Beyond the Lens:

Communicating Context through Sensing, Video, and Visualization

by

Gershon Dublon
B.S., Yale University (2008)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements of the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

Signature of Author

Program in Media Arts and Sciences
August 5, 2010

Certified by

Joseph A. Paradiso
Associate Professor of Media Arts and Sciences
MIT Media Laboratory
Thesis Supervisor

Signature of Author

Mitchel Resnick
LEGO Papert Professor of Learning Research
Academic Head
Program in Media Arts and Sciences

Beyond the Lens:
Communicating Context through Sensing, Video, and Visualization

by

Gershon Dublon

Submitted to the Program in Media Arts and Sciences
on August 5, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Responding to rapid growth in sensor network deployments that outpaces research efforts to understand or relate the new data streams, this thesis presents a collection of interfaces to sensor network data that encourage open-ended browsing while emphasizing saliency of representation. These interfaces interpret, visualize, and communicate context from sensors, through control panels and virtual environments that synthesize multimodal sensor data into interactive visualizations. This work extends previous efforts in cross-reality to incorporate augmented video as well as complex interactive animations, making use of sensor fusion to saliently represent contextual information to users in a variety of application domains, from building information management to real-time risk assessment to personal privacy. Three applications were developed as part of this work and are discussed here: DoppelLab, an immersive, cross-reality browsing environment for sensor network data; Flurry, an installation that composites video from multiple sources throughout a building in real time, to create an interactive and incorporative view of activity; and Tracking Risk with Ubiquitous Smart Sensing (TRUSS), an ongoing research effort aimed at applying real-time sensing, sensor fusion, and interactive visual analytic interfaces to construction site safety and decision support. Another project in active development, called the Disappearing Act, allows users to remove themselves from a set of live video streams using wearable sensor tags. Though these examples may seem disconnected, they share underlying technologies and research developments, as well as a common set of design principles, which are elucidated in this thesis. Building on developments in sensor networks, computer vision, and graphics, this work aims to create interfaces and visualizations that fuse perspectives, broaden contextual understanding, and encourage exploration of real-time sensor network data.

Thesis Supervisor: Joseph A. Paradiso
Title: Associate Professor of Media Arts and Sciences, MIT Media Lab

Beyond the Lens:

Communicating Context through Sensing, Video, and Visualization

by

Gershon Dublon

The following people served as readers for this thesis:

Thesis Reader

Hiroshi Ishii
Jerome B. Wiesner Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Thesis Reader

Frédo Durand
Associate Professor of Electrical Engineering and Computer Science
Department of Electrical Engineering and Computer Science

Acknowledgements

To my advisor **Joseph A. Paradiso**, for supporting the group with all his energy, for his continuing trust, and for sharing his passion for music.

To my readers **Hiroshi Ishii** and **Frédo Durand**, for their constructive critique and support in this work.

To the **Responsive Environments Group**: **Nan-Wei Gong**, for being my ukulele soul-mate, and for being ever-critical and always constructive; **Laurel Pardue**, for being the best friend and ally I could hope for; **Nan Zhao**, for parachuting in from Germany, and being a star; **Bo Morgan**, for being love-struck, with everything and everyone; **Matt Aldrich**, for his refreshing sanity and clear thinking; **Michael Lapinski**, for fighting with me sometimes but always making up in the end; **Brian Mayton**, for being the most capable, patient, and friendly group-mate to everyone; **Mat Laibowitz**, for being my on again, off again west coast pen pal; **Nicholas Joliat**, for being so creative, supportive, and witty; **Noah Swartz**, for walking in one day, ready to learn, build, and teach others; **Mark Feldmeier**, for all his creative uses for junk; **Sebastian Palacios**, for being there to help and learn in any way he can; and **Amna Carreiro**, for being there for everyone, through thick and thin.

Thanks twice, to my close collaborators **Brian Mayton** and **Laurel Pardue**.

To our collaborators across campus: **Stephen Miles**, **Hiroshi Ishii**, and the **Future ENEL Team**.

To my collaborators: **Nathalie van Bockstaele**, **Clovis Thielhaber**, and **Jackie (Chia-Hsun) Lee**, for always challenging me to look deeper, and to **Wendy Jacob** for forming the collaboration, shaping my thinking, and supporting our work.

To **Greg Tucker**, **Linda Peterson** and **Aaron Solle**, for making everything happen.

To UROPs **Anisha Jethwani**, **Patrick Hurst**, **Jeffrey Prouty**, **Turner Bohlen**, and **Tanya Liu**, for their energy and creativity.

To **Peter J. Kindlmann** for his support and guidance, and to **Andreas Savvides**, for so generously including me in his group, and for launching me onwards.

To **Liz Carlin**, **Edwina Portocarrero**, **Pol Pla i Conesa**, **Alex Charrow**, **Tina Zavitsanos**, **Aliza Shvarts**, and **Michael Geraci**; and to my family: **Dina**, **Giora**, and **Amalle**.

To **Johannes Somary** (1935-2011), artist, teacher, friend, and mentor.

Contents

1 | Introduction

With instant speed the causes of things began to emerge to awareness again, as they had not done with things in sequence (in the mechanical age).

Marshall McLuhan, "Understanding Media," 1964 [1]

1.1 From Data to Knowledge, and Back

Our environments have been filled with sensor networks designed to provide specific data and solve pre-determined problems, but information from embedded sensors remains, for the most part, siloed in closed-loop control systems and out of reach of the majority of users. In a technology-driven rush to collect new and ever more streams of information, sensor networks have been deployed so quickly and broadly that users, be they consumers monitoring their personal energy usage or corporations tracking their global supply chains, can barely manage the data influx, much less know how to interpret the data. Moreover, rapid developments in sensor networks research are enabling real-time, mobile and distributed sensing, bringing these technologies into new domains, from homes and offices to construction sites. In many cases, workflows do not easily adapt to the changes and users face challenges integrating real-time information into existing practice.

At the same time, there is a prevailing emphasis within the sensor networks research community on a one-way "from-data-to-knowledge" paradigm, where information from multiple sensing modalities are fused together in a hidden layer to produce high-level inferences. While this approach can be extremely useful, the underlying statistical models require a posteriori knowledge, which in many cases is lacking or non-existent for new deployments. When such models are available, this approach tends to reduce data to single dimensions of interpretation—"a user is consuming too much energy," for example. In addition, this knowledge is often separated from other data streams that may provide valuable context, or even produce differing results if examined together.

What if we could create user interfaces to the hidden world of densely distributed, networked sensors? Relatively little attention has been paid to facilitating open-ended and convergent exploration of these sensor networks, though many new and compelling results may stem from correspondences that span across networks, times, and spaces. Recent efforts to approach this problem theorized cross-reality environments, where ubiquitous sensor networks would interact with pervasively shared virtual worlds. These efforts focused on the point of exchange between physical and virtual environments, and hinged on the concept of massively multiplayer virtual worlds, like Second Life and Sun Microsystem's Project Wonderland. Many of the challenges faced by this research were particular to that concept of pervasive sharing in virtual worlds, chief among them the vacancy problem, which resulted from a combination of asynchronous usage and typically low levels of user interest. In addition, the research was limited by the technological and graphical shortcomings of the specific proprietary platforms to which it was tied.

This thesis focuses on encouraging and enriching individual users' experience of sensor network data. Motivated by growing numbers of sensor deployments and increasing access to previously siloed networks, this thesis lays the groundwork for exploratory modes of user interaction with sensor networks and the data they produce, and seeks interfaces that support this kind of interaction. By enabling users to move fluidly between convergent data streams and the higher-level inferences that fuse them together, this work aims to develop sensor network user interfaces that foster better understanding of sensor-driven context and the data that produces it.

To that end, this thesis documents research and development towards this goal, from sensor deployment to interface design. This work is spread across a set of separate, but related projects, taking a variety of forms and applying to a broad set of domains: first, an installation that composites multiple video sources into a single, salient view that is fed back to portals distributed throughout the environment; second, the introduction of real-time, mobile sensor network data and augmented video to a new domain, construction site safety; and third, a virtual environment that scalably collocates representations of dense, multimodal sensor data, applied to building information management. In addition, an ongoing project described here fuses wearable fine-grained location sensors with cameras to better interpret and filter contextual information for user privacy. These projects culminate in sensor network interfaces that interpret complex, contextual knowledge, like the activity taking place in a building or the state of the environment on a construction site, and present that information to users for exploration and interaction, through a combination of immersive 3-d visualization, animation, and augmented video.

