Comparison of User Performance in Mixed 2D-3D Multi-Display Environments

Abhijit Karnik, Tovi Grossman, Sriram Subramanian

Comparison of User Performance in Mixed 2D-3D Multi-Display Environments

Abhijit Karnik[1], Tovi Grossman[2], and Sriram Subramanian[1]

[1] Department of Computer Science, University of Bristol, Bristol, UK
{karnik, sriram}@cs.bris.ac.uk
[2] Autodesk Research, 210 King St. E., Toronto, Ontario, Canada
firstname.lastname@autodesk.com

Abstract. Stereoscopic displays and volumetric 3D displays capable of delivering 3D views have in use for many years. These standalone displays have been investigated in detail for their impact on users' viewing experiences. Effects like aesthenopia and nausea are well-known for flat-screen based stereoscopic displays. However, these devices have not been tested in the context of multi-display environments (MDEs). The performance cost of repetitive switching between a 3D (stereo or volumetric) display and a standard 2D display are not known. In this paper, we perform a thorough user study where we investigate the effects of using such 3D displays within the context of a MDE. We report on our findings and discuss the implications of the same on designs involving such hybrid setups. Our experiments show that in the condition involving two 2D displays which allow for motion parallax and perspective correction, the participants performed the task the fastest.

Keywords: stereoscopic display, autostereoscopic display, volumetric display, zone of comfort, multi-display environment, performance, mental load.

1 Introduction

Multi-display environments (MDEs) combine multiple display elements into a single coherent system. Such systems have been explored in different combinations of tabletop, wall and hand-held setups delivering 2D content as well as non-stereoscopic content. Of late, with the increase of availability of stereoscopic 3D displays (stereo-3D), such devices are also becoming part of MDEs [14, 16, 23]. Another class of 3D displays, which we term as spatial 3D displays, display true depth and are inherently multi-view and autostereoscopic [10, 25, 38]. Both stereo-3D and spatial 3D displays allow an interesting case for MDEs which allow mixed content delivery.

There are demonstrable advantages of stereo-3D and spatial 3D displays in terms of perception of 3D digital content [13, 34] which been studied in detail. However, there is an associated cost with the use of these devices especially stereo-3D. Prolonged use of stereo-3D has been associated with aesthenopia [3] as a combination of blurred vision, headaches, fatigue, nausea and pain. These symptoms are associated with visual-vestibular conflict and vergence-accommodation conflict.

adfa, p. 1, 2011.

Visual-vestibular conflict arises when stereoscopic content is meant to simulate great depth and movement, such as in cinemas. This triggers the brain into assuming motion of the body. However the vestibular organs (in the inner ear) which detect physical motion indicate that the body is still. This results in the effect termed as visual-vestibular conflict. This conflict is less pronounced in desktop and office environments which are physically smaller (than cinema screens) and also afford other environmental clues pointing to the lack of motion. Thus this effect is not considered within scope of this paper.

Vergence-accommodation conflict is more important with the display sizes relevant to MDEs. Normally, human eyes accommodate (rotate inwards or outwards) such that the lines of sight intersect on an object of interest and the focus is adjusted to the same location. However with stereo-3D displays, there is a disparity between the focal plane and the perceived location of the object. This results in the effect termed as vergence-accommodation conflict.

Vergence-accommodation conflict is relevant to MDEs using stereo-3D displays. While fatigue is well-reported for continuous use of stereo-3D displays, a relatively unexplored area is what impact vergence-accommodation conflict has on performance. As expected in a task spanning across a standard 2D display and stereo-3D display, the user would have to switch context between the two devices on a regular basis. Would this context switch aggravate symptoms resulting from vergence-accommodation conflict?

While stereo-3D displays are known to have issues with vergence-accommodation conflict, spatial 3D displays usually don't suffer from such problems. This makes them ideal for tasks involving localized 3D content while delivering realistic 3D views. However, as a part of MDEs, these devices can also impose a performance penalty as the user has to switch between a virtual 2D view and a realistic 3D view.

In the often cited example of using 3D visualization for air traffic control, the controller may be forced to switch between a 3D visualization of the air traffic to a 2D view listing weather conditions or information about inbound aircrafts. If the hybrid nature of the setup affects the performance of the controller in any form, such effects need to be studied. Thus an evaluation of performance becomes critical if such MDEs are to become part of day to day use. Motivated by this, we performed an experiment involving a 3D task involving three scenarios. The experiment aims to answer the question: "What is the effect of repetitive switching between a standalone 3D display and a 2D display during a task involving content spread across both?"

We perform a study that uses a mental rotation task to investigate the effects of using a 2D display in conjunction with either a spatial 3D or a stereo-3D display. Three conditions are studied. In the first condition, we pair a 2D display with another 2D display that supports motion parallax and perspective correction. The second condition involves a 2D display and a stereo-3D display. The last condition uses a swept volumetric display (a type of spatial 3D) paired with the 2D display.

Thus the main contribution of this paper is a systematic investigation of the effects of a hybrid MDE on user performance for a 3D data intensive task. Our experiments show that in the condition involving two 2D displays which allow for motion parallax and perspective correction, the participants performed the task the fastest. However they also achieved higher accuracy when using the stereo-3D display. Finally, the spatial 3D condition had lowest scores for time as well as accuracy. We conclude by offering some possible explanations for these outcomes.

2 Related Work

We consider three different aspects of related research in this section. We wish to explore 3D displays in the context of MDEs, so it is pertinent to explore the MDE literature. Since spatial 3D display based MDEs are not commonly known, we also look at the literature around standalone 3D displays separately. Finally we explore literature related to the cognitive effects that 3D displays have on users.

2.1 Multi-display environments

Multi-display environments (MDEs) that combine interactive tables with wall-mounted displays provide users with enhanced visualization and interaction capabilities. Such setups have been around for a while now. Earlier examples like VIP [1] and ImmersaDesk [7] have demonstrated that multiple views of a task on different projection planes enhances user experience. Similarly, MDEs have been shown to be useful for a range of tasks such as geospatial applications [11], biomolecular modeling [2] and astronomy [40]. Most of these MDEs only explored the combination of planar 2D displays making them 2D display-based MDEs.

