All Together Now: Visualizing Local and Remote Actors of Localized Activity

Scott Lederer, Jeffrey Heer
Group for User Interface Research
UC Berkeley, Berkeley, CA, 94720
{lederer, jheer}@cs.berkeley.edu

ABSTRACT

We present *All Together Now* (ATN), a tool for visualizing localized activities involving both local and remote actors. ATN presents each user with a webpage containing a common view of a shared virtual space modeled after the physical locus of the activity. Actors signal socially meaningful behavior by manipulating the spatial positions of their representations in this space. Local actors' positions are acquired automatically using computer vision. Remote actors indicate their positions with a mouse. Actors are not expressly identified. ATN exploits people's culturally established notions of spatial position to help them convey contextually relevant social cues to each other. Conveying *just enough* spatial and identity information helps optimize—without needlessly eliminating—the awareness asymmetries intrinsic to *localized distance work*.

Author Keywords

CMC, CSCW, social computing, social visualization, visualization, localized distance work, localized activity

ACM Classification Keywords

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces: Graphical User Interfaces (GUI); H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces: Collaborative computing; Computer-supported cooperative work; Synchronous interaction

INTRODUCTION

This paper presents *All Together Now* (ATN), a tool for visualizing localized activities that involve both local and remote actors. ATN gives users a common, top-down view of a shared virtual space modeled after the physical locus of the activity (Figure 1). Users convey contextually interpretable social cues through the positions of their icons in this space. This helps optimize the asymmetries between local and remote actors' awareness of each others' states.

Below we discuss the notions of localized activity and localized distance work, followed by a description of All Together Now, our tool for supporting this work, including coverage of the computer vision techniques we used to acquire local actors' spatial positions. We then place ATN into the history of research on social visualization and localized distance work. We close with ideas for extending ATN to address issues of scale, identity, and history.

HERE, THERE, AND IN BETWEEN

We designed ATN to address two specific problems. First, remote actors typically lack the sense of *place* that lends meaning to the social cues that come naturally to participants of a localized activity. Further, local actors generally lack the means for interpreting—let alone perceiving—remote actors' behavior within the cultural and architectural context afforded by the activity's place.

Localized Activity and Meaningful Spatial Positioning

Activities often occur in specific places [6]: lectures in lecture halls, planned meetings in meeting rooms, dinner in dining rooms or in front of the television. Much of the coordination work conducted as part of *localized activities* occurs through the finessing and mutual awareness of spatial position. One's spatial position at any moment can signal relevant, culturally meaningful information [3]. For example, it can reflect one's role (*e.g.*, the person at the front of the lecture hall is the lecturer, the head of the household sits at the head of the table) and it can inform interpretations of one's level of interest or engagement (*e.g.*, audience members at the rear of a lecture hall may be less interested or engaged than those at the front).

Localized Distance Work

Computer-mediated communication (CMC) technologies enable rich possibilities for *remote* participation in and observation of localized activities (*e.g.*, conference calls and online lectures, respectively), but remote actors generally lack the cues needed to establish the rich awareness that local actors enjoy. Designed and incidental cues do exist (*e.g.*, presence conveyed through instant messaging systems; background noise in a conference call), but there remains a problematic asymmetry between local

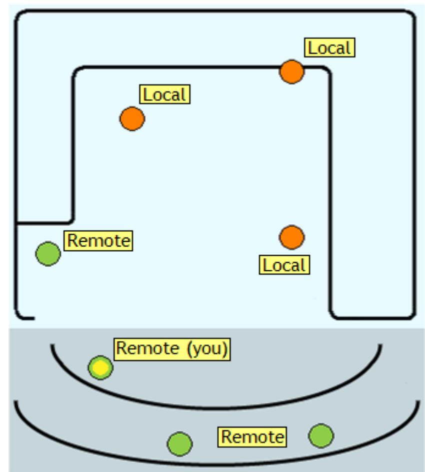

Figure 1. ATN captures a bird's-eye video feed of the physical space (left), locates people using computer vision (middle), and displays local actors' positions (orange) in a virtual space (right) shared with remote actors (green). Non-participating remote actors are placed in an observation deck. Each remote actor's circle is marked with a yellow core in his personal view. (Picture on right is annotated for grayscale printers).

and remote actors' awareness of local state when conducting *localized distance work*.

