
8. The reduced mobility of tripod cameras may outweigh its benefits of stability, with the current sensor configurations used in the study.

5.5.1 Head movement

It may seem obvious that sensor noise and head movement would reduce the precision of the augmented viewport technique. However, with the error analysis of the study, we can conclude that sensor error causes reduction in precision to a greater extent than head movement. Sensor noise caused errors across more combinations of tasks and axes of operations than head movement did. The pairwise t-Tests on manipulation errors show that sensor noise had a significant effect in all but one axes in both moving and scaling tasks, while head movement only caused issues for the moving tasks along the vertical axes of the image plane. This can be explained by the fact that sensor noise included GPS that could report errors in the user's location causing precision errors on the depth axis, because the user's position fluctuated to be closer to or further from the remote location. Head movement does not have this issue in the depth axis.

Further investigation reveals the possibility that the significant effect of head movement in the up axis may have been caused by sensor calibration error instead. The zoom lens camera was mounted on the participant's head using an oval frame while the head's orientation sensor (Intersense InertiaCube) was separately mounted on the sunglass-style immersive display (Vuzix Wrap920AR). When the oval frame sat on top of the head, it was not possible to misalign the horizontal orientation (yaw) of the camera to the InertiaCube's, because the oval frame could not freely rotate left or right. However, it was highly likely that the oval frame could slip back and forth on the head and tilt the camera slightly upwards/downwards, due to the different shapes and sizes of the participants' heads. This caused an offset in the vertical orientation (pitch) between the camera and the InertiaCube. This offset eventually affected the error results in the moving task in the up axis, as noted in the post-hoc t-Test between the Head camera and Tripod camera in Section 0. Therefore, we are confident that head movement almost does not significantly affect precision.

The head movement only reduced the precision in the moving task in the up axes, but not affected scaling at all. For the scaling task, the position of the virtual object was fixed in the correct position, overlaying on top of the physical window sill. Head movement would cause the object to be displaced from the correct position; however, despite the misalignment error, the participants were able to complete the scaling task by estimating the size of the physical window in the background of the augmented viewport. Therefore, head movement does not affect the estimation of size for manipulation task.

During a prior pilot study, it was noticed that the use of a zoom lens camera for the head mounted display worsened the head movement at a distance, by the same ratio as the zoom lens bringing the closer view, causing precision error as well as blurry vision and rendering the background image too unstable to be useful. However, based on the results of the study, the participants completed the scaling task unaffected by head movement. Therefore, head movement together with zoom lens camera does not render the imagery blurry or unstable.

Comparing the analysis of completion times and the number of clicks reveals that head movement took a longer time to complete the task but did not require extra clicks. It can be deduced (and through observation during the study) that the participants spent most of the task time trying to stabilize the head mounted camera. Once a stable viewpoint is achieved, it took a similar number of mouse clicks as the fixed tripod condition. Therefore, it can be concluded that head movement does cause time delay, but does not complicate the manipulation tasks.

5.5.2 Remote cameras and multiple viewports

Using a remote camera caused the participants to use more mouse clicks but did not take longer time. From this discrepancy, we can explain that the participants performed a number of exploratory moving and scaling operations in short succession to get used to the oblique viewing angle. Thanks to the stability of this condition and extra mouse clicks, the participants could complete the scaling task without taking extra time and without sacrificing precision in scaling tasks. However, precision suffered for the moving tasks. As explained earlier in this paper, all augmented viewport suffers from the same image plane limitation of being ineffective along the normal axis of the viewport. The analysis indicates that with the seemingly rapid exploratory succession of virtual object movements, it was easy for the participants to move the object into an incorrect position along the normal axis of the viewport. Such an incorrect position could not be detected easily, which led the participant to believe that the task goal was completed. Therefore, it did not take longer time to perform this task, however, the precision suffered.

The opposite situation happened for the multiple viewports condition: taking longer in time but not extra mouse clicks. The extra time was spent on trying to understand the spatial relationship between the two camera viewpoints. There were not significantly more mouse clicks, possibly because the first few mouse clicks of moving or scaling the object introduced visual changes on both viewports. The confusion of the spatial relationship may have led the participants to conclude that extra mouse clicks may not be useful to comprehend the combination of two viewpoints. Therefore, they did not try any more exploratory extra clicks than the single remote viewport condition.

The multiple viewports, however, did not produce any more improvements in precision. It must be noted that the visualizations as described in Section 3.2.1 were not enabled in the study. We excluded the visualizations to reduce the confounding variables of the study. Therefore, the participants were left with only the two video streams from both cameras (see Figure 1) and the ability to perform exploratory manipulation on the virtual objects to figure out the spatial relationship of the viewports, which is what the visualizations described earlier, are designed to support. There were only a few participants that succeeded in the spatial relationship problem, after a few iterations of tasks. This reduction in performance is a well researched topic relating to situation awareness and mental workload, as explored by Veas et al. [19] in their work to present visualizations in assisting the understanding of multiple camera setups from the first person perspective. Similar works in the area of video surveillance investigate different techniques to improve the spatial understanding of camera setups. Notable examples are the video flashlight technique [20], contextualized video [21], and the DOTS system [22]. Our study provides empirical results proving the needs for extra visualizations in a multiple camera setup. We are interested in applying these techniques to improve on the visualizations for the augmented viewport.

