1. ABSTRACT

We can't teleport and time travel yet, but with the mediation of technology it can be possible to convey one's presence to a spatially and temporally remote location. This thesis proposes to develop and evaluate a new form of technology-mediated communication and interaction that allows people, whether officemates or strangers, to see and interact with each other outside of a time continuum by situating a representational social memory into a built environment. The memory will be materialized as a dynamic re-composition of sensory, audio, and visual information from a space that is determined to be of social significance.

Problems: How can sensor and visual information be evaluated and filtered to extract events of social significance and interest? How should these events later be visualized? Accessed? How can situating these "memories" in a space contribute to the sense of community and social history by creating new types of individual and public interaction?

2. TABLE OF CONENTS

3. MOTIVATIONS

People are compelled to document their lives and the lives of others. The artifacts we create with writing tools and cameras serve both as a record of events and as interface and stimulation for remembering and reliving them. These encapsulated experiences are generally also subjected to archival and organizational methods, giving rise to published periodicals, online blogs, and photo albums. Traditional documentation devices are individually owned and serve an autobiographical impulse, acting as a filter between that individual and the world and can thus be considered an extension of personal memory capabilities. The work I am proposing aims to assist and provide for this impulse to document by creating a new medium for memory. My design goals are focused on creating a collective memory—visual and auditory--situated in a public environment to serve both individual and communal interests, with the expectation that this will also create a new public intimacy and richer connections between people that may be both spatially and temporally remote.

Our communications and interactions with each other are increasingly technology-mediated, whether through email or video conferencing, and physical proximity and shared meeting spaces are quickly loosing their significance in standard personal encounters. Our built environments still strongly affect who occupies them and what activities happen within their space. This creates a conflict and contradiction in our use of architectural space—being neither as public or as personal, nor static or dynamic. This work is an attempt to facilitate reconciliation of this inconsistency by augmenting architectural space with the technology-aided ability to collect, store, and display visual and audio information, so that the information and history it contains contribute to its embodied identity. The augmented space becomes an evolving artifact of the community and a member of the community itself, owing to its ability to store this information created by the individual and the community and to use that information to interact with them.

4. BACKGROUND

4.1. *Memory*

Developing successful architectural designs is not just about structure and efficiency. Spaces are realized and appreciated at many different levels by generating aesthetic, social and cultural experiences. Situating memory into a built environment can likewise be experienced at multiple levels. Henri Bergson's 'crytal-image,' described

in *Matter and Memory*, illustrates the experience this work intends to create. Bergson cites the image as the currency of exchange between perception and memory and can be treated as a mode of transportation between the present (or the perceiving) and the past (or the remembering). The 'crytsal-image' does just this--fusing the past-ness of the recollected event with the present-ness of its viewing (Bergson, 1990). The 'crystal-image' in this work is to be achieved by using the same sensory information to initiate capture and recomposition of the live video as is used to activate playback so that the images seen are not so much an artifact of the past but a representation that changes as the present moves forward.

The factors that affect the function of memory, but not their biological mechanisms are at the base of the research for this situated memory. One group of such factors of particular interest, as Maurice Halbwachs argued (1992), derives from the social arena, which people always inhabit. He introduced the term 'mémoire collective' or collective memory and stressed how strongly social processes influence not only people's personal memories of their own lifetimes, but also a community's shared memories of the past (Kohli-Kunz, 1973). Going beyond Halbwachs's argument, cognitive psychologists, neurobiologists, and sociologists have recently proposed that human memory works radically differently from the traditional archive model and is in fact constructed in the human brain according to the conditions and viability of the remembering subject. (Bolles, 1998). Memory of the past, it is argued, is not only influenced but also constituted by physical and social contexts of the present. As a consequence whether or not a particular event or process remembered corresponds to the actual past becomes less important than the specific conditions under which such memory is created and constructed as well as the personal and social implications of memories held.