While this asymmetry can certainly be reduced, we believe it should not be eliminated. A remote actor is just that: remote. His physical distance is a basic characteristic of his role in the activity [10]; it should be managed, not overcome. While we can improve his access to physical and social context, it may be situationally inappropriate to, say, provide a full video feed, which can encroach on the privacy and comfort of physically present actors [7].

The converse asymmetry also exists: local actors often know little about the state of remote actors. How many are there? What are their roles? What are their levels of engagement? As with the other, this asymmetry might best be optimized but not eliminated, since the very remoteness of remote actors implies an obscurity that is often situationally appropriate. Each of these asymmetries might be optimized by providing *just enough* information about activity and actor state through carefully coordinated cues.

Generally, the cultural, technical, organizational, and perceptual factors at play in any given deployment will determine how much information is *just enough* to optimize these asymmetries. In any case, however, a useful design technique is to exploit people's culturally established notions of spatial positioning to convey meaningful social behavior to and from both local and remote actors.

ALL TOGETHER NOW

All Together Now addresses these awareness asymmetries by creating a shared virtual space in which all actors—local *and* remote—can use spatial positioning to convey contextually relevant social cues. Actors are represented as colored circles on a map representing the physical space that is the locus of the activity (Figure 1). This visualization is presented on a webpage available to local and remote actors. The spatial positions of local actors—represented by orange circles—are automatically acquired through computer vision. Remote actors—represented by green circles—are automatically placed into an *observation deck* upon loading the webpage.

The metaphor here is that remote actors start off *observing* but not actively *participating* in the activity. For an example of observing without participating, consider an office worker using ATN to quickly check the state of a meeting room to determine if it's available. To avoid "disrupting" a potential meeting-in-progress, she appears as an observer, not a participant.

A remote actor signals active *participation* by clicking on a suitable position in the virtual space, thereby repositioning her green circle to those coordinates. By placing her circle in a meaningful position within the virtual space, the remote user can convey interpretable social cues to other actors.

In the language of Erickson and Kellogg, ATN takes a *mimetic* approach to representing social context: it "mimic[s] the social cues being produced (often unconsciously) in face-to-face situations through an interface that is explicit" [3]. While ATN does not require deliberate interaction from local actors to mimic the social cues signaled by their positions (since the vision system tracks their positions automatically), it does require an explicit mouse-click from remote users when they select a new spatial position to signal a social cue. We feel this is a reasonable cost for the enhanced awareness ATN provides.

Lightweight Anonymity

The current prototype makes no attempt to identify actors. The vision system does not perform any biometric analyses of local actors; remote actors' IP addresses are retained for operational purposes but are not conveyed to other actors. Nonetheless, anonymity is not guaranteed. Indeed, identity is often conveyed probabilistically through social context (*e.g.*, an orange circle at the front of an ATN visualization of a lecture room during a scheduled lecture probably represents the scheduled lecturer). This lightweight anonymity contributes to ATN's strategy of conveying *just enough* information to abet optimal social awareness.

Whether ATN should explicitly convey identity is a deployment decision. In the case of our current deployment, we have found non-identification an elegant feature, one that helped motivate our use of non-biometric, vision-based motion analysis to acquire local actors' spatial positions.

Usage Scenarios

ATN makes possible a range of usage scenarios. Our current prototype is installed in the cubicle of one of the authors (Figure 1). One use of this installation is the conveyance of *presence* to remote coworkers. Members of our research group often convey presence by customizing the outgoing status message of Yahoo! Messenger (*e.g.*, "Working from home"). A similar practice emerged around our ATN installation. When the author was working away from his desk, he would place himself virtually at his desk using ATN. Remote colleagues knew that an orange circle at his desk indicated that someone (likely he) was at his desk, while a green circle implied he was away from his desk but interruptible for work-related matters.

