Intelligent architectural settings using a computer vision based visual analytic interface

Eric Sauda, Chris Beorkrem, Richard Souvenir, Donna Lanclos and Scott Spurlock

University of North Carolina at Charlotte
{ejsauda, CBeorkrem, souvenir, dlancllos, sspurloc}@uncc.edu

Abstract. This paper presents a framework to enable the understanding and designing of interactive architectural settings. We present our work in interactive public displays in the lobbies of university building, demonstrating both the design and evaluative dimensions. We identify the need for a method to understand meaningful behavior in architectural settings. We then present a unique approach combining computer vision and ethnography in a visual analytic interface using the SENSING Toolkit, a computer vision framework for collecting and storing long-term, large-scale human motion, and VALSE (Visual Analytics for Large-Scale Ethnography)an interactive, visual analytic interface called designed to allow domain experts to query and understand the data. Finally, we propose a new concept of media rich spaces that we call intelligent architectural settings.

Keywords: Smart buildings, computer vision, ethnography, visual analytics.

1 Introduction

How can the new forms of computational interaction and surveillance be understood in the context of architectural design? How might such tools be evaluated relative to the intentions of the designer? How might interactive spaces be responsive to the changing and particular needs of its users?

Within architectural practice, answers to these questions have remained either anecdotal or fragmentary. Within architecture practice, the idea of a new role for digital media (cameras and monitors) as part of the space of building often included as part of the programming of a building, but little thought has been given to the way in which this may transform our understanding of the architectural setting. For example, in the case of (removed for blind review) at our university, a media wall is included in the building without any programming of either the hardware or the interaction. In other cases, architects have designed installations using interactive media, but usually with very specific and restricted form of interaction.

The (removed for blind review). research group includes faculty from architecture, computer vision, and anthropology. As part of our on-going work, we have developed methods for the analysis of media rich environments. During the course of this research, we realized the need for a more fulsome and meaningful understanding of
architectural settings, and we understood the possible role that applied ethnography and computer vision could play. This method can capture meaningful behavior in architectural settings over long time frames, and would afford designers the opportunity to study the programmatic and human-centric performance of such settings with higher levels of accuracy and assurance. It also holds the promise of unique insights for architectural design and performance.

The goal of our work is to use the VALSE (Visual Analytics for Large-Scale Ethnography) system to create intelligent architectural settings, which in contrast to existing interactive systems are programmable, adaptable and focused on meaningful behavior.

2 Evaluation of Interactive Architectural Settings

We have conducted research and design of interactive public displays over the last five years. Our most recent test bed for this work was the lobby of a new campus building, located in (removed for blind review). This building serves primarily as a combination of both office space and educational classrooms, and the lobby is a 4,000 sq. ft. space that serves as the primary circulation route from the main entrance to the elevators. The space is adjacent to a coffee shop and serves a variety of purposes including, a lounge space and venue for hosting art exhibits and corporate events.

![Fig. 1: Lobby of the (removed for blind review) (as viewed from one of nine ceiling-mounted cameras).](image)

Our beginning studies of nearly 100 installations and public information displays from fields including architecture, art and computer science led us to conclude that there was the need for a systematic method of design and evaluation. We developed such a system that included \textit{space, hardware, processing and behavior} as the four critical components.

\textit{Space} was described using standard orthographic representations of plan and
section. Since these are the best understood within the architectural community, we began with these as a standard description onto which we would map the other analysis. Our objective was to work toward a more useful and fulsome vocabulary for the description of interactive spaces that would supplement the normative representations.

*Hardware* analysis included a surprisingly wide variety of components, ranging from data projectors to flat panel displays to microphones to speakers to cameras to motion sensors. We quickly determined that it was important to understand the spatial implication of each component; for example, data projectors and LED displays both can be used for image display, but critical differences emerged from the case studies that include size of the image, sensitivity to ambient light, ability to project on floors and ceilings and sensitivity to occlusion. We developed a set of graphic icons that captured the spatial extensions of each hardware device. We were then able to map these symbols onto the floor plans and sections in a way that spatialized their effects.

![Diagram](image)

**Fig. 2** Representational diagrams of the proxemic relations of individuals and groups to the Digital Tunnel and Digital Tunnel flow chart. Upper left is a pair of sections with devices indicated. Upper right is the menu of hardware symbols. Below is a diagram of the processing logic of one installation. Note the inclusion of device symbol with the flow chart of the processing.

*Processing* covered all variable connections between the hardware, the space and behavior, typically controlled by digital devices of varying complexity and programmability. To capture the central flow of decision-making we used flow charts, a graphic device often used in computer science. We placed each hardware device
within the box of a flow chart, cross-referencing the spatial description with the flow of information. In almost all of the installations that we studied, the processing was relatively simple, often including only rudimentary motion triggers and image capture and projection.