Although these projects take very different forms, as a set of case studies they share a common set of system and interface design principles, which are elucidated in this thesis. While this thesis takes time to motivate each individual application, the larger aim is to analyze each as an instructional example of a different way of conceiving of new user interaction with sensor network data. From a systems perspective, the applications sense, interpret, and communicate information that relates people and their environments. These tasks involve integrating systems of wearable devices and embedded sensing infrastructure with servers and database architectures that can together accommodate and analyze heterogeneous data streams. This work explores strategies for processing and serving data from multiple sources in real-time.

At a higher level, each user interface makes use of the unique affordances of its medium, aiming to produce concise representations of contextual information from sensors through visual metaphors equipped by those affordances. In a game engine, for example, they might include first-person perspective, 3-d visualization, physics, lighting, and interactive animation, among others. In their designs, these interfaces leverage such affordances to relate and combine the perspectives provided by the data streams on which they operate. Although each interface organizes and presents sensor data differently, all make sensor fusion central to the user experience, encouraging users to see and explore relationships across sensing modalities. These inferential visualizations expose the process of the fusion that produces them, while layers of interactivity provide a means for exploration of this information. Finally, this thesis raises questions about how sensor data might be organized and presented to users in light of its goals, and looks for answers in discussion, preliminary user evaluation, and expert interviews.

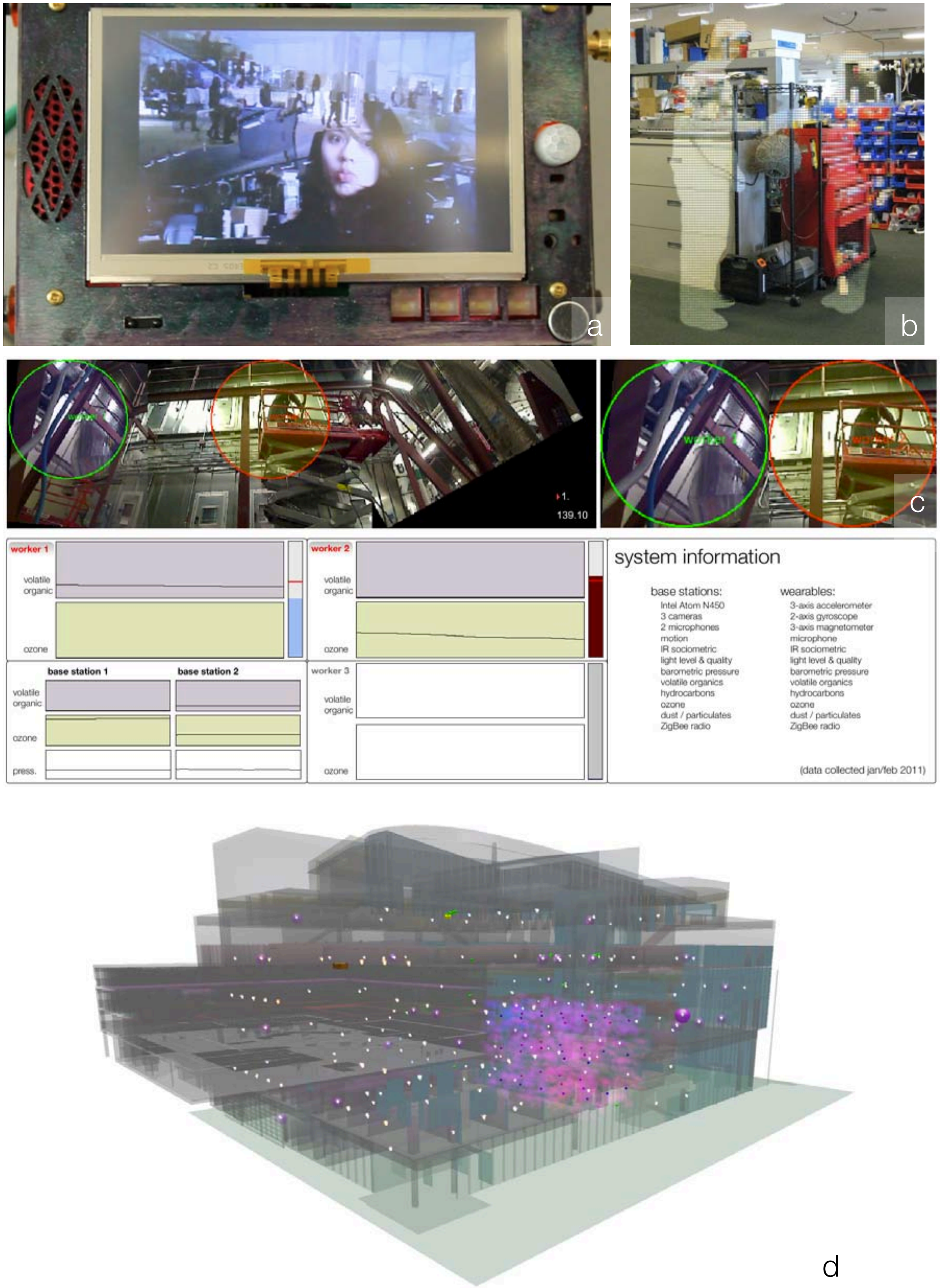
Building on documentation of these three projects and discussion of their common system and interface design goals, this thesis proposes a framework for thinking about user interfaces that support and encourage exploration of dense sensor networks, and begins to map the vast landscape of this application space through the examples given, as well as past work. More broadly, in a world of increasingly pervasive sensing and sharing, this thesis seeks to trouble the role of sensor networks as simple tools of either surveillance or closed-loop control by advancing technologies that support open-ended exploration. By encouraging users to move fluidly between convergent data streams and the higher-level inferences that fuse them together, this thesis seeks to foster better understanding of sensor-driven context and the data that drives it.

1.3 Thesis Goals and Outline

This thesis develops sensor network user interfaces that interpret, visualize, and communicate context, and proposes a common framework for their design that supports open-ended exploration while emphasizing saliency of representation. A number of interface examples are presented in detail, each targeted at a very different application—interaction with real-time video through a portal, construction site safety, personal video privacy, and open-ended, cross-reality browsing of dense sensor network data. The first examples hinge on video as the key channel for communicating sensor data, fusing distributed sensors and cameras to augment video streams with contextual information derived from the sensors. The last example, called DoppelLab, has a broader reach, and attempts to intuitively and scalably collocate visual representations of dense, multimodal sensor network data. In this document, each project is made to stand on its own in terms of application and motivation, but ideas and designs are developed, evolved, and shared across all.

This thesis builds on a diverse body of related work spanning sensor networks and sensor fusion, cross-reality, visual analytics, augmented and structured video, and sensor-network user interfaces. Chapter two situates this thesis within the related fields, and reviews the literature for insight. Presentation of the examples is split into several chapters—chapter three introduces the camera-wearable fusion and video augmentation interfaces, and chapter four the cross-reality browser. Chapter five draws from the previous chapters to propose a common design framework and set of goals for these applications, generalizing from the specific examples to a conceptual common ground. Chapter six details the user studies and evaluation done with expert professionals in the construction and building management industries. Chapter seven details ongoing work and future plans, and chapter eight looks ahead to a new front in cross-reality interfaces.