Moving beyond the prototype installation, one can imagine other uses. An ATN installation in a lecture hall that offers the lecturer a view of the virtual space could reflect the quantity and interest level of remote students through the positions of their green circles in the audience area. This anonymous feedback could help the lecturer optimize his presentation. In another scenario, a remote observer could assess the current availability of a meeting room by loading its ATN view. By positioning his green circle at the door of the room, he could signal to the occupants that someone else is waiting to use the room. Yet another scenario could involve a family sharing dinner around the dinner table. Glancing at the ATN view on the wall, they notice the presence of a green circle in the position of an empty chair: the daughter at college studying and simultaneously—albeit virtually—sharing dinner with her family.

Architecture

The current ATN prototype has a straightforward architecture. The spatial positions of local actors are determined using computer vision to process the feed from an overhead camera (more below). This information is stored in a MySQL database and updated continuously.

Remote actors load a webpage containing a regularly updated view of the virtual space. By default, their icons are placed in the observation deck, signifying non-participating observation. If they click on the view, the click's coordinates are sent to the web server using an HTML image map, these new coordinates are stored in the database with the user's IP address, and the page is refreshed.

The view is constructed dynamically upon each page load using PHP. The program loads a static PNG image of the virtual space, acquires the coordinates of all active actors from the database, and draws appropriately colored and positioned circles onto the background image according to those coordinates. Local actors have orange circles, remote actors have green. Each remote actor's circle has a yellow center to distinguish it in his personal view.

Local actors share a single view in the physical space. This view is identical to that seen by remote actors except their circles are not distinguished by a yellow core. Each local actor can determine which orange circle represents her simply by noting its spatial alignment with her own.

Acquiring Position Using Computer Vision

The spatial coordinates of local actors are obtained by analyzing the video feed of an overhead camera installed in the ceiling, providing a birds-eye view of the physical space. Candidate humans are determined through motion analyses, which are then fed to a tracker that makes the final classifications. Most of this work was written as custom components on top of Klemmer *et al.*'s Papier-Mâché physical interface toolkit [9].

Incoming raw images are first subject to background subtraction. We initially used an "exponential forgetting" model [4], in which the background estimate is updated as a weighted combination of the current frame and the current background estimate, but found this did not produce results robust enough for our purposes. To remedy this, we augmented the background subtraction routine to force a complete update of the computed background image whenever the total motion between frames exceeds an empirically determined threshold.

The background subtracted image is then thresholded to produce a simplified bi-level (two color) image. A median filter is applied to reduce noise in the thresholded image. The connected components algorithm (already implemented by Papier-Mâché [9]) is used to identify regions in the image as candidate people. The candidate regions are additionally filtered using size and color constraints before being submitted to our tracker.

The tracker compares the positions of a frame's candidate regions to the positions of currently tracked people. If a candidate is sufficiently close to a tracked person, the tracker assumes the person has moved to that region and updates her position accordingly. The tracker assumes that people that move beyond a specified boundary (the yellow bounding box in Figure 1) are leaving the space and stops tracking them. High levels of movement can generate false candidates that erroneously pass through the tracker's filters, causing phantom orange circles to appear in the visualization. These are easy to identify, since they never move; a garbage collector removes them periodically.

RELATED WORK

ATN exists squarely in the tradition of social visualizations pioneered by Erickson, Kellogg, Donath, Viegas, and their associates. These researchers have investigated the design and deployment of socially translucent interfaces, which elegantly visualize online social activities like chats and auctions [2, 3, 12]. These projects focus on visualizing *online* activities. ATN differs by visualizing *localized* activities. We based ATN's visual design on Erickson *et al.*'s designs and suggestions (*e.g.*, all users see the same third-person view; portray actions, not interpretations; allow ambiguity and deception) [1].

Goldberg *et al.* built *Tele-Actor*, a system that lets remote people use a web browser to view and direct the actions of a local actor participating in physical activities like playing Twister or attending an awards ceremony [5]. Tele-Actor differs from ATN in that the local actor has to don special camera-equipped headgear and remote actors see a full-fidelity video feed of the local actor's perspective.