Behavior proved to be the most difficult aspect of the design to capture. We began with an idea of “ant trails”, showing the path of each user individually of the aggregate. These quickly proved to be very difficult to display as multiple paths overlapped and obscured a clear reading. We added “heat maps” of a regular floor grid. This allowed us quickly compare the patterns of occupation generated by different installations for individual occupation and aggregated behavior over much longer periods of time.

Fig. 3: Mapping of behavior in an installation. Note “ant trails” and heat maps, as well as the qualitative evaluation based on observation and activity/gesture collection.
The objective information generated by this approach did not allow us to understand a large part of the reactions to the installations. Therefore, we decided to include a narrative of the experience of the user. We included general observations of the behavior of users as well as specific activities and gestures. While this method allowed us to record aspects that would be otherwise impossible to capture (surprise, delight, fear, anxiety), it did not provide us much help with the variability of the experience of any installation.

![Interactive installations in the lobby of (removed for blind review).](image)

Fig. 4: Interactive installations in the lobby of (removed for blind review).

Based on our study of existing installations and armed with this method for analysis, we did a series of 20 installations in the available locations. We used a combination of an array of large scale display screens, a forward facing camera, a rear mounted computer running Quartz composer or Processing to capture image sequences and provide interaction. Each of these installations was in place for at least two weeks, allowing us to study the operation of the systems and the reactions by users.

These interactive experiments identified some anticipated behavior (approaching the screen, making noise, degree of motion, etc.) and a possible response by the installation (alteration of the mirrored images, generation of diagrams of movement on the screen, etc.).

It became clear that the capture of behavior was a critical shortcoming in both our experimental work and in the field of interactive displays. The subtlety and variety of behavior we observed in the space was reduced to one or two preset responses. We needed a way to understand and use meaningful behavior in these settings, and to integrate them into the designs. In collaboration with our colleagues in the (removed for blind review), we began to develop a computer vision based method for the
collection and analysis of ethnographic data.

3 Applied Ethnography

A member of our team, (removed for blind review) staff anthropologist at the University Library at (removed for blind review), had led our ethnographic studies of the public interaction in university buildings.

The anthropological understanding of space has moved from the colonial notion that a physical location shapes culture and to the notion that there is instead a sense of place, created by certain spaces. Place is constructed from a complex interaction of everyday practices [1], memories, and imagination, and is overlain on and related to physical spaces, but also distinct from them [2-4]. Particularly prominent in anthropological studies are concerns about social structures that infuse and create public spaces. These social structures inform the ways that people encounter not only physical spaces, but also the people they find and interact with within those places [5-7]. Despite this focus, anthropologists have struggled to address the fact that human behavior is not only taking place in physical spaces, but virtual ones as well, where observation requires going beyond traditional social science field methods.

Place, then, is produced by the dwelling of people in space. Place is an artifact, a processual entity generated by the occupation, interpretation, imagination, and memory of people associated with/living within the space. The interaction of people with space gives meaning to place. Sense of place cannot exist in the absence of people; for example, home is a specific kind of place. The meanings people associate with place rise from individual interactions that are situated in larger cultural structures and identities.

Contemporary discussions within anthropology are also informed by the experiences of people who are immigrants, or otherwise displaced. Their attachment to and creation of place is often accompanied by a dislocation, or an inability to feel at home in places where they are forced to try to live their lives. In these cases, notions of place become even more separated from physical spaces, to the extent that some places only exist in the minds and memories of people, no longer to be found in the physical world at all.

Particularly prominent in anthropological eyes are concerns about the social structures that infuse and create public spaces, and so inform the ways that people encounter not just the physical spaces that underlie the place, but also the people they find and interact with within those places.

Ethnography provides an analytical method of seeing culturally and socially situated practices [8, 9]. Architectural settings become places by virtue of the interactions that take place in them, and the meanings inscribed upon those spaces in the course of those interactions. Thus, a critical step in our research process is to get at the meaning of those places and interactions.
Fig. 5. Example of typical collection of ethnographic data in library setting, using direct observation and hand annotated plans. Study includes position, motion and activity data, collected at 20 minute intervals over a four hour period.