The systems presented in this document are the result of a number of close collaborations, both within the Responsive Environments Group at the MIT Media Lab and across the Institute. Wearable sensing hardware for construction workers was developed by Brian Mayton as part of a larger collaboration between the Responsive Environments Group, the Tangible Media Group, and the Senseable City Lab at MIT. DoppelLab is a joint effort with Laurel S. Pardue and several undergraduate researchers.



system information

base stations:
Intel Atom N450
3 cameras
2 microphones
motion
IR sociometric
light level & quality
barometric pressure
volatile organics
hydrocarbons
ozone
dust / particulates
ZigBee radio

wearables:
3-axis accelerometer
2-axis gyroscope
3-axis magnetometer
microphone
IR sociometric
light level & quality
barometric pressure
volatile organics
hydrocarbons
ozone
dust / particulates
ZigBee radio

(data collected jan/feb 2011)

Figure 1: Communicating context from sensors through interactive video and visualization. (a) Flurry; (b) Disappearing Act; (c) TRUSS for Safety; (d) DoppelLab

2 | Related Work

This thesis develops systems that integrate pre-existing and custom-designed sensor network systems with new user interfaces to multimodal sensor data, and combines information from multiple sensors to produce salient representations of contextual information in the form of augmented video as well as 3-d visualization and animation. This work integrates literature from a wide variety of related fields that engage sensor networks, including human-sensing, sensor fusion (in particular fusion of cameras and distributed mobile devices), cross-reality, visual analytics, augmented and structured video, and sensor-network user interfaces. Further chapters review some of this literature as it relates to the individual applications described within each one. This chapter introduces selected work from those fields, providing a high-level overview of prior work relevant to the larger thesis.

2.1 Human-sensing and Sensor Fusion

Human-sensing describes a broad class of problems relating to the application of indirect sensing to the extraction of spatiotemporal properties of people as they move through their environments. This section briefly touches on human-sensing challenges, and relates them to the larger problem of sensor fusion. Teixeira and Dublon provide a comprehensive, multi-disciplinary survey of techniques for human-sensing in [2], covering work spanning detection of presence, counting, extraction of indoor location, tracking, and identity. Figure 2, excepted from [2], illustrates these spatiotemporal properties and the traits that can be used to infer them.

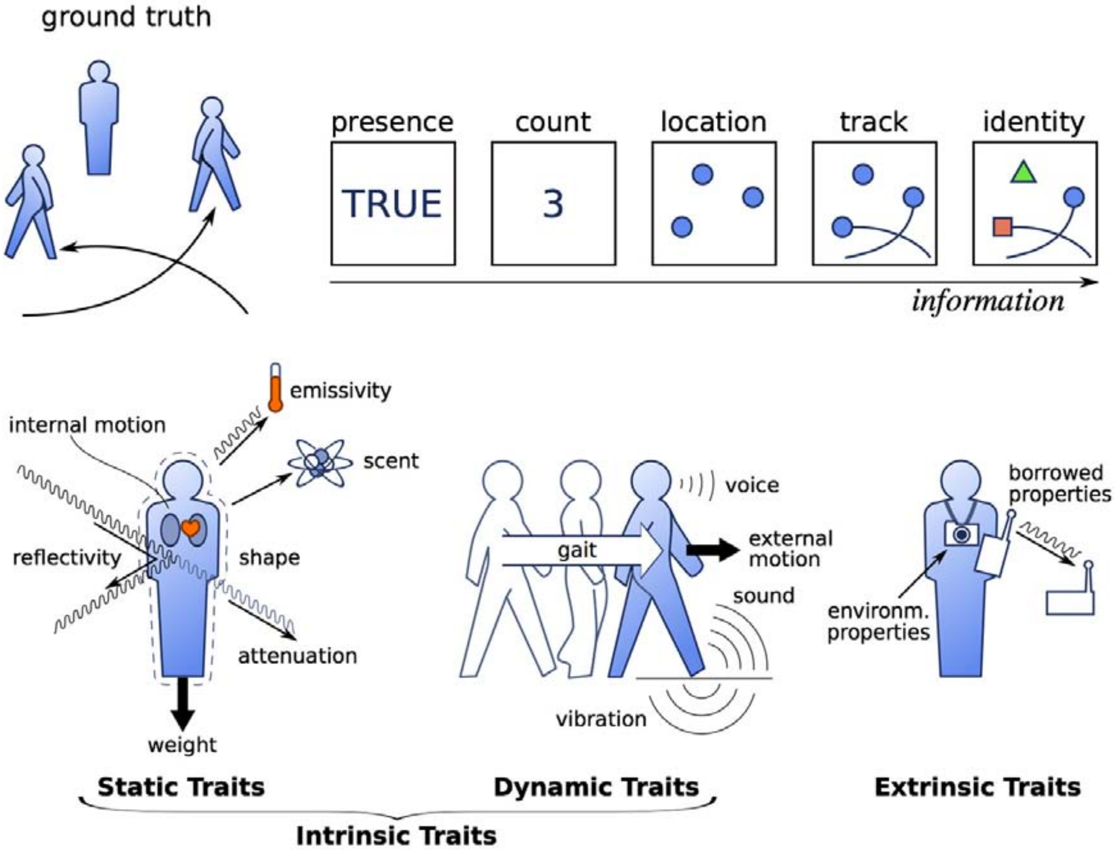


Figure 2: Top: human-sensing goals; bottom: traits linked to a variety of sensing modalities (From A Survey of Human-Sensing, Teixeira, et al.)

Approaches to the problem are split between those that instrument people with physical sensors, like accelerometers or radio-location tags, and those that rely purely on sensors in the infrastructure, like cameras and microphones. When instrumentation is a possibility, fine-grained radio-location using ultra-wide-band radios and time-of-flight ranging best suits all five spatiotemporal properties; however, this approach requires significant radio infrastructure, facilities for distributing and managing tags, and a willingness on the part of users to carry the devices. Still, the technology is advancing, and costs are dropping rapidly.

Video is the most effective infrastructural sensing modality covering these properties. Recent approaches to detection extract specialized features, such as

SIFT (scale-invariant feature transform) [3] or HoG (histogram of oriented gradients) [4], and perform classification using support vector machines or other methods. These features have significantly improved performance of vision-based detection and tracking. Still, the performance of most algorithms tends to decrease dramatically when the environment is cluttered or changing (e.g. inconstant light, frequent occlusions, or large numbers of people). This shortcoming has prompted researchers to develop increasingly complex models and methods that can work under a variety of conditions. Even so, despite significant advances in research, and admirable tracking performance when people stay within the field of view of a camera, vision alone cannot reliably count people under many real-world conditions, much less consistently identify them as they move through space.

2.1.1 Sensor Fusion

Sensor fusion algorithms combine multiple sensors or sensing modalities to produce inferences, leveraging the emergent benefits of the data fusion for improved performance. Fusion techniques can be used to build higher-level features that capture the affordances of each modality or avoid pitfalls associated with each one. In this thesis, the term sensor fusion is used broadly to describe the combination of multiple sensing modalities towards some better result than either could have independently produced. Under this definition, a good deal of applied sensor networks research makes use of fusion, in particular for activity recognition but also for human-sensing and other general sensing problems. The projects in this thesis aim to apply sensor fusion to produce intuitive interfaces to sensor data.

There are many approaches to sensor fusion, ranging from relatively simple flowcharts, confidence weighted averaging [5], or distributed average consensus methods [6] to others that employ Bayesian networks for the data fusion, applying

Kalman or particle filters, or building hidden Markov models. This section reviews several papers that apply fusion to problems related to the larger thesis.

Lukowicz, et al. develop a system for recognizing tasks like sawing, hammering, and turning a screw, among others in a wood workshop using a combination of wearable microphones and accelerometers mounted on a user's arm [7], [8]. The authors assume that in their application, loud sounds detected near a user's hand indicate shop activity, and use the difference in intensity between the hand and the upper arm to segment relevant activity from the background. Their system then applies linear discriminant analysis (LDA) to the frequency spectrum (FFT) of the audio data and classifies segments by their Euclidean distance to the mean of each class in the reduced feature space. At the same time, the system extracts features like peak counts within a time window and mean peak amplitude from the accelerometer data, as well as a single axis of raw data from each accelerometer, and applies hidden Markov models for classification. The authors test a number of methods for fusing these classifications: a simple comparison of top choices, which designates as valid only those classifications that agree; and class rankings, which examine each classifier's per-class confidence. Several class ranking methods are tested, and the authors report the best results for class ranking using logistic regression.

Camera-wearable sensor fusion is motivated by the premise that significant algorithmic challenges and computational costs of parsing video can be mitigated by the labeling of tracked objects with lower-dimensional and ID-linked sensor data that are synchronized with the video capture. Candidates for this fusion include inertial measurement units (IMU) that carry some combination of accelerometers, gyroscopes, and magnetometers, or any other sensors that can capture some shared state also available to a camera.