With these setups, 3D data is displayed by 3D rendering on a 2D surface [4, 11, 18]. The '3D content' is non-stereoscopic in nature. With user tracking systems (e.g. Kinect) it is possible to provide motion parallax as well as perspective correction [29] for the 3D content. While planar devices are capable of delivering a rich rendering of 3D views, they lack the realism delivered by displaying true 3D objects in terms of accuracy of depth estimation [13] and orientation [15].

More recently, MDEs with inbuilt stereoscopic capabilities have been demonstrated. Systems such as Toucheo [14], Holodesk [16] and PiVOT [23] are capable of generating stereoscopic views collocated with 2D views thus providing a composite MDE that is capable of delivering mixed content. In case of Toucheo and Holodesk, the 2D content appears spatially below the 3D content. With PiVOT the content is collocated but accessed by leaning forward or back. While these are special examples of such MDEs, a simpler example would be one that involves a desktop setup where one display is a 2DD and another is either a stereo-3D or a spatial 3D display. These desktop setups are currently feasible given the availability of 3D monitors and could present the mixed content side by side.

2.2 Standalone 3D devices

Standalone 3D devices fall into two broad categories: Planar stereo-3D displays and spatial 3D displays. With stereo-3D displays, the binocular disparity is generated by delivering two different views to the user's eyes looking at a static planar screen.

Examples of stereo-3D displays include ones using lenticular arrays [28], microlens arrays, parallax barriers [33] or a hybrid combination of these [26]. While there are some glasses-free stereo-3D displays [22], the commercially available state of the art relies mainly on shutter-glass based systems. All these devices work by providing different views to each eye of the user. Thus vergence-accommodation conflict affects all these displays.

On the other hand spatial 3D displays generate views such that the visualized object has real spatial depth and dimensions. To achieve real spatial depth, the relevant points in the volumetric space are turned into point sources of light. The relevant points are representative of the reflective surface(s) of the object(s) allowing perception by the eye. Different methods have been utilized to solve the problem of lighting the volumetric pixel points (voxels). A stack of static but sequentially switched diffusers achieve the true 3D effect in DepthCube [38]. Other approaches apply different physical properties like plasma bubbles generated by a pulsed laser as shown by AIST, Japan, laser-triggered fluorescence [8] and laser-induced damage glass [30]. The swept diffuser technique used by LightField [21], Vermeer [5], Perspecta [10] and its anisotropic implementation [6] with a view-point driven autostereoscopic view have also been demonstrated.

2.3 Visual comfort in 3D setups

It is necessary to first make a case in support of 3D displays (both stereo-3D and spatial 3D) as standalone devices. Price and Lee [34] have shown that performance in spatial cognitive tasks for students improves with stereoscopic imagery. Also, Jin et al. [19] found that stereo-3D provided an advantage when presenting complex structures and spatial relationships. A study involving volumetric displays by Grossman and Balakrishnan [13] showed that volumetric displays (i.e. spatial 3D displays) can provide better depth perception in some tasks, in comparison to stereo-3D displays.

However, since visual discomfort arising from vergence-accommodation conflict is well known for stereo-3D displays, extensive research work has investigated these effects in a purely single display context. Kooi and Toet [24] explored how binocular disparity affects viewing comfort while Tam et al. [39] explored the visual discomfort with respect to a 3D TV setting. The dynamic accommodative response to stimuli corresponding to stereo-3D display was studied by Oliveira et al. [31] showing effects due to vergence-accommodation conflict. Emoto et al. [9] showed that repeated vergence adaptation leads to decline in visual functions. Lastly, work by Shibata et al. [36] and Hoffman et al. [17] have explored as to how the visual performance degrades while working with stereo-3D displays.

While these visual fatigue effects associated with stereo-3D displays have been shown in detail by prior research, there is little literature regarding effects on cognitive load when a 3D display is used in tandem with a 2D display. Paas et al. [32] describes that cognitive load can be measured through three properties: mental load, mental effort, and performance. They also mention that mental load and mental effort are more difficult to measure since they involve the use of secondary tasks. However, performance can be measured in terms of item accuracy and completion time. Thus, a task can be designed to measure performance and extrapolate it to cognitive load resulting from a particular setup. This leads us to our experiment.

3 Experiment

Our experiment is used to determine the impact that integrating 3D displays into MDE environments has on a user's viewing experience, visual comfort, and task performance. While there are multiple options of configuring a 3D MDE, we consider a desktop configuration with two displays side by side; as such systems could be easily adopted into existing workplaces today. We studied three possible display combinations, as described in the experiment conditions below.

3.1 Conditions

It has already been established that 3D images, irrespective of display type, are better for performing shape understanding tasks [37]. Instead, our goal is to determine the impact of hybrid 2D-3D display environments. The following three display combinations were used:

- **2D-2D:** In this condition the first display was a static 2D display. The second display was also a 2D display, but head tracking was used to present perspective corrected views that would also respond to motion parallax. Thus the view would be regenerated based on the head-position of the user and would seem three dimensional whenever the user moved their head.
- **2D-3D:** In this condition, we combined a static 2D display (as described above), and a stereo-3D display. The stereo-3D display also used head tracking to provide perspective corrected views.
- **2D-VO:** This condition combined the static 2D display with a 3D volumetric display.

While an additional condition 3D-VO would be possible, it was not considered since our assumption is that in the MDE setup, the task always has a 2D display element.

3.2 Task

We wished to identify a task which presented a significant amount of cognitive load on the participant and also required frequent switching between the two displays. We chose a modified form of the Shepard-Metzler Mental Rotation test (SMT) [35] as the experimental task. The ability to rotate two and three-dimensional objects in the mind is known as mental rotation. The SMT is used to test the ability of a participant to accurately and rapidly mentally rotate three dimensional objects. The original SMT used two images each containing a 3D shape made up of cubes connected at the face. The shape in the second image is either a) the same 3D shape but rotated along one of the 3 axes or b) the mirrored (along one of the 3 axis) version of the 3D shape and then rotated.

In our task, the shape is made up of 1 unit diameter spheres connected to each other as a chain. The shapes that were select satisfied the criteria that they were three dimensional (having 3-4 non-coplanar 90° bends) and fit inside a 5×5×5 grid. The number of spheres per shape ranged from eleven to seventeen. Like the original SMT, the task was to identify if the two shapes presented on the displays were same or mirrored. Our task differs from the original SMT in terms of rotation. While the original SMT rotates the shapes along one axis, in our task, the rotation can happen along all three axes at the same time. We defined two difficulty levels, easy and hard. With the 'easy' level, the one shape is rotated by <30° along a random axis as compared to the other shape. With the 'hard' level, the second shape was rotated along all three axes such that the sum of the rotations was >60°. Thus, based on *match* state (*same* or *mirror*) and *difficulty level* (*easy* or *hard*), four different combinations (as shown in Figure 1) were possible.