4.2. *Socially Responsive Environments: Responding to Presence and Action*

Sensor technologies can "perceive" human presence and activity and thus social conditions--optical proximity sensors can be used to detect how long ago your officemate left their desk, piezoelectric accelerometers can be used to detect that people always go to a certain painting first in a museum, and infrared and Galvanic Skin Response sensors can be used to detect someone's anxiety in a social situation. Developments in computer vision can also be used to evaluate the profuse amount of visual information a camera can provide in a manner analogous to the human visual system. The wide number of algorithms for object and face recognition, segmentation,

focus of attention, and motion tracking, perhaps developed for different initiatives, can be deployed to detect, track and analyze Humans and their behavior. Palimpsest (Figure 1, B. Piper + S. Agamanolis) used segmentation to detect and extract passersby in a space at Media Lab Europe from their background. Their images were then layered to form a visual that looped with a slight delay so that individuals could interact both with oneself and with other passers-by from earlier points in time that also served as a compression of the recent social happenings of the space. While the dynamic video was successful in raising the awareness of the social history of the space, it used only physical presence as the cue for image capture and response--offering minimal control to the user. Also, the response it gives does not vary with happenings of fluctuating social interest and significance, as it uses this segmentation to detect presence and not to understand how that presence might fit in the context of a more complex event or interaction.

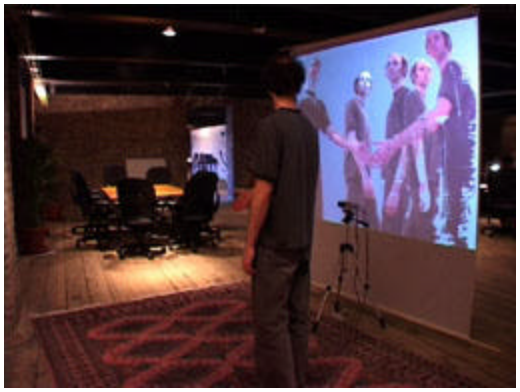


Figure 1. Human Connectedness' Palimpsest.

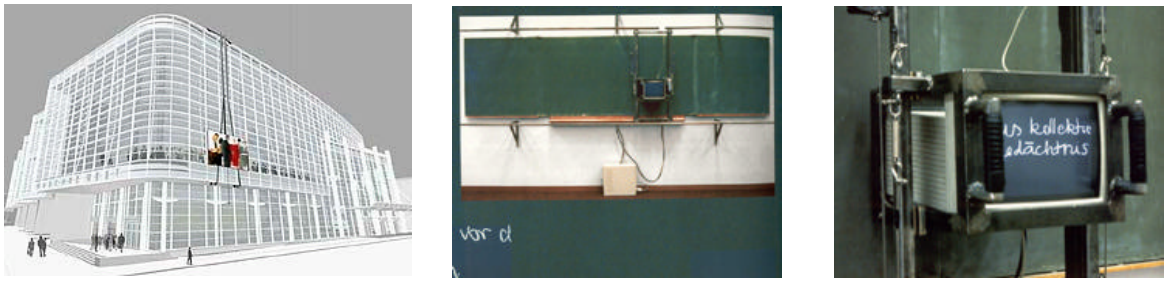
Mere detection of an activity proves insufficient to create a system that can react meaningfully and evocatively. That an environment is responsive is not enough—it needs to respond to a variety of particular activities in a particular way. That is, there needs to be a methodology for integration of these sensed elements to give the system an embodied intelligence. Cross discipline collaborations such at Smart Rooms (Darrell et al., 1994) developed at the MIT – Media Laboratory have found accomplishment using vision-based tracking to provide an unencumbered user-interface to virtual and virtually enhanced environments including a circus and theatrical performance space. The applications were focused on individual experiences and tasks, so multi-user issues were not addressed but these applications highlighted the need and developed methods for integration of real time performance, attention mechanisms, reconciliation of heterogeneous intentional behavior, and context-dependent gesture recognition. The Kids Room (Figure 2, Bobick et al.) was the first multi-person application in this vein. Vision based action recognition was coupled to a narrative, exploiting the context of the story in determining both what needed to be seen and how to see it. This project exploited the room's ability to effect the users actions—for example, enticing children to