Teixeira, et al. show in [9] that people carrying wearable accelerometers can, under certain conditions, be reliably linked to their corresponding camera tracks by matching gait timing parameters between the accelerometers and the camera. Heel-strike and mid-swing features are extracted from the accelerometer using the extrema of the vertical acceleration of the walker, and from the camera using the extrema of the standard deviation of the walker's foreground blob, in the direction of the ground plane. The authors achieve average recognition rates of over 85% in simulations of 10 people in a scene, even as subjects leave and re-enter the field of view. However, this strategy fails when subject are not walking, and when gait timing cannot be realistically extracted from a camera, which is often the case as a result of occlusions or non-ideal camera angles.

In [10], the authors generalize this idea, forming a hidden Markov model for each person consisting of a measurement from the inertial sensor and a measurement from the camera and applying maximum a posteriori estimation to generate the most likely matches. On the camera side, acceleration magnitude and yaw are computed from the raw tracks; on each person, the magnetometer supplies the direction of motion, and the accelerometer is used to discover whether the person is walking or stopped (classified by the amount of vertical bobbing seen by the accelerometer). The authors report remarkably good tracking results in simulation that juxtaposes multiple, real tests using a dense network of overhead cameras and dedicated IMUs.

In [11], Laibowitz, et al. apply multi-sensor techniques towards the automated creation of personalized documentary video. Their system, called SPINNER, labels feeds from many distributed cameras throughout an environment with synchronized data and audio gathered from wrist- and lapel-worn mobile sensor devices, and uses those data to select and edit video clips in response to user queries. The automated annotation of the video covers subjects' identities, activities, and social dynamics, enabling queries that hinge on those properties without requiring the intractable amount of manual labor that would be needed to manage a system that multiplies every hour of video by the number of nodes. While the system does not strictly perform sensor fusion with the video, the synchronized annotation is sufficient to extract those properties for the application. However, the system does not label individual pixels in the video—only the clips themselves. Some application of fusion to associate wearable sensor data with individual actors, along the lines of [9] or [10], could enable much finer-grained editing and selection, as well as automated video augmentation or other pixel-level, actor-specific manipulation.

2.2 Cross-reality

Paradiso coin the term *cross-reality* in [12], to describe the mixed reality environment that results from the fusion of densely distributed sensing and pervasively shared virtual worlds, like Second Life. Cross-reality is distinguished from Milgram and Kishino's classical taxonomy of mixed realities [13] by its seamless bridging of the physical and virtual worlds through ubiquitous sensor and actuator networks, to form a *dual reality* [14], in which the these worlds meet. The two worlds are punctuated by so-called "wormholes" through which information and actuation bi-directionally tunnel, making use of available interfaces, from video portals to smart plugs [15].

In [15], Lifton, et al. develop this idea through several examples that cross a network of sensor-laden, web-controllable smart plugs into Second Life. In their *ShadowLab* project, a floor-plan of the MIT Media Lab is augmented with visualizations of data from this network. In another example, the authors cross audio, video, and sensor data from a network of media portals [16] into Second Life, and creatively leverage the 3-d world to map time to space—allowing users to "walk" from the present moment back through time in a video. Consistent with their focus on shared worlds, the authors develop animated metamorphoses of the player-controlled avatars that respond to sensor data. Other work, like [17], has focused on supporting collaboration in these environments. This thesis presents works that builds on cross-reality, but focuses on individual users' experiences of sensor network data through interactive visualization in virtual environments.

2.3 Interactive Information Visualization

The fields of information visualization (Infovis) and interaction design are closely linked, though research efforts have generally treated them as separate endeavors. The literature of Infovis has tended to focus on building and cataloging salient visual representations of data, evaluated through a kind of "cognitive information economics" that measure the amount of information conveyed against the cognitive load associated with the observation [18]. While the separation between representation and interaction has been prevalent in the literature, there are notable exceptions, and increasingly so. Over a decade ago, Woodruff, et al. developed a framework for the use of *zoomable* user interfaces to would present visualizations with constant information density across a variety of scales [19]. Even earlier, Bederson, et al. conceptualized a zoomable ("stretchable") graphical sketchpad on which all the information stored on a computer would be organized and represented at different scales [20].

More recently, an offshoot of the Infovis community has developed a field of *visual analytics* to represent interactive visual interfaces that support analytical reasoning and decision-making. Visual analytic designs aim to transmit large data sets concisely, enable comparison between multiple streams of information, and encourage user interaction with visualizations [21]. These approaches apply strategies such as dynamic highlighting and obfuscation of subsets of visualized data to support hypotheses or call attention to outliers in a data set. Users choose between modular representations, assess the efficacy of various designs, and aid in the fabrication of new designs. This ease of interaction facilitates feedback workflows from design to inference and back.

In [22], Yi, et al. explore the role of interaction in Infovis, surveying the field to build a taxonomy of low-level interaction techniques for visual analytics, as well as a collection of higher-level interaction categories:

- Select: mark something as interesting
- Explore: show me something else
- Reconfigure: show me a different arrangement
- Encode: show me a different representation
- Abstract/Elaborate: show me more or less detail
- Filter: show me something conditionally
- Connect: show me related items

Yi, et al. [22]

The authors catalog users' generic intents in visual analytic applications, served by a variety of specific visualization strategies from the literature, such as interactive pie charts and semantic zooming, which, though quite different in form, both serve towards *elaboration*. These specific strategies can then be incorporated into a toolbox for interaction design in these applications.

2.3.1 Information Visualization using Game-engines

Outside of cross-reality, a small number of researchers have examined the use of game engines for interactive information visualization. [23] provides a clear introduction to the use of game engines for scientific research, outlining the modular system architectures that can support realism in simulation when precision is not critical, but exploration may be helpful.

Brown-Simmons, et al. use game engine affordances to encourage exploration of environmental data and communicate a specific point of view [24], advancing an artistic and scientific framework for thinking about game engines in this context. Building on this work, [25] catalogs a small set of particle emitters for the visualization of Earth science data in a game engine, towards tools for education. Kot, et al. apply a game engine to information visualization, in [26], but focus on gaming-derived metaphors such as shooting, limiting their scope. In general, these efforts have focused on applications of game engines to specific simulations or applications; in contrast this thesis presents a game-engine based tool for open-ended exploration and development.

2.4 Augmented Video

Augmented reality (AR) spans a large field of work beginning in the early 1990s, in which a view into the physical world is mediated by or annotated by data from outside the original sensory channel constituting the view. A sub-field of augmented reality, augmented video renders layers of information from the physical world in the image domain, often on top of a live camera view; a broad collection of work falls into this category, and several examples are included here. Bimber and Raskar treat the subject thoroughly in [27].

Early on, researchers theorized head-mounted displays that would augment the user’s view with knowledge about the world. Feiner, et al. looked at ways of delivering knowledge about maintenance and repair processes to users wearing such displays, tracking users’ activities and adjusting the rendered plans accordingly [28]. Computer vision and graphics play a central role in AR research. In [29], Mann, et al. develop diminished reality, which selectively removes objects from a user’s camera-mediated field of view. Herling and Broll track user-selected objects, remove them, and perform image completion using patches from the surrounding texture to fill in the gap [30]. In [31], Sebe, et al. track objects and insert them into a 3-d environment for a surveillance application, applying segmented blobs as textures to polygons in the virtual space.

Structured video encompasses work in which some axes of structural context from video are meaningfully encoded in the image domain through augmentation, most often for compression or human consumption. This structural context is extracted using vision techniques like motion estimation and segmentation, and includes information like scene content and camera motion [32]. This information is then relayed to users through interactive interfaces that make use of video augmentation or other manipulation. In [33], Elliott constructs an augmented spatiotemporal volume from the video that reveals information about its content on its side; the viewer shows automatically detected scene cuts alongside a color frequency histogram for each frame. In [34], Teodosio and Bender define a class of images called “Salient Stills” which capture and convey information from across times in single, still images. In [35], Correa, et al. develop interactive, dynamic video narratives, which extend the concept of salient stills to moving panoramic video that tracks one or more actors to form salient video.