The participants had to identify if the shapes were same or mirrored and indicate their answer via key presses. The detailed process is described in Section 3.5 (Procedures).

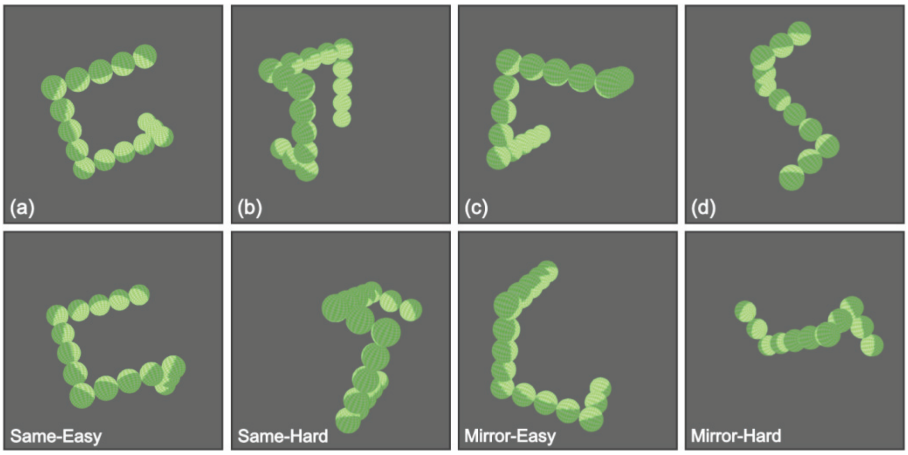

Fig. 1. Shapes used for the task. Top row shows the four shapes used. Bottom row shows the paired shape with *difficult level* (*easy* or *hard*) and *match* state (*same* or *mirrored*)

3.3 Apparatus.

The setup consisted of three displays placed next to each other in front of a plain background.

1. **2D display:** This was a Dell 21" 1920×1080 pixel monitor. Usually, the shutter glasses used for the 3D TV interact with some 2D monitors. This causes the screens to appear black through the glasses. The monitor we used did not get affected in the same way.
2. **Stereo-3D display:** A 40" (1920×1080 pixel) Samsung UN40ES6500 3D TV was used as the stereo-3D display. The display operates in a side-by-side mode for 3D

thus allowing an effective resolution of 960×1080 pixels for 3D mode using active shutter glasses.

3. **Volumetric 3D display:** A Perspecta display [10] from Actuality Systems was used as the volumetric display. Perspecta is a swept volumetric display with a resolution of 100 million voxels and a 10" spatial display diameter. It is one of the few spatial 3D displays that was ever available commercially, and has been used for numerous other experiments [6, 12-13].

The first two displays were connected to an Intel Core i7 machine running Windows 7. The Perspecta display was connected to a standalone Windows XP machine. The two machines were networked together for exchanging experiment state information. For tracking the user's head-pose, we used NaturalPoint's Optitrack Duo. The Optitrack Duo uses a marker constellation to provide spatial position of the tracked object along with its orientation in space with sub-millimeter accuracy. The head-pose information was used to present motion parallax such that the users could 'look around the corners' of the 3D shapes used for the experiment. Head-pose information was received by the master program on the first machine via VRPN. The master program intercepted the user's inputs and communicated updates to the slave program running on the second machine via OSC messages. The setup is shown in Figure 2.

The three displays have very different physical dimensions. Thus it was necessary to ensure that the field of view (FOV) coverage of voxels of Perspecta should be comparable or similar to the FOV coverage of the pixels of both the stereo-3D and the 2D display. The actual positions of the shape and the pixel dimensions of the shapes were adjusted on the stereo-3D and the 2D displays to match that of Perspecta. In each case the effective physical dimension of the shapes was 9" and they were all aligned horizontally, spaced 25" apart. The shapes shown on the stereo-3D display were always within the Percival zone such that the experiment was run within the specifications of Shibata's results [36]. Also since the Perspecta requires low light operation, the whole experiment was run in a darkened room.

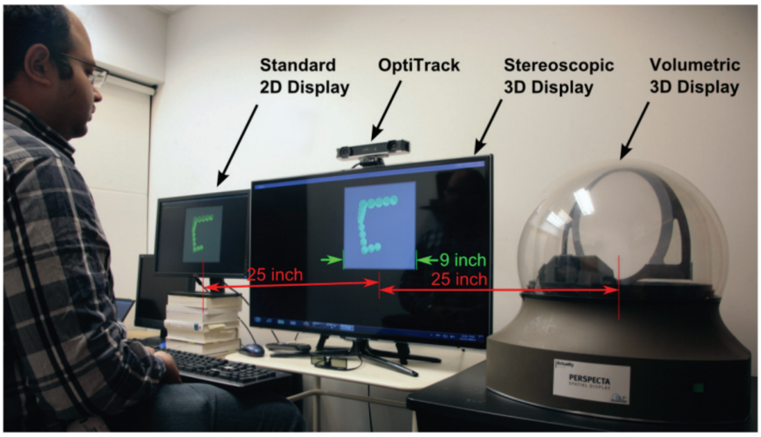

Fig. 2. Experimental setup: All three shapes (one on each display) were aligned horizontally and had the same physical dimensions.

3.4 Participants

A total of twelve participants (7 male, 5 female) were recruited from a local university and via Craigslist. Participants' ages ranged from 18 to 40 years (nine participants from the 21-30 years age-group). All had normal stereo-acuity which was verified through the Titmus-Wirt Fly test. Participants who regularly wore optical correction (5) wore the correction during the experiment (2 wore contact lenses, 3 wore spectacles). All subjects had no prior experience with the task. Also none of them had any experience with Perspecta and had never used a stereo-3D display in a work setting. The participants were compensated for their time with a gift voucher.

