AnnoScape: Remote Collaborative Review Using Live Video Overlay in Shared 3D Virtual Workspace

Austin Lee*[1] Hiroshi Chigira*[3] Sheng Kai Tang[1] Kojo Acquah[2] Hiroshi Ishii[1]

[1]MIT Media Lab
Cambridge, MA 02139 USA
{aslee, tonytang, ishii}@media.mit.edu

[2]MIT EECS
Cambridge, MA 02139
kacquah@mit.edu

[3]NTT Service Evolution Lab
Yokosuka, Japan
chigira.hiroshi@lab.ntt.co.jp

ABSTRACT
We introduce AnnoScape, a remote collaboration system that allows users to overlay live video of the physical desktop image on a shared 3D virtual workspace to support individual and collaborative review of 2D and 3D content using hand gestures and real ink. The AnnoScape system enables distributed users to visually navigate the shared 3D virtual workspace individually or jointly by moving tangible handles; simultaneously snap into a shared viewpoint and generate a live video overlay of freehand annotations from the desktop surface onto the system's virtual viewports which can be placed spatially in the 3D data space. Finally, we present results of our preliminary user study and discuss design issues and AnnoScape's potential to facilitate effective communication during remote 3D data reviews.

Author Keywords
3D review; remote collaboration; hand-drawn annotation; video overlay.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces - Interaction styles.

INTRODUCTION
When collaboratively reviewing visual content, leaving quick annotations and using natural hand gestures such as finger pointing, often helps participants communicate effectively [2]. However, in remote settings for reviewing shared contents, it is difficult for the participants to make use of physical workspace and take full advantage of traditional freehand sketching. This constraint pertains to reviewing both 2D and 3D content in most screen/desktop sharing programs [6, 19] or groupware [3, 4]. For example, in architectural practice, a common method for leaving annotations during a remote review of 3D information includes importing renderings of digital snapshots on drawing software or printing paper plots from the 3D model [15]. In this paper, we aim to improve the remote collaborative review experience of 2D and 3D information by integrating the individual physical desktop with the virtual shared 3D workspace using spatial video overlay technique. Our system configuration enables sharing views of the digital 3D models and supports collaborative hand-drawn annotation in the virtual 3D space (Figure 1a-b). Providing a real-time shared view, tools for quick annotations on 3D digital information and a connection between the physical and the digital workspace has a great potential for remote collaboration between professionals in fields such as architectural/landscape design, planetary science or medical practice.

* The first two authors contributed equally to this work

SUI '14, October 04 - 05 2014, Hololulu, HI, USA
Copyright 2014 ACM 978-1-4503-2820-3/14/10…$15.00.
http://dx.doi.org/10.1145/2659766.2659776

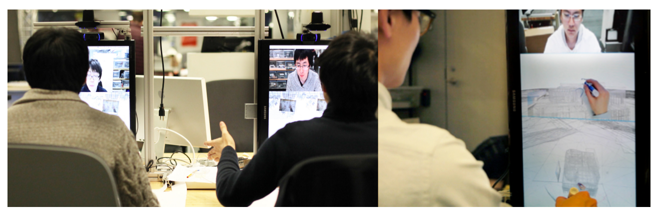

Figure 1. (a) AnnoScape prototype setup, (b) Shared 3D space.

Contributions
We propose a new system design called AnnoScape that focuses on remote collaborative review of 3D digital data using a live video overlay of the physical desktop image on the viewports of the 3D scene. AnnoScape provides the capability to merge multiple work platforms into a shared virtual 3D workspace through spatial video overlay technique. This approach allows AnnoScape system to maintain the benefits of traditional sketching tools on the physical desktop workspace, which includes the tactility and familiarity provided by the basic desktop tools. Our contributions include:

- Application of video overlay techniques for 3D review, using hand-drawn annotations and materials on the physical desktop.
- Interaction techniques to support various work modes based on viewport configurations.
- A preliminary user study on the usage of physical artifacts and freehand annotations during the collaborative review of 3D information over a distance,

to provide guiding principles for the future direction of 3D remote collaboration.