2.5 Sensor Network User Interfaces

There have been a number of recent efforts to create general purpose sensor network user interfaces. Lifton, et al. developed the Tricorder, a location- and orientation-aware handheld wireless sensor network navigator and data browser, based around an early smartphone [36]. Later work brought about the Ubicorder, a tablet-based graphical user interface (GUI) for sensor data browsing as well as defining inference rules [37], [38]. Both systems featured a graphically-augmented

2-d floor plan on which sensor data is illustrated. Other work has focused in particular on user interfaces for defining inference rules on data streams, either through GUIs or scripting [39], [40]. Several commercial enterprises have developed user interfaces to sensor data, for consumer applications ranging from home energy monitoring [41], [42] to personal fitness [43]. These tend to focus on single modalities and center on 2-d graphical representations that can become quite complex as networks grow.

3 | Sensor-driven Augmented Video

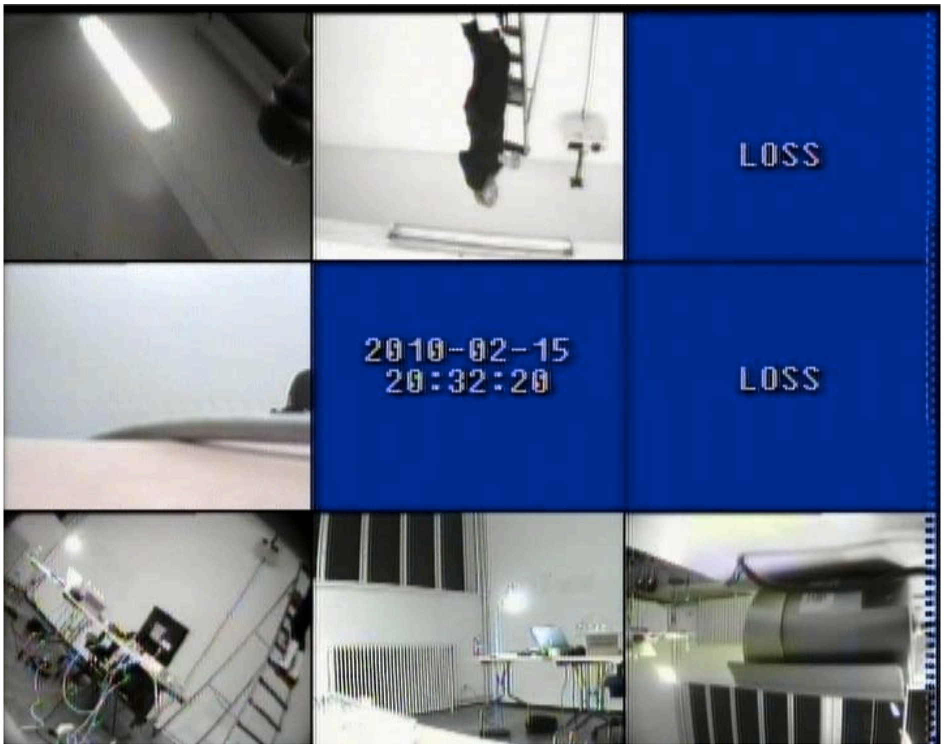


Figure 3: Multiple camera views, organized as adjacent rectangular windows in the style of most video surveillance interfaces. [Image credit: Scott Fitzgerald, February 15, 2010 via Flickr, Creative Commons Attribution]

3.1 Introduction and Motivation

Streaming video is by far the most ubiquitous and effective medium for high-bandwidth communication of contextual information to remote users, but most interfaces to live video in use today, like the one depicted in Figure 3, remain woefully antiquated. At the same time, there are limits to the reach of a camera, as well as to the information a viewer can realistically extract from many separated, adjacent views. Outside the application of video surveillance, which makes extensive use of the split format and which this thesis does not consider, this section investigates ways of creating more compelling and concise user interfaces to remote context through video.

As noted in chapter two, a number of research efforts over nearly two decades, like [34] and [35], have originated ways of collapsing time in video to create concise representations of complex movie sequences, composed of still images or short video summaries. In contrast, this section develops sensor-driven techniques for composing concise, interactive visualizations of multiple, concurrent video streams. This work is concerned with the relationships between camera perspectives and the search for salient details in each view, or across multiple. How can distributed sensing being used to estimate what the camera cannot see, or what is happening just outside the field of view? How are the cameras situated? Finally, how can a user be transformed from passive observer into involved participant?

Questions like these are inspired by artwork like the paintings in Figure 4. In Magritte's "The Blank Cheque," layers of depth are interwoven to link spaces and viewpoints, rendering foreground objects partially transparent. The oddly compelling cognitive dissonance brought on by the painting references the viewer's perceptual link to the hidden contexts of occluded objects, and suggests a way of thinking about occlusion in next-generation video. In the Pistoletto work, called "The Visitors," figures painted on tissue paper and affixed to floor-to-ceiling, mirror-finished steel create the impression that the surrounding environment flows through the work. This simple layer of interactivity turns the work into a portal that incorporates its ever-evolving context, causing the work to reach out of the 2-d plane and into the real (outside) and imagined (inside) spaces that it continuously creates and transforms.

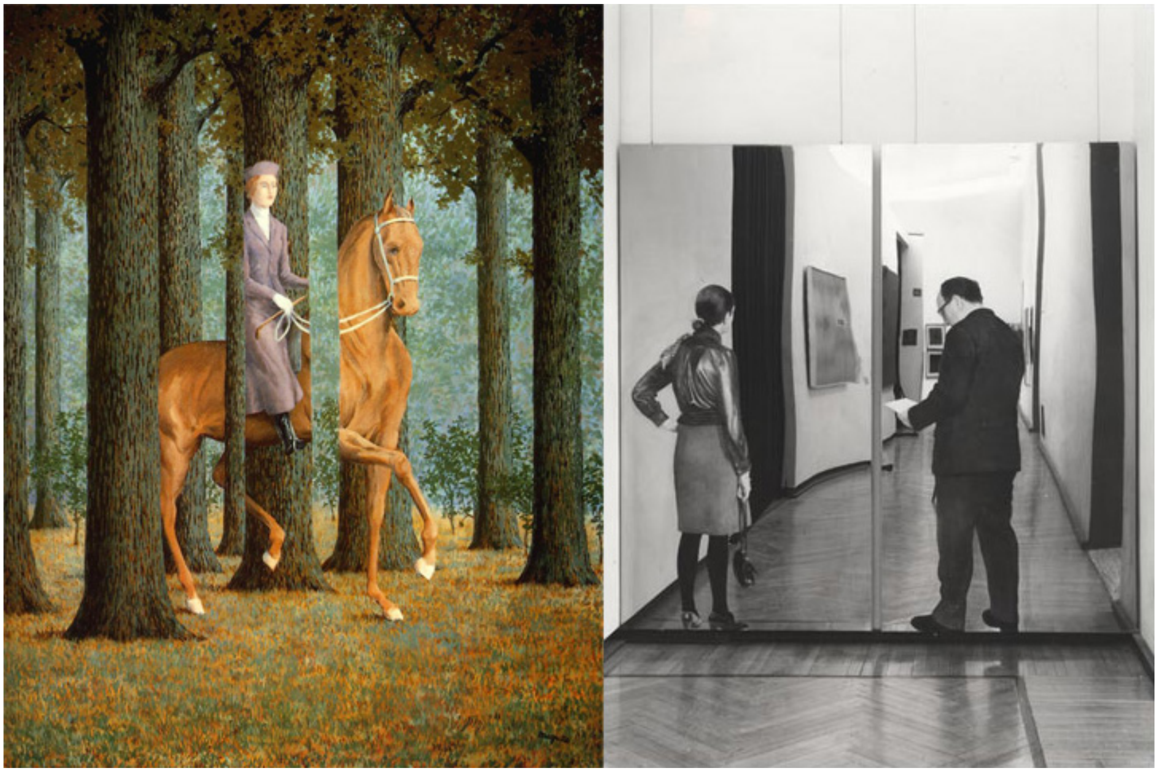


Figure 4: From left: René Magritte, “The Blank Cheque” (1965, oil on canvas) [National Gallery of Art, Washington, D.C.], and Michelangelo Pistoletto, “The Visitors” (1962-1968, figures painted on tissue paper and affixed to mirror steel) [cite: Galleria Nazionale d'Arte Moderna, Rome, Italy].