Figure 2 View from one of the four cameras used in the implementation of Kids Room used to recognize actions on the red and green rugs.

stand on one of the three colored rugs--to simplify the computer vision problems that the multi-user, interactive, and dynamic narrative features of the project pose. But where the Kids Room and Smart Rooms were looking for activities of particular physical significance to affect a decision tree in real time, this research intends to look at physical actions for their social significance to *simultaneously* record and compose playback for one or multiple users. It also would lead to different responses to the same physical action at different points in time, where as the responses in the Kids Room (in that it was intended as a one time experience) and Smart Rooms are time-independent.

4.3. *Architectural Interfaces*

Visual accounts of events are transformed into publicly accessible, engaging, and dynamic social records in both Diller + Scofidio's *Fascimile* and Frank Fietzik's *Tafel* with their use of architectural elements as an interface. *Fascimile* is essentially a window where a live and fictionalized private space—the lobby of San Fransisco's Moscone center—becomes public by being displayed life size on a monitor suspended by a vertical armature that rides on a track along the surface of the Moscone's exterior façade. Media artist Frank Fietzek uses a similar user-controllable interface [Figures 3b and 3c] that can track and scan the surface of a smudged chalkboard and reveal text and images once present, effectively accumulating the palimpsestic chalkboard's



Figures 3a-c. 2a) Fascimile for Moscone Center 2b) Frank Fietzek's Tafel 2c) Interface for Fietzek's Tafel

information by using the images and display device as a new form of mediated recollection. The uses of these interfaces successfully merge the physical and present with the virtual and the past.

5. PROPOSED RESEARCH

5.1. *Past Work*

There are three projects that I have been a part of that have individually addressed the problems that this work intends to approach. audioPrint was an experiment in physically manifesting information about the social make up of a space. Sounds that were generated using individual characteristics of a person as they entered the space were layered on top of each other so that the ambient music was a product and indicator of who was in the room. Reflexion, borne from the Human Connectedness Group, used auditory and visual cues to compose a scene that exploited auditory and visual attention for the purposes of creating a both responsive and interactive "magic mirror" between two people in spatially remote locations. Double Exposure was to be photo booths in two separate locations (MIT – Media Lab and MLE) in which people could go in, have there picture taken and have it composed with people that had either previously been in that booth or simultaneously been in the booth in the cross Atlantic location.

5.2. *Methodology*

This work intends to be deliverable as an environment in the Media Lab's Garden and conference room created by enhancing it with projections and sound samples that are recomposed scenes and sounds captured from past events there. Simple sensor and camera information described below will serve to both construct and trigger recreation of events so that at any given point, a 'user' is both seeing and experiencing the social history and adding to it. The Isis programming language, which is specially tailored to such multimedia applications, will be used in a Linux environment with these sensors and cameras to do the image and sound capture, manipulation and playback.

There are three experimental social variables that will be investigated in order of complexity along Aaron Bobick's taxonomy of machine perception (Bobick, 1997), from presence to control to activity and action. Initial tasks will be to detect the presence, the Bobickian "atomic primitive," of individuals using proximity and motion sensors so that the information captured and conveyed is who was where when (and with whom else) and which areas have the most activity. To accommodate for the documentary and narrative side of memory the next design step will be to provide users with a physical interface that will allow them to initiate image and sound capture and "playback" of the archived footage. While presence and movements require no contextual or sequence knowledge to be recognized, and activity—which refers "to sequences of movements or

states"—requires statistics of the sequence, it is detection and recognition of action—larger scale events which typically include interaction with the environment and causal relationships that straddles the gray division between perception and cognition (Bobick, 1997) and separates this work from installations like Fascimilie and Tafel. The eventual goal for the system is to detect *select* complex activities and actions of social significance, as it is beyond the scope of this research to develop new sensory and computer vision methods for the detection of all such activities. Suitable existing methods will be appropriated to focus the space's attention on information exchanged in a group encounters, moments of laughter and heightened excitement, types of movement in and uses of the space and additions and changes to the space.