3.5 Procedure

For the 2D-2D condition, the 2D display displayed one of the shapes as a fixed 2D shape. This shape did not respond to the user's head movement. The stereo-3D display was operated as a 2D display and it showed the second shape as a single flat 2D shape. This shape responded to the user's head movement allowing motion parallax based viewing. For the 2D-3D condition, the second shape was displayed as a stereo-3D shape while still allowing for motion parallax. The participant had to wear the shutter glasses to view the shape correctly. For both conditions, the program reoriented the shape to a perspective correct orientation thus giving a sensation of 3D. For the third condition (2D-VO), given the physical arrangement of the three displays, the stereo-3D display was operated as a normal 2D display and the first immobile 3D shape was displayed on it. The second shape was displayed on Perspecta. Since Perspecta is autostereoscopic, there was no need to use the head-pose information to reorient the shape.

Before performing the experiment, the participant's stereo-acuity was confirmed by the Titmus-Wirt Fly test. We intended to reject participants who failed the test, however all the participants passed the test. After the test, the participants were acclimatized with setup (especially Perspecta) and then the task was explained. The participants were seated 30" from the display plane. They were encouraged to move their head right and left but asked to limit the motion towards or away from the displays. The actual experiment consisted of 2 phases. During the initial phase, consisting of 20 trials, the users were allowed to get used to the experimental procedure. Data from this phase was discarded. The second phase was the actual experiment. The participants were asked to perform the task as quickly as possible and were made aware that the accuracy of their answers was also being recorded.

For a single trial, two shapes would be displayed on the two displays (relevant to the test condition). To clearly demarcate switching of context between the two devices, only one display would should its shape at any given time. To switch to the other shape, the participant had to press the spacebar. The participant had to indicate the match state (same or mirror) through a single key-press. Once an answer was given, the experiment moved onto the next trial. Since a minimum of one switch would be required to see both the shapes, the program would not allow the experiment to move to the next trial if the answer was indicated without a single switch. This prevented participants from accidentally or intentionally skipping through the trials.

The experiment consisted of a total of 300 trials divided into blocks of 100 trials per condition per participant. For a single block of 100 trials, there were 50 *hard* and 50 *easy* trials. Also, the block had 50 *same* and 50 *mirror* shape pairs. The order of *easy* and *hard* trials was randomized to prevent monotony. The participants were not allowed to pause between trials but were allowed to take a 10 minute break between conditions. The participants filled out questionnaires before and after each block and at the end of the experiment. The learning and order effects were counterbalanced by changing the order of the conditions per participants using a Latin square design.

3.6 Measures

The experiment was run as a within-subjects design for the three conditions being compared. For the task metrics, *difficulty level* was used as an additional variable. The following data was collected from the participants:

Task metrics.
We measured four details per trial. The number of *switches* between the two displays was logged along with the answer given by the participant. *Accuracy* of the answer was binary, either right or wrong. The *total time* taken to perform each trial was also logged. Since it was possible to differentiate between the time spent on one display versus the other (only one display was active at a time), the time spent viewing the 2D shape and the time spent viewing the 3D shape were logged separately. We wished to investigate if the participants spent more time on either the 2D display or the other display. So *time non-2D* metric was computed as the percentage time spent on the non-2D display (100×Time spent on non-2D/Total time). The time metrics were recorded in milliseconds with the accuracy derived from the system clock.

Head Pose.
The head-position information was being used by the program to render perspective correct views for the 2D-2D and the 2D-3D conditions. However, we also recorded the head-pose (as to where the head was pointed) as another parameter. Given the distance between the two displayed shapes (25”), simple saccadic motion of the eyes was not enough and the participants resorted to turning their heads to view the displayed shapes. This turning of the head could be detected by the tracking system and could be sampled 60 times a second. We used this to generate heat maps of where the participant was looking and for how long.

Questionnaires.
We used three different questionnaires during the experiment. Similar to Shibata et al. [36] we wished to record any occurrence of symptoms usually associated with stereo-3D displays. We used a *symptom questionnaire* adapted from their study as shown in

Figure 3 left. Since each participant could have a different starting symptom state, we administered a pre-trial questionnaire and a post-trial questionnaire for each condition. The questionnaires were answered on a digital form requiring the users to click on the desired answers. We also administered a NASA TLX questionnaire after the end of each block. The NASA TLX questions gauged the mental demand, physical demand, pace of task, perceived success, effort and irritation for each block.

The last questionnaire was a *ranking questionnaire* (as shown in Figure 3 right). We asked the participants to rank the three conditions as per their preference for each question. The participant answered this five-question questionnaire after the completion of the experiment. The first three questions measured the participants' perception about the three conditions as compared to each other for the symptoms of fatigue, eye irritation and headache. The last two questions measured the participants' preference for a particular condition and the associated easiness enabled by the condition.

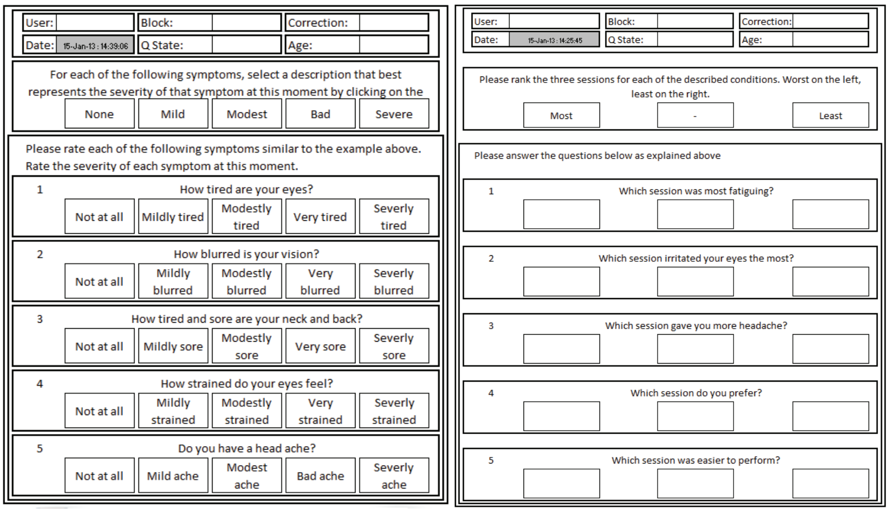

Fig. 3. Questionnaires used for evaluation of symptoms and task load. (left) Symptom questionnaire (right) Final ranking questionnaire. Both tests were administered digitally.