Seeking to harness this transformative potential of perspective-hacking works like the Magritte or Pistoletto, this chapter looks at ways of creating multi-perspective video interfaces that leverage distributed sensing to extend a camera’s reach beyond the lens and into the world. Recent advances in sensor fusion, wearable radio location tracking, and embedded computer vision point to new possibilities in distributed smart cameras and sensor networks, where applications can start to assume persistent correspondence between tagged objects and their image representations across multiple, networked cameras. This chapter develops systems that fuse cameras and wearable sensors to support such capabilities, and documents a set of projects that apply this thinking to the design of augmented video-based user interfaces. Mirroring the larger thesis, each of the projects in this

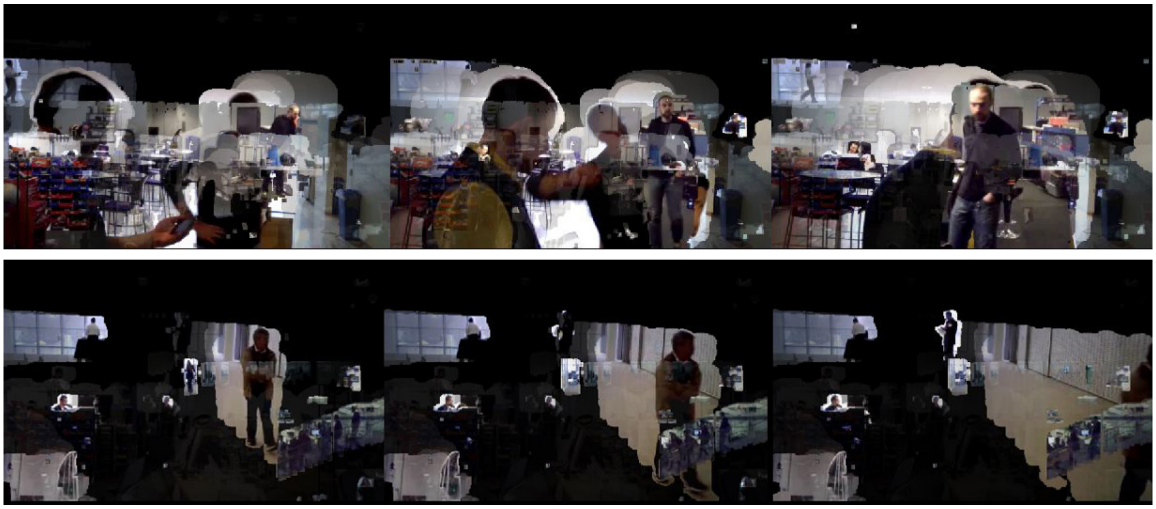


Figure 5: Still frames captured from Flurry at two different times of day, showing composites of building-wide activity captured by sixteen distributed cameras and fed back to screens on each camera device.

chapter stands on its own in terms of motivation and application, but ideas are developed, evolved, and shared across all three.

3.2 Towards Concurrent, Multi-view Salient Video

As noted above, nearly two decades of research have produced powerful methodologies for condensing visual information from pre-recorded video into single frames or short sequences, but for the most part, these techniques do not extend to the fusion of multiple concurrent videos, especially in real-time. Outside of the surveillance industry, which to date has been the only real consumer of distributed, real-time video, increasing deployments of smart camera networks and systems of ubiquitous media portals like [11] have brought new interest and attention to the problem. Assuming a large number of distributed cameras, how

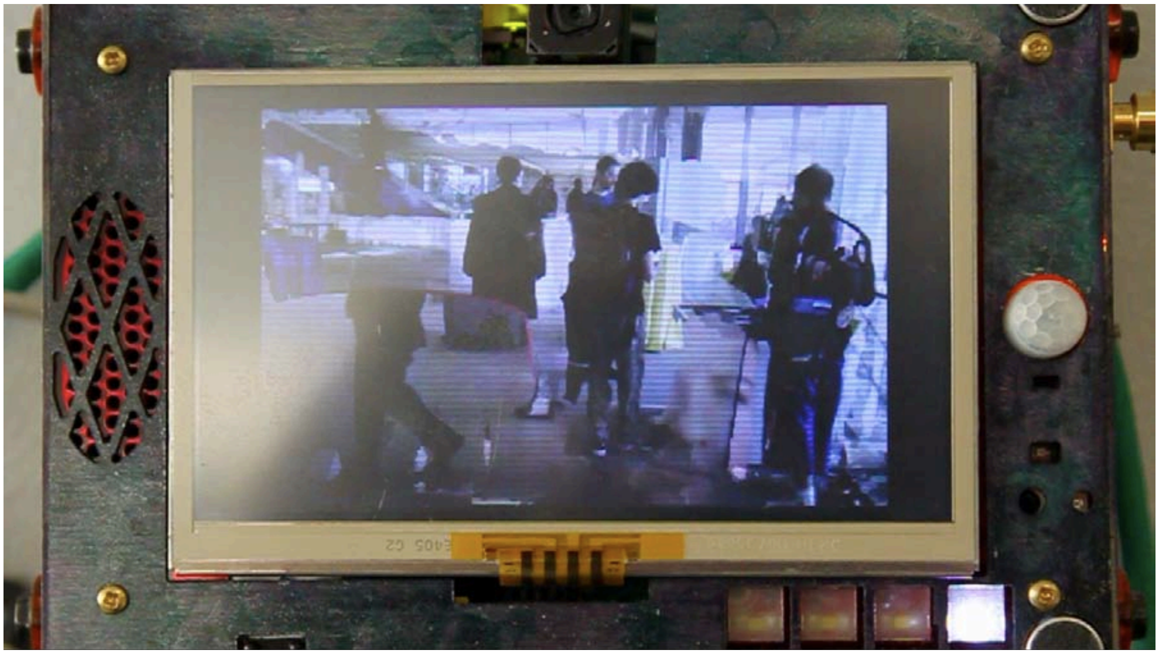


Figure 6: Flurry, fed back in real time to the media portal

can we create salient visual representations of the activity taking place throughout a building?

An early attempt to answer this question produced *Flurry*, an interactive video installation that used the ubiquitous media portals [11] distributed throughout the MIT Media Lab complex to capture fragments of activity from multiple perspectives and weave them together. In the application, video frames from a large number of viewpoints are collected on a central server and composited into a single stream using linearly decaying motion-history images [45] to key the sources and blend them accordingly. Objects that move more than others are made more visible in the output video, and fade over time when their motion stops.

In the installation, which debuted during a conference at the MIT Media Lab in March, 2010 and continued running for several subsequent weeks, the composite images were streamed back to the devices' screens, turning them into permanently open video portals. Looking into one device, viewers would see themselves mixed with anyone else doing the same on another device. Because the nodes were distributed across public and social spaces in the lab, activities in each space were automatically broadcast to others, bringing about immediate engagement through the portals as well as inviting physical attendance. Users treated the devices as portals, interacting with others who happened to simultaneously look into the multi-perspective, multi-user window formed by each node.

Through camera-mediated engagement with its subjects, who encountered the installation at the portal itself, Flurry proposed accidental correspondences within the lab-wide fabric of activity that might give rise to unexpected and alternative lines of communication. By acting as a mirror and bringing its viewers into the interface, Flurry invited exploration of the content it was offering—watching meant contributing, and vice versa. In this way, Flurry recalled the Pistoletto mirror-paintings and their transformative, incorporative effect on both space and audience. After the installation, Flurry's visualization was stored as a static record of these input-output user interactions as well as the everyday happenings of a research lab.