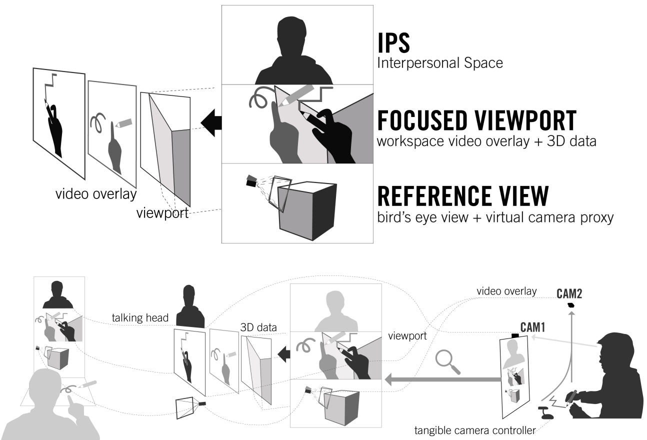

Figure 2. AnnoScape interface consists of interpersonal space (IPS), focused viewport, and reference view (Top), AnnoScape interface flow and infrastructure (Bottom).

RELATED WORK

Shared Surface Video Overlay Applications

Sharing live video of the physical workspace over distance has been explored mainly in reviewing 2D information [7, 8, 16] ClearBoard introduces the concept of seamlessly connecting remote drawing surfaces through a transparent surface. TeamWorkStation shows various workspace overlay techniques. IllumiShare [17] directly projects the remote collaborator's shared surface on the local workspace.

3D Review Using Tangible Controllers

Based on the concept of using physical artifacts as input devices [5], BUILD-IT demonstrates the use of brick-based navigation in a virtual camera in the 3D environment [14]. Other relevant projects using tangible controllers for 3d navigation include Augmented Surfaces [10] and DeskCube [12].

Freehand Drawings in 3D Space

Virtual Notepad [9] and Boom Chameleon by Tsang et al. [13] utilize a spatially aware display to define the 2D plane for digital annotation. Second Surface is a tablet-based application that uses image-based AR recognition technology to enable participants to digitally annotate 2D content in user's real environment [18].

The AnnoScape System

The key attribute of the AnnoScape system is providing tools for quick annotation that enhances the remote 3D review process. To reach this goal, our system design aims to integrate traditional freehand 2D annotations in the 3D space by combining the video overlay-based shared drawing techniques with virtual 3D viewport navigation system. The benefit of the spatial video overlay technique is that the system provides participants the freedom to utilize the individual desktop workspace. In our system, users can navigate in 3D and spatially arrange annotations in various locations from multiple angles primarily by manipulating the virtual camera proxy that is synced to the Tangible Camera Controller. This approach allows the virtual camera proxy to capture natural hand movements of the remote participant. Also, the camera proxy provides video thumbnails of the participants' live annotation session (Figure 3b). The subtle awareness of the remote participant's activity enables the users to naturally monitor the focus of the partner's attention.

DESIGN INTERACTIONS

Hardware Setup and The Interaction Design

Each desktop station consists of a monitor and two webcams connected to a computer (Figure 3a). The prototype configuration includes one webcam (Video 1) facing the user to capture the talking head and another overhead mount video source to stream the information from the physical desktop (Video 2). We place a physical Tangible Camera Controller with markers attached on the desktop that can be detected by Video 2. The interface of the AnnoScape system consists of 1) Inter-Personal Space (IPS) for face-to-face conversations [8]; 2) focused viewport that can merge into a synced workspace; and 3) a reference view for navigation, which shows the location of the focused viewports in the 3D environment from a bird's-eye view (Figure 2) (Figure 4). In AnnoScape setup, we enable users to swap between canned 3D models by having them place paper print of the model with AR markers that trigger the 3D model to change and the scene to reset and clear out the annotations. In our demo, associated contents such as a time-lapse video of the building's construction site are placed in the parallel location where the video was actually taken. Users can snap their viewports onto the 2D content's location and review both 2D and 3D information.

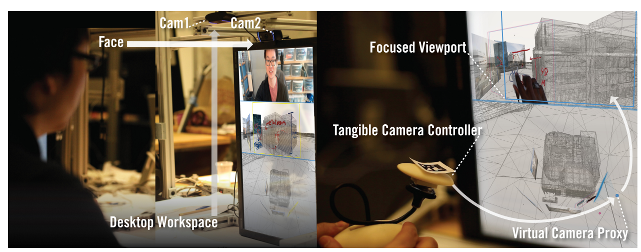

Figure 3. (a) AnnoScape system hardware configuration, (b)Tangible Camera Controller for navigation.

Navigation

The Tangible Camera Controller enables remote participants to visually navigate the shared virtual 3D space individually or jointly and share viewports either asynchronously or in real time. When the collaborator's viewport is close to an existing annotation session, the scene snaps on top of the existing workspace. After navigating in 3D to a desired scene, participants can capture the focused viewport location by pressing the capture button. In our system, capturing the personal focused viewport means locking the scene into an area for annotation and individual virtual workspace. Once the scene is locked, the overhead webcam simultaneously captures the physical individual desktop workspace, and the system overlays the extracted live video image onto the focused viewport. This allows users to place real-time annotation on the captured 3D scene from specific angles in the virtual space. When participants leave the workspace, the annotation is archived spatially in the virtual 3D space.