5.5.3 Tripod cameras

As can be seen from the results of the error analysis pairwise t-Tests, the sensor error caused the worst and most widespread effect on precise manipulation, agreeing with Holloway's error model [11]. Within the area of interaction research, it is more feasible to attempt to correct the head movement error instead, using vision-based image stabilization, for instance. In Figure 3 showing the variants of camera location, the head mounted camera is affected by head movement and sensor error, but providing the most flexible control of the camera. The tripod mounted camera introduces only sensor error, but takes a longer time to adjust. We also concluded from the study that head movement did not cause any more significant error than the sensors alone, and that the tripod camera required more mouse clicks to complete the same tasks, as well as a bulkier and less mobile setup (this condition required a physical tripod to be mounted next to the participant). Therefore, the advantage of the head mounted camera outweighs its drawbacks when compared to the tripod camera. *Based on this observation, it is suggested that we can focus the augmented viewport techniques on only using the head mounted cameras and existing remote cameras in the environment, thus making the technique more mobile and suitable for outdoor wearable computer systems.*

6 Conclusion

We have presented an extension of the augmented viewport framework of techniques and visualizations for the discovery and utilization of a range of physical cameras for precise action at a distance. The augmented viewport utilizes a range of cameras, including remotely located cameras, head mounted zoom lens cameras, and tripod mounted zoom lens cameras, to offer several potential benefits: closer views of the scene of interest, novel and complementary viewing angles with multiple viewports, stability against sensor noise, and view-dependent interaction to enhance precision. We also presented a user study to investigate the effects of different viewpoint, head movement and sensor noise on zoom lens cameras, as well as the multiple viewport visualization, on the usability of the augmented viewport action at a distance technique. The results of the study showed that head movement only causes minor reduction in precision, and extra visualizations are required to assist the user in understanding the spatial relationship among physical cameras.

Acknowledgement

We would like to thank Ross Smith for his help with building the Tinmith backpack for the study, Wayne Piekarski as the developer of Tinmith, and members of the WCL lab for proofreading this paper.

References

[1] T. N. Hoang and B. Thomas, "Augmented Viewport: An action at a distance technique for outdoor AR using distant and zoom lens cameras," in ISWC, 2010, pp. 117 - 120.

[2] J. S. Pierce, et al., "Image plane interaction techniques in 3D immersive environments," in I3D 1997.

[3] W. Piekarski and B. H. Thomas, "Augmented reality working planes: a foundation for action and construction at a distance," in ISMAR 2004, pp. 162-171, 2004.

[4] D. Schmalstieg and G. Schaufler, "Sewing Worlds Together With SEAMs: A Mechanism to Construct Complex Virtual Environments," Presence: Teleoperators & Virtual Environments,vol.8, pp. 449-461, 1999.

[5]S. L. Stoev, et al., "The through-the-lens metaphor: taxonomy and application," in IEEE VR 2002, pp. 285-286.

[6] K. Hirose, et al., "Interactive Reconfiguration Techniques of Reference Frame Hierarchy in the Multi-viewport Interface," in 3DUI 2006,pp 73-80.

[7] M. F. Deering, "HoloSketch: a virtual reality sketching/animation tool," ACM (TOCHI), vol. 2, pp. 220-238, 1995.

[8] T. V. Do and J. W. Lee, "3DARModeler: a 3D Modeling System in Augmented Reality Environment," International Journal of Computer Systems Science and Engineering, vol. 4, p. 2, 2008.

[9] P. Fiala and N. Adamo-Villani, "ARpm: an augmented reality interface for polygonal modeling," in ISMAR 2005, pp. 196-197, 2005.

[10] S. Frees and G. D. Kessler, "Precise and rapid interaction through scaled manipulation in immersive virtual environments," in IEEE VR, 2005, pp. 99-106.

[11] R. L. Holloway, "Registration error analysis for augmented reality," Presence: Teleoperators & Virtual Environments,vol.6, pp. 413-432, 1997.

[12] J. D. Hol, et al., "Sensor Fusion for Augmented Reality," in International Conference on Information Fusion, 2006, pp. 1-6.

[13] S. You, et al., "Orientation tracking for outdoor augmented reality registration," IEEE Computer Graphics and Applications, pp. 36-42, 1999.

[14] G. Schall, et al., "Global pose estimation using multi-sensor fusion for outdoor Augmented Reality," in ISMAR 2009, pp. 153-162.

[15] R. Behringer, "Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors," in IEEE VR 1999, pp. 244-251.

[16] W. A. Hoff, "Fusion of data from head-mounted and fixed sensors," in IWAR, 1998, pp. 167-182.

[17] S. Guven, et al., "Mobile augmented reality interaction techniques for authoring situated media on-site," in ISMAR 2006, p. 235.

[18] J. C. Chung, "A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams," in I3D 1992, pp. 193-196.

[19] E. Veas, et al., "Techniques for view transition in multi-camera outdoor environments," in Proceedings of Graphics Interface 2010,.

[20] H. S. Sawhney, et al., "Video flashlights: real time rendering of multiple videos for immersive model visualization," presented at the Proceedings of the 13th Eurographics workshop on Rendering, 2002.

[21] W. Yi, et al., "Contextualized Videos: Combining Videos with Environment Models to Support Situational Understanding," IEEE TVCG, vol. 13, pp. 1568-1575, 2007.

[22] A. Girgensohn, et al., "DOTS: support for effective video surveillance," presented at the Proceedings of the 15th international conference on Multimedia, Augsburg, Germany, 2007.