To address privacy concerns, the devices could be switched off with clearly marked lamp switches. Separately, the media portal system faced a great deal of privacy-related challenges, and nodes were often turned off by building inhabitants. In [46], the system was used to test privacy preferences and develop novel solutions; in a number of experiments, the authors test users' willingness to give up personal privacy in exchange for different kinds of applications and services. In informal observation of Flurry, it was found that nodes were left on by users in abnormally large numbers. This is consistent with results in [46] that suggest that more explicit transactions of personal information mitigate users' mistrust; the installation made its use of cameras clear by feeding the video output back to the nodes in real-time—users could see what they were sharing as they shared it.

There were a number of shortcomings to the installation that informed future work. First, the motion-history masks in themselves were not ideal for creating scalably legible results; as activity increased to a maximum, the composite became convoluted and messy. Second, by maintaining the spatial arrangements of pixels all the way from the imagers to the output of the compositing process, Flurry never brought the camera beyond its traditional field of view, and wasted precious screen real-estate that could have been used to make the its composition clearer. Finally, though the work inherited modes of interaction from Pistoletto, it did not summon Magritte; as a result of its reliance on the camera as its sole input modality, there was no hacking or redirection of perspective, and no rendering of the invisible. The video interfaces documented in the next sections attempt to address these concerns while taking lessons from Flurry's involvement of users through interaction and natural incorporation.

3.3 Sensing and Video for Construction Site Safety

One of the motivations for this thesis at the outset was increasing numbers deployments of sensor networks in domains where real-time information collected from sensors and delivered remotely has never been part of users' workflows. Moreover, there are a number of critical application domains which demand decision support, as opposed to purely algorithmically-driven automation through sensor fusion. In these cases, the question becomes how sensor fusion can play a supporting, visual analytic role for users, especially for better understanding remote context.

This section develops a system for one domain, construction site safety management, that is not only new to real-time data but also requires expert,

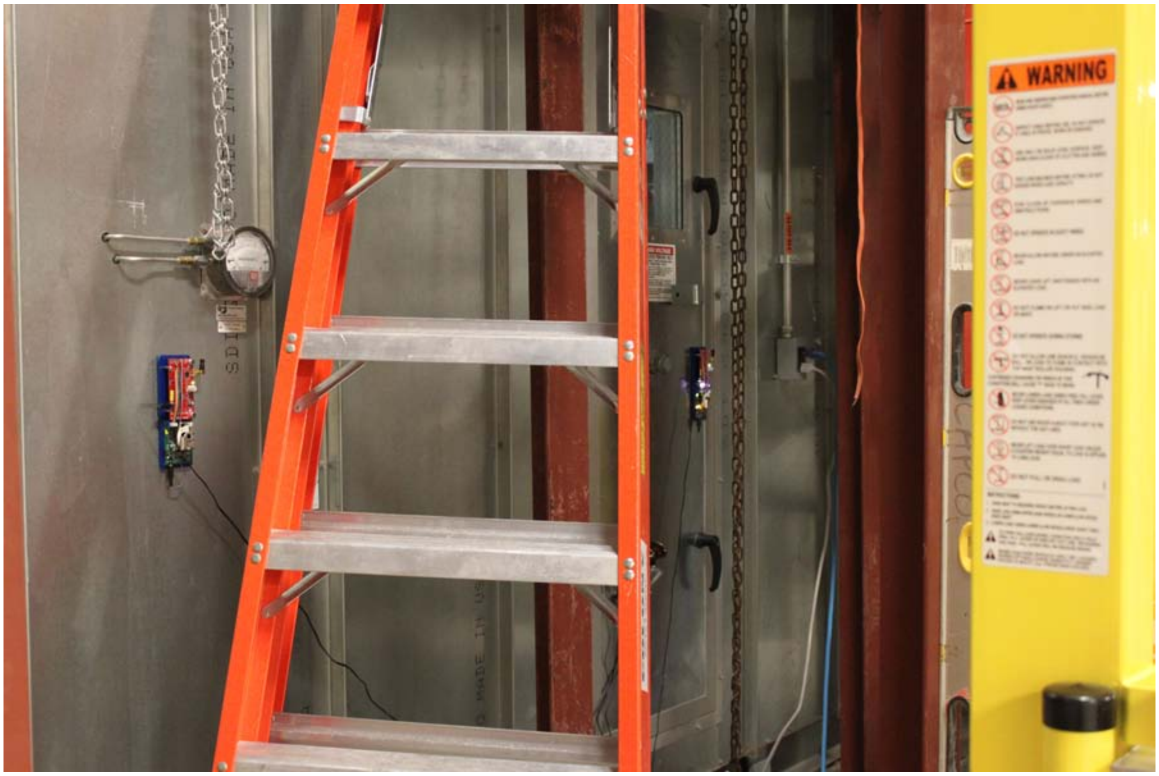


Figure 7: Cluttered construction site with two TRUSS sensor base stations magnetically mounted on either side of the ladder

human-in-the-loop engagement that cannot be replaced by automation. This project contributes to the larger body of work in this thesis in its application of camera-wearable sensor fusion towards an interactive and intuitive interface to sensor network data and video. Like the installation described in the last section, this project collects video from multiple sources and mixes it to distill and communicate context, but in this case, the compositing process is driven by data from sensors, mixing the cameras' perspectives with environmental sensing of otherwise invisible properties.

This project conceives of a remote exploration and decision-support system, called Tracking Risk with Ubiquitous Smart Sensing, or TRUSS, that infers and renders safety context on construction sites by fusing data from wearable devices, distributed sensing infrastructure, and video. Wearable sensors stream real-time levels of dangerous gases, dust, noise, light quality, precise altitude, and motion to base stations that synchronize the mobile devices, monitor the environment, and capture video. At the same time, small, low-power video collection and processing nodes track the workers as they move in and out of the field of view, attempting to re-identify the tracks using information from the sensors. These processes together connect the context-mining wearable sensors to the video; information derived from the sensor data is used to highlight salient elements in the video stream; the augmented stream in turn provides users with better understanding of real-time risks, and facilitates remote human decision support.

To test the system, data was collected from workers erecting and welding steel catwalks in a building during active construction; the results of these tests, as well as user evaluation from industry experts, are in chapter six of this thesis. The first iteration of the system, used in this user study, was not run in real-time (though the hardware would allow it), but rather to test the hardware and software in the challenging environment of a real-life construction site. For clarity, it is important to note the distinction between the system architecture, detailed in the next section, and the data collection exercise, which was not supporting a real-time risk assessment interface. The term “real-time” is used in context of the architecture to illustrate the design thinking, as well as plans for future testing and deployment. The depictions of interfaces and sensor data in this chapter show the results of the data collection as a proof-of-concept towards the near-term goal of real-time sensing.

This work is ongoing, with next generation hardware and software in active development. The second design iteration addresses many of the shortcomings of the first, and will run in real time. The gas sensor daughter board was designed for the project by TRUSS collaborator Brian Mayton in the Responsive Environments Group. Base station and mobile badge devices were originally designed by Mathew