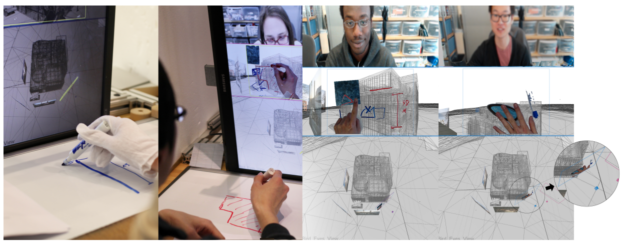

Figure 4. Various annotation methods include physical ink, finger pointing, arbitrary objects and virtual camera proxy with live video stream of the individual workspace.

Annotation

In AnnoScape setup, we use dry erase markers on a horizontal white board for annotation (Figure 4). The system automatically extracts live image of the ink, hand, and physical object from the white foreground to provide legible information. Through background subtraction image processing [1], the pixels in the white area of the streaming image are set to be transparent. For example, hands in white gloves are invisible inside the virtual scene (Figure 4). Rarely, there were times when unwanted objects in the foreground would create markings after automatic background subtraction. We implemented a manual calibration feature using one click keyboard to achieve clean background subtraction.

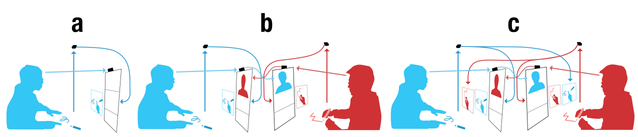

Figure 5. (a)AnnoScape individual review mode, (b) AnnoScape loosely coupled collaboration mode, (c) AnnoScape tightly coupled collaboration mode.

Work Modes

In relation to how the viewports are tied to one another, the AnnoScape system presents the following three types of work modes (Figure 5):

1) Individual Review Mode

Individual Review mode is the asynchronous 3D review session done by an individual user. The user can spatially navigate in the virtual 3D workspace, capture viewports and place annotations on multiple 3D scenes. We also allow users to hide Interpersonal Space UI elements for better view (Figure 6a).

2) Loosely Coupled Real-time Collaboration Mode

The remote collaborators can both log into the system and synchronously work on separate parts of the shared 3D data. This is called Loosely Coupled Real-time Collaboration.

3) Tightly Coupled Real-time Collaboration Mode

When all users are logged into to the system, a user can join the remote collaborator's annotation session on the same 3D scene synchronously by navigating to the activated focused viewport location. The shared viewport generates a strict WYSIWIS platform [11] that is well suited for a real-time shared review. Each participant's viewport is synchronized in the Tightly Coupled Real-time Collaboration mode. In this mode, we allow users to switch the interface between the default interface and the custom interface for the mode, which is optimized for larger view of the shared workspace (Figure 6b).

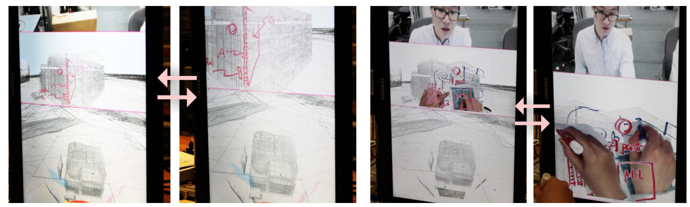

Figure 6. (a)Custom individual review interface switch, (b) tightly coupled collaboration interface switch from the default interface.

PRELIMINARY USER STUDY

We recruited eight participants including three individuals with architectural and one with industrial design backgrounds. We grouped participants into pairs in random order and gave them tasks for each study, such as 1) navigation and capturing the Focused Viewport in the Individual Review Mode, 2) annotate information on the 3D model, 3) Communicate both with annotations and hand gestures, 4) Display of Physical Materials. All participants successfully performed the given tasks. We had the participants rate the fluidity of interaction and the ease or difficulty of completion for the tasks. A two-tailed paired t-test with a significance value of $p = 0.05$ was used for questions rated 1-7. A similar sign test was conducted for binary responses of preference questions. White gloves versus no white gloves test didn't result in any major differences. The responses when the material samples were used were significantly better than not using them.