3.7 Results

The data logged for the test and collected from the questionnaires was analyzed with SPSS 19. These are presented below.

Task metrics.

For the task metrics, the experiment presented as a within-subjects repeated measures design with two independent variables as the *conditions* (3 groups: 2D-2D, 2D-3D and 2D-VO) and *difficulty level* (2 groups: *Easy* and *Hard*). We averaged the results (except for accuracy, which was summed) per user for each *condition* and then analyzed the results through two-way repeated measures ANOVA with Bonferroni correction. We looked at the main effects of *condition* and *difficulty level* as well as interaction between the two. We expected that the main effect of *difficulty level* would present as lower accuracy and higher time as well as switches for the *hard* trials.

1. *Accuracy*: The accuracy metric measured if the participants correctly identified the trail pair of shapes to be mirrored or the same. A maximum score of 50 was possible for each combination of *condition* versus *difficulty level*. There was a significant main effect of *condition*, $F_{(2, 22)} = 8.51$, $p<0.005$. The 2D-3D condition had highest average accuracy while the 2D-VO condition fared the worst. As expected, there was a significant main effect of *difficulty level*, $F_{(1, 11)} = 6.34$, $p<0.05$, with *easy* trials having higher average accuracy. Also, there was a significant interaction between *condition* and *difficulty level*, $F_{(2, 22)} = 5.84$, $p<0.05$. The results are shown in Figure 4 left.
2. *Switches*: The metrics were calculated for the average number of switches performed by user per combination of *condition* and *difficulty level*. There was no significant main effect of *condition* on the number of switches. However there was a significant main effect of *difficulty level*, $F_{(1, 11)} = 9.28$, $p<0.05$. The participants performed more switches for *hard* trials as compared to *easy* trials. There was a significant interaction between *condition* and *difficulty level*, $F_{(1.24, 13.66)} = 8.2$, $p<0.05$. The results are shown in Figure 4 right.

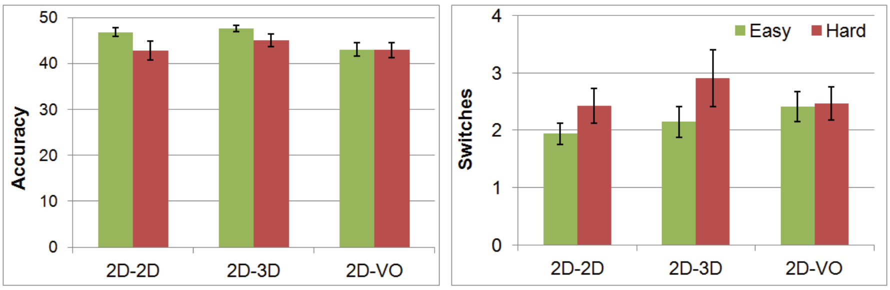

Fig. 4. Task metrics analysis with standard error-bars. The green bars for each condition correspond to the Easy task and the red for Hard task. (left) Average accuracy achieved by participants with standard error-bars. Maximum possible score was 50. (right) Average number of switches performed before arriving at the answer

3. *Total time*: The average of total time taken per trial was used for analysis across the six combinations. We did not find any significant difference between *conditions* for total time thus implying a lack of significant main effect. For *difficulty levels*, however, we found that there was a significant main effect, $F_{(1, 11)} = 7.37$, $p<0.05$. We also found a significant interaction between *condition* and *difficulty level*, $F_{(1.33, 14.61)} = 5.12$, $p<0.05$. Testing for within-subjects contrasts, we found significant difference between 2D-VO and 2D-2D *conditions*, $F(_{1, 11}) = 10.54$, $p<0.05$. The results are shown in Figure 5 (right).

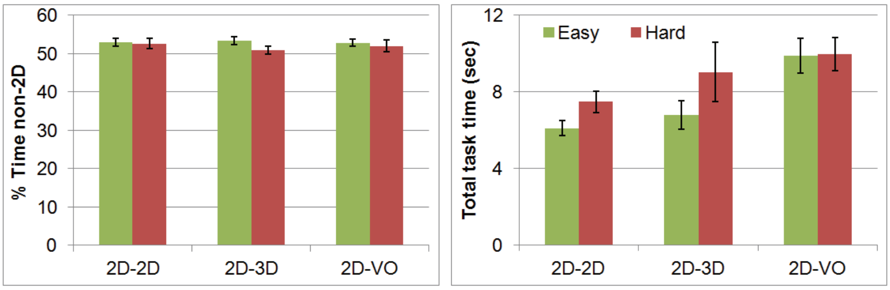

Fig. 5. Task metrics analysis with standard error-bars. The left bars for each condition correspond to the Easy task and the right for Hard task. (left) Average of percentage of time spent by participants looking at the non-2D display. (right) Average of total time in Seconds taken by participants per trial.

4. *Time non-2D*: We found no significant difference between the three conditions for the amount of time spent on non-2D display. Similarly, no significant difference was observed for main effect of *difficulty levels* as well as interaction between *condition* and *difficulty levels*. The results shown in Figure 5 (left) show near similar averages across all combinations.

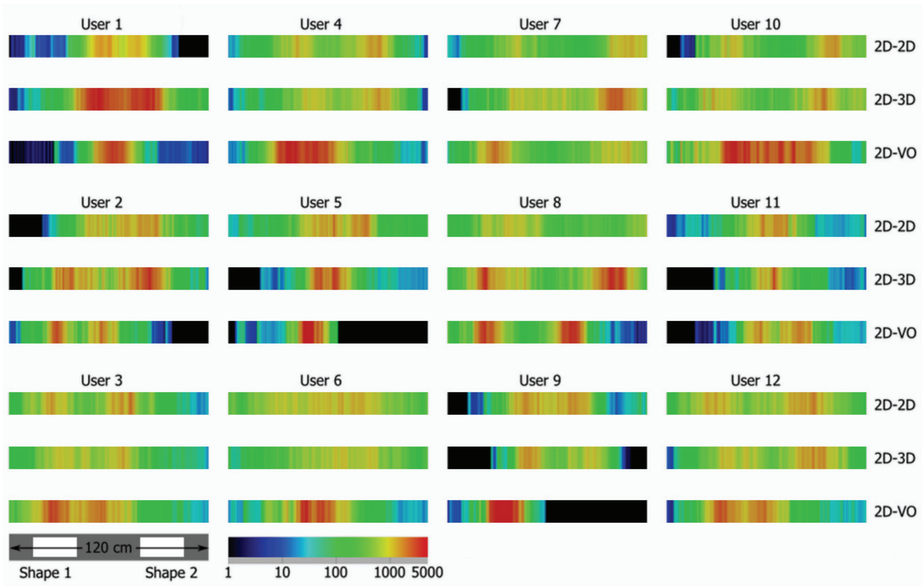

Fig. 6. Heat maps showing how much time a participant spent looking at what part of the screens. Color scale is logarithmic and clamped at 5000 samples.