Karahalios and Dobson's *Chit Chat Club* allows remote actors to virtually inhabit an anthropomorphic robot anchored at a café table, through which they can interact with people seated at the table [8]. Paulos and Canny developed *PRoPs*, mobile anthropomorphic robots through which remote operators can participate in localized activities [11]. Like ATN, these projects allow remote actors to participate in localized activity. Unlike ATN, they rely on a physical avatar to represent the remote actor in the physical space. ATN requires no special hardware to represent remote actors; it represents all actors—local and remote—with virtual icons in a virtual space accessible through any web browser.

There is a sizable corpus of related computer vision research. To be clear, ATN does not offer any new contributions to computer vision. We believe it is unique in using computer vision to acquire the spatial positions of local actors to parameterize a mimetic social visualization.

FUTURE WORK

There are a number of possible directions to take ATN in the future. First among them involves robustness and scale. ATN is currently deployed in a specific, small location. We would like to explore the challenges of making ATN generally deployable, scaling it to larger groups and spaces, and making it more robust to high levels of movement.

As mentioned, ATN currently—and deliberately—does not identify users. While it intrinsically supports *awareness* and *visibility*—two of Erickson and Kellogg's three characteristics of socially translucent systems [3]—this non-identifiability limits awareness and inhibits *accountability*. This third characteristic could be introduced through various mechanisms, including authentication, chat, audio, and the option to transition to a full video feed.

Finally, by maintaining a record of actors' spatial positions, we could create All Together *Then*, a tool for visualizing the histories of local and remote actors' spatial positions in the course of localized activities.

CONCLUSION

We have presented *All Together Now*, a tool for visualizing localized activities involving both local and remote actors. ATN presents each user with a common view of a shared virtual space modeled after the physical locus of the activity. Actors signal socially meaningful behavior by manipulating the spatial positions of their representations in this space. Local actors' positions are acquired automatically using computer vision. Remote actors indicate their positions by clicking accordingly in the view; by default they are placed in an *observation deck*.

By exploiting people's culturally established notions of spatial positioning, ATN helps people intuitively convey contextually relevant social cues to remote co-participants. Providing *just enough* spatial and identity information helps optimize—without needlessly eliminating—the awareness asymmetries intrinsic to localized distance work.

ACKNOWLEDGMENTS

We thank Anind Dey and Scott Klemmer for their ideational and technical contributions to this project.

REFERENCES

1. Erickson, T. Designing Visualizations of Social Activity: Six Claims. *Ext. Abstracts CHI 2003*, ACM Press (2003), 846-847.
2. Erickson, T., *et al*. Social Translucence: Designing Social Infrastructures that Make Collective Activity Visible. *Communications of the ACM*, 45 (4), 40-44.
3. Erickson, T. and Kellogg, W.A. Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction*, 7 (1), 59-83.
4. Friedman, N. and Russell, S., Image Segmentation in Video Sequences: A Probabilistic Approach. *Proc. UAI 1997*, Morgan Kaufmann (1997), 175-181.
5. Goldberg, K., *et al*. Collaborative Online Teleoperation with Spatial Dynamic Voting and a Human "Tele-Actor". *Proc. ICRA 2002*, IEEE (2002), 1179-1184.
6. Harrison, S. and Dourish, P., Re-Place-ing Space: The Roles of Place and Space in Collaborative Systems. *Proc. CSCW 1996*, ACM Press (1996), 67-76.
7. Hudson, S.E. and Smith, I., Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. *Proc. CSCW 1996*, ACM Press (1996), 248-257.
8. Karahalios, K. and Dobson, K. Chit Chat Club, http://smg.media.mit.edu/projects/ChitChatClub/.
9. Klemmer, S.R., Li, J., Lin, J. and Landay, J.A. Papier-Mâché: Toolkit Support for Tangible Input. *Proc. CHI 2004*, CHI Letters 6(1).
10. Olson, G.M. and Olson, J.S. Distance Matters. *Human-Computer Interaction*, 15 (2/3). 139-178.
11. Paulos, E. and Canny, J., PRoP: Personal Roving Presence. *Proc. CHI 1998*, ACM Press (1998), 296-303.
12. Viegas, F.B. and Donath, J.S., Chat Circles. *Proc. CHI 1999*, ACM Press (1999), 9-16.