As an experiment in how the social history of the place can become a narrative unto itself, I also intend to have a viewing station situated in a location that is remote from the Garden, where both outsiders and inhabitants of the Garden can view the video of the collapsed recent past of the space. The intended form factor is a flat monitor that the user can slide along one axis to move to different points of the footage or history in the same vein as Diller + Scofidio's Facsimile [Figure 2a] and Frank Fietzek's Tafel interface [Figures 2b and 2c].

The visualization will no doubt go through much iteration. Composition and image properties (color, saturation, contrast, alpha values, noise, and blur) will be the two main variables experimented with. They will be chosen for the visual and linguistic metaphors that they carry (Lakoff, 1983), for example placing an image left in relation to another could imply it took place prior to the image on the right, alpha values affect transparency and could be used to indicate a lingering presence and fading memories, and saturation of a clip can increase at each successive viewing.

5.3. *Evaluation*

As an engineered tool this work should clearly and engagingly present information, present information of communal significance, and should be used casually and formally by everyday and occasional users with the effect of inducing new social awareness, communication, and interaction. As an object of design, it should have intuitive controls and aesthetic appeal and significance. The objective evaluation will be carried out by observing differences in individual and social patterns before and after the introduction of the system, observing the manner it is used by both new and regular users and the differences in what each gleans from it, and soliciting critique on its aesthetic impact. This type of evaluation can influence the design of such a system,

thus at least one iteration cycle will be conducted before such a final evaluation is conducted.

6. TIMELINE

1.03: Image capture, processing, and visualization methods coding.
2.03: Hardware development for sensory memory retrieval and physical interfaces.
3.03: Fabrication of physical interfaces. Design and execute first round of evaluations.
4.03: New reading based on evaluation, technical and theoretical. Begin authoring thesis document.
5.03: Refining space, interfaces, and visualization.
6.03: Temporary installations. Design and execute more evaluations.
7.03: Final revisions. Authoring of thesis document.
8.03: Authoring of thesis document

7. RESOURCES

For the realization of this project, I will need:
PC Workstations with video capture, graphics, and sound cards; Cameras, monitors and projectors; Machine shop and stock; Various sensor and electronics components; People: Mike Bove, William Mitchell, Joseph Paradiso, and the student body.
Most of these resources are readily available at the lab or are easily acquirable.

8. REFERENCES

J. Baudrillard. "The Ecstasy of Communication." In *The Anti-Aesthetic*, edited by H. Foster, The New Press, 1998.

A. Bobick. "Movement, Activity, and Action: The Role of Knowledge in the Perception of Motion." Proceedings of the Royal Society, 1997.

A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schütte, A. Wilson. "The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment." *PRESENCE* (pp. 367 – 391), 1999.

E. B. Bolles. *Remembering and Forgetting. An Inquiry into the Nature of Memory*. Walker & Co, 1988.

H. Bergson. *Matter and Memory*. Zone Books, 1990.

C. Cullinan + S. Agamanolis. "Reflexion: a responsive virtual mirror." *Proceedings of UIST*, 2002.

T. Darrell, P. Maes, B. Blumberg, A. Pentland. "A Novel Environment for Situated Vision and Behavior." *IEEE Workshop for Visual Behaviors*, CVPR, 1994.

M. Halbwachs. *On Collective Memory*. University of Chicago Press, 1992.

A. Kohli-Kunz. *Remembering and Forgetting*. Duncker & Humbolt, 1973.

G. Lakoff, M. Johnson. *Metaphors We Live By*. University of Chicago Press, 1983.

W. Mitchell. *E-topia: Urban Life, Jim-But Not as We Know It*. MIT Press, 2000.

B.Piper + S. Agamanolis. http://www.medialabeurope.org/~stefan/hc/projects

H. Schwartz. *Media Art History*. Prestel and the Media Museum ZKM, 1997.

B. Witzgall + W. Burleson. "Socially Responsive Environments". 2002.