Heat maps.

The head-pose was available as a 3D spatial coordinate of the tracker and a local rotation of the tracker. Since the tracker was mounted on the participant's forehead, it gave a good representation of where the user was looking. The participants did look down at the keyboard when they had to enter their answer. The head-pose records related to these events (which showed the tracker orientation almost parallel to the ground) were removed from analysis. The remaining records were used to generate the heat maps. To compare the head-pose characteristics of different participants, we chose to only represent the horizontal location of the head-pose for drawing the heat map as the x-axis of the map. The tracker provided 60 samples/second and the number of samples at each specific location were used to draw the map. As shown in Figure 6, the x-axis shows the 120 cm region centered on the two shapes. The color scale is a logarithmic scale and is clamped at 5000 samples.

Questionnaires.

NASA TLX.
The NASA TLX questionnaires were administered for each condition after the trial block ended. Hence, we did not look for differences due to *difficulty levels* and were looking for effects of the *conditions* only. The questionnaire data was analysed using Kruskal-Wallis test. For all questions excluding the first one (How mentally demanding did you find the task?), we found no significant difference between the three conditions. For *Mental Demand*, there was a significant difference, $H(2) = 6.87$, $p < 0.05$. Mann-Whitney tests were used to follow up this finding (3 pairs) with Bonferroni correction, (significance level at *0.0167*). The *Mental Demand* was not significantly different between 2D-2D and 2D-3D ($U = 71$, $r = -0.01$) as well as 2D-3D and 2D-VO ($U = 37.5$, $r = -0.41$). However, *Mental Demand* for 2D-VO as compared to 2D-2D was significantly higher ($U = 29$, $r = -0.51$).

Symptom Questionnaire.
The symptom questionnaire results were analyzed separately for each condition and question using Wilcoxon signed-rank test. We found no significant change in symptoms for *Vision clarity* and *Headache* for any of the conditions. Also, for *Eye Tiredness* there was no significant change for the conditions 2D-2D and 2D-3D. However, we found significant increase in *Eye Tiredness* for 2D-VO, $z = -2.236$, $p < .05$, $r = -0.46$. Similarly, we found that the participants reported a significant increase for *Neck and Backache* for 2D-3D, $z = -2.0$, $p < .05$, $r = -0.41$ and 2D-VO, $z = -2.236$, $p < .05$, $r = -0.48$. Lastly, we found significant increase in *Eye Strain* for all three conditions: for 2D-2D, $z = -2.24$, $p < .05$, $r = -0.48$; for 2D-3D, $z = -2.65$, $p < .05$, $r = -0.54$ and for 2D-VO, $z = -2.07$, $p < .05$, $r = -0.42$.

Ranking Questionnaire.
The ranking questionnaire results are presented in Figure 7 (right). For *headache*, the participants consistently ranked 2D-VO condition as the worst (a lower average rank). This does not tie in well with the symptom questionnaire where we found no significant difference between the three conditions as well as the verbal feedback of the participants stating that they didn't get a headache from any of the conditions. For overall *fatigue*, the 2D-VO condition again was ranked the worst while 2D-3D condi-

tion was ranked the best. For the remaining questions (*Eye Irritation*, *Easier* and *Preferred*), the results were tied. We did not see a clear trend of preference for any of these questions. All participants ranked *Preferred* for the three conditions in exactly the same order as they ranked the conditions with respect to *Easier*.

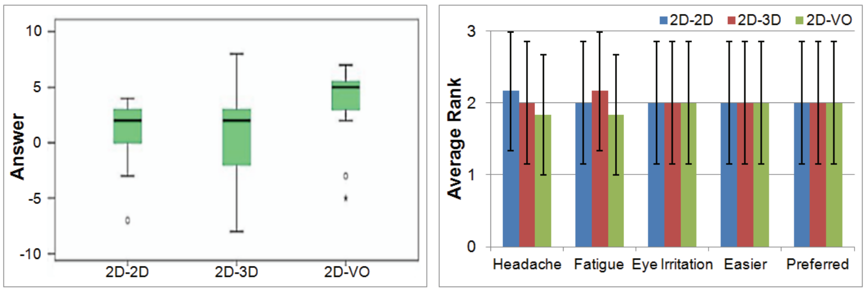

Fig. 7. Questionnaire analysis. (Left) Boxplot for NASA TLX *Mental Demand* shows higher indications for 2D-VO condition. (Right) Ranking questionnaire analysis.

4 Discussion

4.1 Interpretation of results

2D-2D condition.
On average, participants performed less switches within the 2D-2D condition. When factored for *difficulty level*, the 2D condition had higher accuracy than the 2D-VO condition. The participants also took less time to complete the task.

The 2D-2D condition also presented the least change in symptoms. Only the *Eye Strain* symptom was aggravated by the condition, but again a similar change was seen in the remaining two conditions. *Mental demand,* as measured by NASA TLX, was significantly lower than 2D-VO. In the overall ranking, 2D-2D was ranked lower for *headache* and *fatigue.*

While preference scores were tied, the 2D-2D condition performed better and caused fewer symptoms. Thus, it is possible that 3D cues provided by simple motion parallax are somewhat better suited than those afforded by stereo-3D. This is comparable to the results of Johnston et al. [20] where multi-frame motion parallax is shown to work better than stereopsis.

2D-3D condition.
The 2D-3D condition has the best accuracy as compared to all the other conditions. As per the *final questionnaire*, the 2D-3D condition was ranked the least in terms of *fatigue.* However on average, participants performed more switches and took longer than the 2D-2D condition. The condition also aggravated the symptoms of *Eye strain* and *Neck & Backache*.