As a background to developing the VALSE tool, we used our research in interactive environments and conducted regular observations within a university library [10, 11], thus providing qualitative descriptive data of the contexts and content of the interactions. Researchers devised structured interview questions arising from the observation data, and the cycle of interviews and observations continued throughout the project. The interview texts as well as the observation notes were analyzed and coded by the anthropologist and a graduate assistant, and the data helped inform the design of the VALSE toolkit and its adaptations, serve as ground truth to test the toolkit, and deeply analyze the impact of the technology in the spaces. In turn, we expect the deployment of the toolkit to enable researchers to examine meaningful behavior at a scale not possible without automated tools.
Intelligent architectural settings using a computer vision based visual analytic interface

Fig. 6: Example of typical collection of ethnographic data in library setting, using direct observation and hand annotated plans. Study includes position, time and use of both digital (computer, laptop, tablet, phone) and analog (books, paper, pencil) devices. Data was collected over a 24 hour period.
4 Computer Vision

We began our studies with interactive architectural settings, by extending our observational methods to include a computer vision system to supplement the ethnographic studies. This allowed us to expand our interests to include an investigation of interactive computing in public spaces, supplemented by a system capable of analyzing the ways in which users reacted and used the environment surrounding the interactive place.[27-36].

By overlapping nine different cameras within the (removed for blind review), we are able to compensate for occlusion by objects and other people within a space.

Fig. 7. Diagram showing the camera view cones for the computer vision array. By overlapping nine different cameras within the (removed for blind review), we are able to compensate for occlusion by objects and other people within a space.

After our installation, the entire lobby space can be observed from a network of nine ceiling-mounted cameras. Preliminary versions of the applications have been deployed to the displays and computers, and initial versions of the computer vision algorithms described have been implemented on the camera network. In the lobby, interactive applications were re-deployed to an eight-panel LED display with co-located microphones and speakers. We conducted small-scale ethnographic observations to confirm the content and meaning of human interaction within the lobby space. The users included students, faculty, and visitors, and the areas around these displays are open and unrestricted, affording opportunities to engage large audiences. In addition, as part of this effort, we analyzed the intersection of technology, space and behavior, continuing to produce a series of analytic diagrams both for existing installations and for our test in this space.

Two members of our team, (removed for blind review) of the (removed for blind review) have developed tools that enable our research group to capture and interpret
Intelligent architectural settings using a computer vision based visual analytic interface

ethnographic data in architectural spaces. Traditionally, video has been used in scientific research primarily for manual data collection and analysis [12, 13]. In the social sciences, automated methods for video analysis are not widely used [14].

Recent advances in computer vision have brought automated analysis of human motion within reach. We have developed the SENSING Toolkit, a framework for collecting and storing long-term, large-scale human motion. Our framework is designed to use a computer vision system, including a network of cameras installed in large indoor spaces, such as building lobbies. The system uses ray-tracing logic to triangulate locations, track movement and collect metrically accurate locations of where people have been over time, as well as identify activities that were undertaken while they occupied a location. All of this data is collected while protecting the anonymity of the user, by never recording their actual image.

Data collection is just the start, however. To facilitate analysis over very long timeframes, we have developed an interactive, visual analysis interface called VALSE (Visual Analytics for Large-Scale Ethnography), designed to allow domain experts to query and understand the data. VALSE follows a many-coordinated-views paradigm to present a user with motion summaries such as heat maps, scatter plots, and motion trails, as well as avatar-based, animated reconstructions of activity. These visualizations can be dynamically customized with user-defined time intervals for targeted analysis, and replayed using a familiar DVR-style interface. This type of tool can expand the reach of traditional social science analysis. VALSE leverages the ability of human ethnographers to recognize meaningful patterns of behavior with the computational ability to apply these insights for very extended periods of time.

Fig. 8. Diagram of the system architecture. The SENSING Toolkit will provide an API for applications to analyze and respond to activity in large spaces. VALSE is built upon the SENSING Toolkit provides tools for visualizing and annotating motions.

Large-scale analysis of human behavior in architectural settings requires an understanding of more complex activities than simple gestures. For example, the sequence of a person waving then walking towards another person, then stopping may correspond to the behavior meeting a familiar person. Methods for detecting atomic human actions (e.g., walking, waving) from video have become increasingly accurate; however, there has been less success at recognizing activities (e.g., meeting, dancing, studying).

Ontologies [18, 19], which are widely used in AI, knowledge engineering, and informatics, provide a representation of the concepts and relationships of a particular domain. In computer vision, ontologies have been used to understand hierarchical semantic relationships. However, since our ultimate goal is to develop tools for
analysis and practical inquiry/design, this type of top-down categorization could limit the possibilities of data exploration. We are implementing a data-driven approach for learning behaviors from gestures. Domain-specific data (e.g., people moving through a lobby) contains common, repeated, related motion patterns. We use the base gestures to learn higher-level behaviors using data stream mining methods. We extend the Episode Discovery algorithm[26] for finding overrepresented gesture sequences in large data collections and discover clusters of interactions that are closely related in time and apply significance testing on discovered clusters to generate sets of significant episodes based on the frequency of occurrence, length, and regularity. This data-driven approach is used to learn new behaviors that will likely correspond to a single semantic concept. We develop higher-order models by encoding the discovered sequences of gestures and continuing to build a hierarchy of behavior patterns. In addition to validation of automatically discovered behaviors, we allow investigators to use the visual analytics tools to identify motion behaviors of interest. By selecting an example of a motion or activity pattern from video, its constituent gestures can be used as the definition of a new motion motif. Collectively, these tools support the aim of having a user, untrained in computer vision, define and query for problem-specific motifs of complex behavior.