DISCUSSION AND FUTURE WORK

Based on the findings from the preliminary user study, we designed additional applications and features that could improve the AnnoScape system. Regarding the reference image for direct annotation, while options such as using overhead projection may have been a possibility; it also generated problems such as casting harsh shadows on the workspace during the annotation session. Later we learned the issue of indirectness could be drastically resolved by applying regular technical drawing tools such as a ruler to provide the reference point in the virtual scene. We included hidden features such as visually tilting or zooming the 3D view using keyboard input in current setup. We also implemented the preview mode, which allows users to transition to the archived annotation workspaces through smooth animation with simple keyboard input. As a separate application, we created tools for content-filtering options that can display captured images or associated contents with different alpha values. This was designed for scenarios with more than two users. The tool would help identify the contents based on the owners or the time in which it was created. We plan to evaluate the potential of these additional features further in future work. Although the primary driving idea of the AnnoScape is to make use of the live overlay of desktop in 3D space, we also saw great potential in a scalable multimodal system. As a proof of concept, we integrated the AnnoScape into a pair of tablet device. The mobile AnnoScape can possibly integrate a real-time live video feed from various physical locations into the 3D digital workspace. For example, the onsite construction can be streamed live inside the virtual 3D workspace in the parallel location using the camera from a tablet device. Achieving a seamless transition across a variety of work modes and enriching the representation of shared information are our long-term goals.

REFERENCES

1. Ahmed M. Elgammal, David Harwood, and Larry S. Davis. 2000. Non-parametric Model for Background Subtraction. In *Proceedings of the 6th European Conference on Computer Vision-Part II (ECCV '00)*, David Vernon (Ed.).
2. Austin Seugmin Lee. 2013. Use of Live Video Overlay on 3D Data for Distributed Collaborative Review. M.S. thesis. Dept. of Architecture. Program of Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.
3. Autodesk Revit. Retrieved May, 2013, from: http://www.autodesk.com/products/autodesk-revit-family/overview
4. Digital Project, Gehry Technology. Retrieved May, 2013, from: http://www.gehrytechnologies.com/digital-project
5. George W. Fitzmaurice, Hiroshi Ishii, and William A. S. Buxton. 1995. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '95), Irvin R. Katz, Robert Mack, Linn Marks, Mary Beth Rosson, and Jakob Nielsen (Eds.).
6. Google Hangouts. Retrieved May, 2013, from: https://www.google.com/+/learnmore/hangouts/
7. H. Ishii. 1990. TeamWorkStation: towards a seamless shared workspace. In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work* (CSCW '90).
8. Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '92), Penny Bauersfeld, John Bennett, and Gene Lynch (Eds.).
9. I. Poupyrev, N. Tomokazu, and S. Weghorst. 1998. Virtual Notepad: Handwriting in Immersive VR. In *Proceedings of the Virtual Reality Annual International Symposium* (VRAIS '98).
10. Jun Rekimoto and Masanori Saitoh. 1999. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (CHI '99).
11. M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. 1987. WYSIWIS revised: early experiences with multiuser interfaces. ACM Trans. Inf. Syst. 5, 2 (April 1987), 147-167. DOI=10.1145/27636.28056
12. Michael Glueck, Sean Anderson, and Azam Khan. 2010. DeskCube: using physical zones to select and control combinations of 3D navigation operations. In Proceedings of the 2010 Spring Simulation Multiconference (SpringSim '10). Society for Computer Simulation International, San Diego, CA, USA, Article 200 , 4 pages.
13. Michael Tsang, George W. Fitzmzurice, Gordon Kurtenbach, Azam Khan, and Bill Buxton. 2003. Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. In *ACM SIGGRAPH 2003.*
14. Morten Fjeld, Fred Voorhorst, Martin Bichsel, Kristina Lauche, Matthias Rauterberg, and Helmut Krueger. 1999. Exploring Brick-Based Navigation and Composition in an Augmented Reality. In *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing* (HUC '99)
15. Moum, A. 2010. Design Team stories, Automation in Construction vol. 19 issue 5 August, 2010. p. 554-569.
16. Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. Commun. *ACM 36, 7* (July 1993), 87-96.
17. Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. 2012. IllumiShare: sharing any surface. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (CHI '12).
18. Shunichi Kasahara, Valentin Heun, Austin S. Lee, and Hiroshi Ishii. 2012. Second surface: multi-user spatial collaboration system based on augmented reality. In *SIGGRAPH Asia 2012 Emerging Technologies* (SA '12).
19. Timbuktu(software). Retrieved May, 2013, from:http://en.wikipedia.org/wiki/Timbuktu_(software)