Thus we can say that the 2D-3D condition has an advantage in case of tasks where accuracy is crucial. Our participants reported it to be less fatiguing but also agreed with the previously observed results indicating *eye strain* resulting from vergence-accommodation conflict. It is important to note that in our setup the objects did not present extreme negative or positive parallax. This could be one of the reasons as to why other symptoms were not reported. Our results are in line with Shibata et al.'s [36] findings. We can conclude that for the 2D-3D condition, if we limit extreme parallax we gain on accuracy with lesser fatigue.

2D-VO condition.
The outcomes of the experiment suggest that 2D-VO condition fares worst in terms of accuracy and average task time. Even if we only consider the *hard difficulty level*, we see that the results are just marginally better than the 2D-2D condition for accuracy and still worse for average task time. From the NTLX scores for *mental demand*, we find that it has higher values than the other two conditions. These results are unexpected. We assumed that a more realistic representation of objects would help recover more information and hence help accuracy.

4.2 Implications

A general and direct implication of these results is that an MDE consisting of a spatial 3D display and a 2D display should be avoided for spanned tasks involving high cognitive load. This was contrary to our initial expectation that 2D-VO condition would be significantly better than the rest. However, as pointed out by Grossman and Balakrishnan [13], there are a few mitigating factors for poor performance of a spatial 3D display. Even as of today, the display quality of such devices is not at par with that of 2D and stereo-3D displays. There are artefacts in the display (for e.g. the central spine of the display cannot show any information and colour quality) which can influence the results. However, it is also important to note that in our case, the experimental setup was designed such that these effects were minimized. The visual size of the output of all three displays was matched and there were no extra visual cues provided by the shapes shown on the 2D display or the stereo-3D display.

When compared to the 2D-2D condition, it is possible that there are other focus based factors affecting the performance for 2D-VO condition. For the side-by-side 2D displays, the user has a fixed reference to a focal point on the plane of the display. However with a spatial 3D display, there is no central plane and thus no central point of focus. This can add to mental load when there is switching between the displays.

Also as the visualization of the shape in true 3D makes it look more real, it is possible that the switching process becomes one where the user has to switch contexts (from virtual world to real world) and they potentially do not view the two displays as a part of the same system. Such a situation also arises when the user has to switch focus between a physical object and a virtual object. Surprisingly, we could not find any research that investigates performance effects while comparing a purely virtual context to one with mixed context. The closest work is in the tangible literature by Marshall et al. [27] where they cautiously suggest that in a single user instance, a tangible interface is not necessarily better. Thus further investigation in this regard is warranted.

It is also possible that users find it difficult to compare a true 3D shape with a 2D shape that the first display shows. We refrained from using perspective correction on the 2D shape for the 2D-VO condition as the static 2D served as a common control shape to all three experimental conditions.

Lastly, as a recommendation for selection of 3D display elements for MDEs, we feel that there is a possible benefit of amalgamation of the 2D-2D and 2D-3D condition. The 2D-2D condition allowed perspective corrected views and motion parallax for the second shape. Comparing its accuracy results with 2D-3D, the overall results for symptoms and NTLX we can argue that this may be an ideal configuration for prolonged use tasks. With the availability of low cost desktop based head tracking systems, it might be beneficial to have a stereo-3D display which is operated mainly in 2D mode but allow perspective corrected views for presenting 3D. Only when the task involves high density of 3D elements, the device can switch to stereoscopic mode thus adding binocular disparity as another cue.

4.3 Future Work

We do not believe that a spatial 3D display is unsuitable for MDE setups. For tasks similar to our experimental task, our results hold true. However there may be other tasks with minimal cross-device contexts wherein a spatial 3D display may prove to be more beneficial. Future work could be used to explore the impact which the task has on our results. Furthermore, effects like aesthenopia and nausea were not a major factor in any of our conditions. However, with prolonged use, the effects may become most prevalent in the 2D-3D condition. Another factor of our tested MDE setups is that they all contained only 2 displays. In the future, it may be interesting to understand the impact of having a greater number of displays, or including displays with larger form factors and with arrangements that favor the type of 3D display used.

5 Conclusion

We have investigated the performance cost of repetitive switching between a 3D (stereo or spatial) display and a standard 2D display in context of a MDE. The experimental results prove that there is a cost involved with the scenarios involving a 2DD and a spatial 3D display which is higher than other scenarios. The results should provide a guideline for the design of MDEs utilizing either spatial 3D or stereo-3D elements.

6 Acknowledgments

We thank Daniel Wigdor and DGP Toronto for helping out with the participant recruitment. We also thank Anne Roudaut for her inputs. This work is supported by the European Research Council (Starting Grant Agreement 278576) under the Seventh Framework Programme.