Our public lobby display test bed allowed us to test all the elements necessary for constructing an intelligent architectural setting and to coordinate their operation, but there also were significant problems with this setting. The range of behaviors contained within this space was heterogeneous and ill defined, and has made the ethnographic study of behavior difficult. The requirement that the display be the sole venue for response also limited the range of implementation strategies for the system. Based on these insights, we believe that more focused architectural settings would be more effective for future studies.

5 Toward an Intelligent Architectural Setting

The work of our group combining computer vision and ethnography into the VALSE system offers us the opportunity to understand behavior in architectural settings in a meaningful way over long time frames, including the possibility of real time understanding and response. The combination of human insight from ethnographers and the computational extension from computer vision is a powerful example of a human-computer analytic system.

How might this new ability change the way in which architectural settings are designed and analyzed?

In our work, we draw the distinction between “interactive” architectural settings and “intelligent” architectural systems. Interactive systems have a preset logic of response to the behavior of users, often limited by the capabilities of the display hardware and software. Intelligent systems have the ability to identify and understand meaningful behavior and can be programmed to respond with a wide variety of responses.

An intelligent architectural setting would allow us to understand space in several new ways;

First, we have the ability to get “hard” data on how human behavior is affected by
the design strategy employed in a particular space. Designers have often used spatial and programmatic organizational strategies to justify experimentations in architectural form. For the first time we would be able to monitor the results of these experiments and evaluate their results. This offers us the possibility to create a broad analysis of design intentions: imagine a school design that can actually be analyzed for its ability to encourage positive and engaging behavior from its students and teachers, and that can be generalized over hundreds of design examples. Particularly those architectural sub-fields that build multiple instances of a common design would be able to learn from each new design and improve subsequent instances.

We can also monitor how changes to a facility, for example a hospital waiting room, affect the behaviors of the occupants. If a hospital were to reorganize the furniture in their waiting room, does that have an impact on the way in which users occupy the space? Do the users change the space on their own? How might we use this knowledge to make more useful and comfortable settings?

Second, there is the possibility of the development of real time responsive environments, able to interpret behavior and suggest short term or immediately adaptive settings. Imagine an Alzheimer facility that is not just bricks and mortar but is able to track and understand the meaningful behaviors of the patients 24 hours a day, 365 days a year, understanding individual variation in behavior over both the short and the long term. For example, Alzheimer patients typically have trouble with spatial and visual issues, unable to distinguish between a shadow cast on the floor and a step. Intelligent architectural settings will be able to maintain surveillance and understand the changes in meaningful behavior. This could be a flexible and programmable architectural setting that could become a part of the therapeutic setting, allowing adjustments to individual patient’s therapy, creating an improved quality of life for the residents.

6 Future Work

The characteristics for potential investigations include facilities and institutions that are often designed based on previous iterations and organizational strategies. Certain building typologies fall into this category specifically, housing, hospitals and health care clinics and schools. As design has adapted to the information age, it has identified these programs as having an architectural and organizational character based upon learned observation. For example, hospitals and clinics have identified efficient floor plate strategies, which create better working environments for their employees and more effective healing spaces for their users. Based on previous examples, firms will make subtle changes to these layouts to attempt to improve upon previous strategies but they often are reusing proven strategies.

Given these primary evaluation criteria we have been working with several national firms specializing in health care and education who are potential partners for future work. We have identified several immediate targets with these firms.

Areas such a waiting rooms are usually ill-defined spaces, without any accurate way to assess their operation. Gathering data over long periods of time with ethnographic insight will allow the design to understand how these areas are used and how they might respond. It might also lead to the incorporation of physical and ubiquitous computing into these settings.
VALSE provides for an anonymous strategy for monitoring users without compromising their identities. This would allow us to capture valuable information about the operation of the facility, identifying areas of operational concern and enhancing positive behaviors. We envision scenarios where RFID tags could be linked to the VALSE data to allow for a more articulated understanding of the movements and activities of team members. This would allow for the general public to move anonymously through the system while employees are identified to provide more fulsome understanding of their movement and activities.

References

11. Lanclos, D. and L.S. Connaway. *"I find Google a lot easier than going to the library website." Imagine Ways to Innovate and Inspire Students to Use the Academic Library.*
Intelligent architectural settings using a computer vision based visual analytic interface