7 References

1. Aliakseyeu, D., Martens, J.-B., Sriram Subramanian, M., Vroubel, W.W.: Visual Interaction Platform. In: Proc. INTERACT '01, pp. 232-239. IOS Press, (2001)
2. Bailly, G., Nigay, L., Auber, D.: NAVRNA: visualization - exploration - editing of RNA. In: Proc. AVI '06, pp. 504-507. ACM, New York (2006)
3. Berezin, O.: Digital cinema in Russia: Is 3D still a driver for the development of the cinema market. 3D Media 2010, (2010)
4. Bowman, D.A., North, C., Chen, J., Polys, N.F., Pyla, P.S., Yilmaz, U.: Information-rich virtual environments: theory, tools, and research agenda. In: Proc. VRST '03, pp. 81-90. ACM, New York (2003)
5. Butler, A., Hilliges, O., Izadi, S., Hodges, S., Molyneaux, D., Kim, D., Kong, D.: Vermeer: direct interaction with a 360° viewable 3D display. In: Proc. UIST '11, pp. 569-576. ACM, New York (2011)
6. Cossairt, O.S., Napoli, J., Hill, S.L., Dorval, R.K., Favalora, G.E.: Occlusion-capable multiview volumetric three-dimensional display. Appl. Opt. 46, 1244-1250 (2007)
7. Czernuszenko, M., Pape, D., Sandin, D., DeFanti, T., Dawe, G.L., Brown, M.D.: The ImmersaDesk and Infinity Wall projection-based virtual reality displays. SIGGRAPH Comput. Graph. 31, 46-49 (1997)
8. Ebert, D., Bedwell, E., Maher, S., Smoliar, L., Downing, E.: Realizing 3D visualization using crossed-beam volumetric displays. Commun. ACM 42, 100-107 (1999)
9. Emoto, M., Niida, T., Okano, F.: Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. Display Technology 1, 328-340 (2005)
10. Favalora, G.E.: Volumetric 3D Displays and Application Infrastructure. Computer 38, 37-44 (2005)
11. Forlines, C., Esenther, A., Shen, C., Wigdor, D., Ryall, K.: Multi-user, multi-display interaction with a single-user, single-display geospatial application. In: Proc. UIST '06, pp. 273-276. ACM, New York (2006)
12. Grossman, T., Balakrishnan, R.: The design and evaluation of selection techniques for 3D volumetric displays. In: Proc. UIST '06, pp. 3-12. ACM, New York (2006)
13. Grossman, T., Balakrishnan, R.: An evaluation of depth perception on volumetric displays. In: Proc. AVI '06, pp. 193-200. ACM, New York (2006)
14. Hachet, M., Bossavit, B., Cohé, A., Rivière, J.-B.d.l.: Toucheo: multitouch and stereo combined in a seamless workspace. In: Proc. UIST '11, pp. 587-592. ACM, NY (2011)
15. Hancock, M., Nacenta, M., Gutwin, C., Carpendale, S.: The effects of changing projection geometry on the interpretation of 3D orientation on tabletops. In: Proc. ITS '09, pp. 157-164. ACM, New York (2009)
16. Hilliges, O., Kim, D., Izadi, S., Weiss, M., Wilson, A.: HoloDesk: direct 3d interactions with a situated see-through display. In: Proc. CHI '12, pp. 2421-2430. ACM, NY (2012)
17. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Vision 8, (2008)
18. Jiang, H., Wigdor, D., Forlines, C., Borkin, M., Kauffmann, J., Shen, C.: LivOlay: interactive ad-hoc registration and overlapping of applications for collaborative visual exploration. In: Proc. CHI '08, pp. 1357-1360. ACM, 1357266 (2008)
19. Jin, Z., Zhang, Y., Wang, X., Plocher, T.: Evaluating the Usability of an Auto-stereoscopic Display. In: Jacko, J. (ed.) Human-Computer Interaction. Interaction Platforms and Techniques, vol. 4551, pp. 605-614. Springer Berlin Heidelberg (2007)

20. Johnston, E.B., Cumming, B.G., Landy, M.S.: Integration of stereopsis and motion shape cues. Vision Research 34, 2259-2275 (1994)
21. Jones, A., McDowall, I., Yamada, H., Bolas, M., Debevec, P.: An interactive 360° light field display. In: Proc. SIGGRAPH '07 ETech, pp. 13. ACM, New York (2007)
22. Karnik, A., Mayol-Cuevas, W., Subramanian, S.: MUSTARD: a multi user see through AR display. In: Proc. CHI '12, pp. 2541-2550. ACM, New York (2012)
23. Karnik, A., Plasencia, D.M., Mayol-Cuevas, W., Subramanian, S.: PiVOT: personalized view-overlays for tabletops. In: Proc. UIST '12, pp. 271-280. ACM, New York (2012)
24. Kooi, F.L., Toet, A.: Visual comfort of binocular and 3D displays. Displays 25, 99-108 (2004)
25. Langhans, K., Bahr, D., Bezecny, D., Homann, D., Oltmann, K., Oltmann, K., Guill, C., Rieper, E., Ardey, G.: FELIX 3D display: an interactive tool for volumetric imaging. In: Proc. Stereoscopic Displays and Virtual Reality Systems IX, pp. 176-190. SPIE, (2002)
26. Lin, Q., Qiong-Hua, W., Jiang-Yong, L., Wu-Xiang, Z., Cheng-Qun, S.: An Autostereoscopic 3D Projection Display Based on a Lenticular Sheet and a Parallax Barrier. J. Display Technology 8, 397-400 (2012)
27. Marshall, P., Rogers, Y., Hornecker, E.: Are tangible interfaces really any better than other kinds of interfaces? In: Proc. Tangible User Interfaces in Context and Theory Workshop, CHI '07. ACM, New York (2007)
28. Matusik, W., Pfister, H.: 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. ACM Trans. Graph. 23, 814-824 (2004)
29. Nacenta, M.A., Sakurai, S., Yamaguchi, T., Miki, Y., Itoh, Y., Kitamura, Y., Subramanian, S., Gutwin, C.: E-conic: a perspective-aware interface for multi-display environments. In: Proc. UIST '07, pp. 279-288. ACM, New York (2007)
30. Nayar, S.K., Anand, V.N.: 3D Display Using Passive Optical Scatterers. Computer 40, 54-63 (2007)
31. Oliveira, S., Jorge, J., González-Méijome, J.M.: Dynamic accommodative response to different visual stimuli (2D vs 3D) while watching television and while playing Nintendo 3DS Console. Ophthalmic and Physiological Optics 32, 383-389 (2012)
32. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.M.: Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychol. 38, 63-71 (2003)
33. Perlin, K., Paxia, S., Kollin, J.S.: An autostereoscopic display. In: Proc. SIGGRAPH '00, pp. 319-326. ACM, New York (2000)
34. Price, A., Lee, H.-S.: The Effect of Two-dimensional and Stereoscopic Presentation on Middle School Students' Performance of Spatial Cognition Tasks. J Sci Educ Technol 19, 90-103 (2010)
35. Shepard, R.N., Metzler, J.: Mental Rotation of Three-Dimensional Objects. Science 171, 701-703 (1971)
36. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: Predicting visual discomfort with stereo displays. Vision 11, (2011)
37. St. John, M., Cowen, M.B., Smallman, H.S., Oonk, H.M.: The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks. Human Factors 43, 79-98 (2001)
38. Sullivan, A.: DepthCube solid-state 3D volumetric display. In: Proc. Stereoscopic Displays and Virtual Reality Systems XI, pp. 279-284. SPIE, (2004)
39. Tam, W.J., Speranza, F., Yano, S., Shimono, K., Ono, H.: Stereoscopic 3D-TV: Visual Comfort. IEEE Trans. Broadcasting 57, 335-346 (2011)
40. Wigdor, D., Jiang, H., Forlines, C., Borkin, M., Shen, C.: WeSpace: the design development and deployment of a walk-up and share multi-surface visual collaboration system. In: Proc. CHI '09, pp. 1237-1246. ACM, New York (2009)