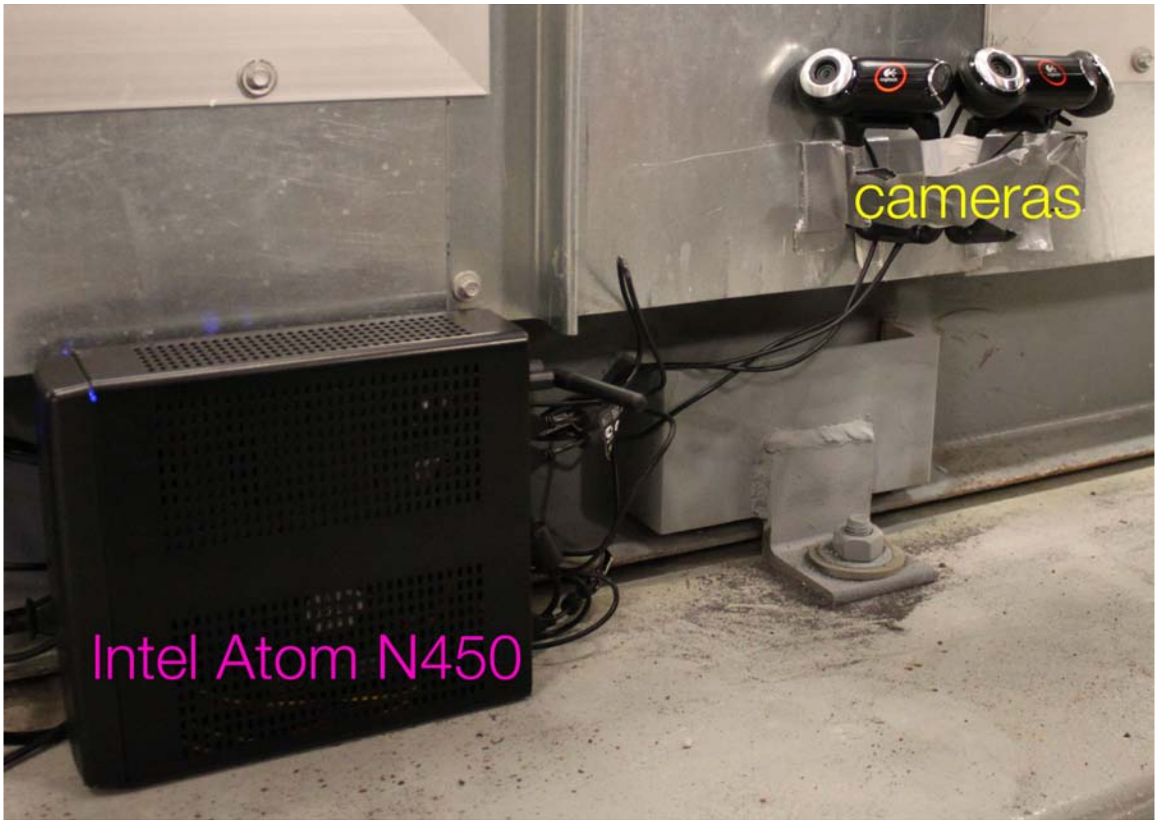


Figure 8: Low-power computer performs NTP time synchronization, bridges ZigBee network to the Internet using WiFi, and manages camera/vision subsystem.

Laibowitz for the SPINNER ubiquitous media portal system [11] and reprogrammed for this application.

3.3.1 TRUSS System Architecture

The next sections focuses principally on the TRUSS system architecture and interface design, and relate the latter to the general design principles introduced in the thesis introduction and earlier sections. The system is composed of three main hardware components, organized into a tiered, networked architecture: battery-powered wearable sensor devices, small, externally powered radio base stations

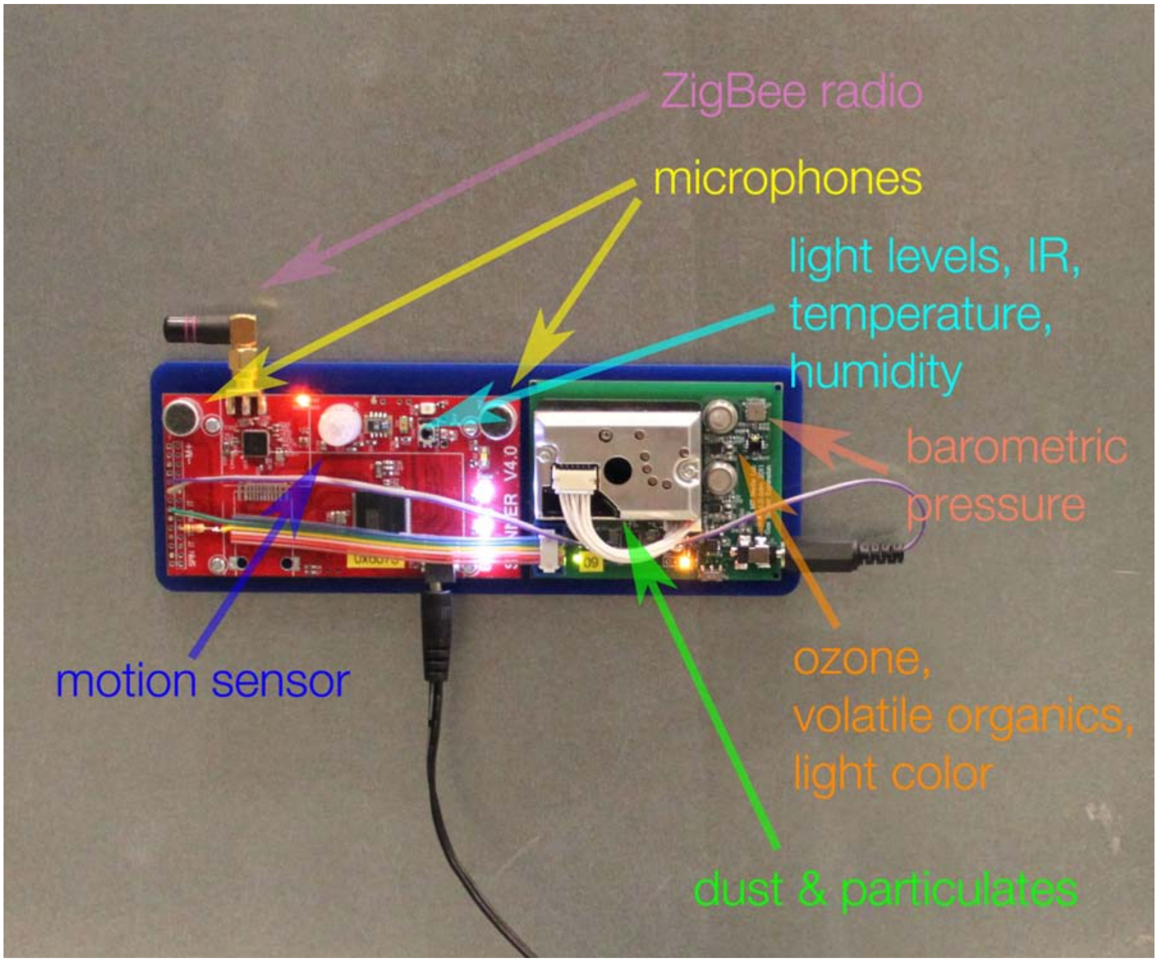

ZigBee radio
microphones
light levels, IR, temperature, humidity
barometric pressure
motion sensor
ozone, volatile organics, light color
dust & particulates

Figure 9: Base station sensor node, with “red board” streaming sensor hub paired to gas-sensor board.

with onboard sensors, and cameras attached to embedded Intel Atom-based computers running Linux. The embedded machine consumes approximately an order of magnitude more power than the base station, and the base station 2-3 times that of the mobile node. The components are time-synchronized using the ZigBee radios, enabling applications that fuse data from the independent sources; without synchronization the system becomes effectively non-functional, as none of the data can be correlated and confirmed across the nodes. In the absence of some models, and particularly in the unpredictable and challenging environment of a construction site, the uncalibrated signals from the gas sensors and drifting (though calibrated) signals from the pressure sensors cannot be reliably linked to activities.

The embedded computer acts as a network bridge between the low-power ZigBee network and a fixed-infrastructure WiFi (or wired LAN or GSM cell network, in the general case), performing network time protocol synchronization (NTP) over the Internet, synchronizing the local ZigBee network with NTP time, and enabling remote connections. The computer also hosts a video subsystem consisting of a camera, a computer vision library, a video encoder, and a streaming server. Another, significantly more powerful remote server could perform further operations on the video before it reaches users, adding another tier, though this is not the case in the existing architecture.

The fixed base stations shown in Figure 7 and Figure 9 are made up of a general purpose radio and sensor node (called the “red board”) connected to a daughter board designed for environmental monitoring. The latter is inspired by Angove and O’Flynn in [47]. The onboard sensors include a PIR motion sensor that can be used to trigger data collection or processing when workers are detected. Stereo microphones pick up loud crashes and yelling. An infrared “sociometric” sensor [48] on both the base stations and mobile nodes is used to detect where workers are facing, and when and how they work together. Light level and color sensors keep track of lighting conditions to detect welding or anomalous flashes of light. A barometric pressure sensor on the base station, in tandem with one on each mobile node, can together provide a precise (~10cm resolution) measure of relative altitude, measured between the base stations and mobile nodes [49]. Both base stations and mobile nodes carry the environmental monitoring system, with sensors that measure un-calibrated levels of volatile organic gases, hydrocarbons, ozone, and particulates.

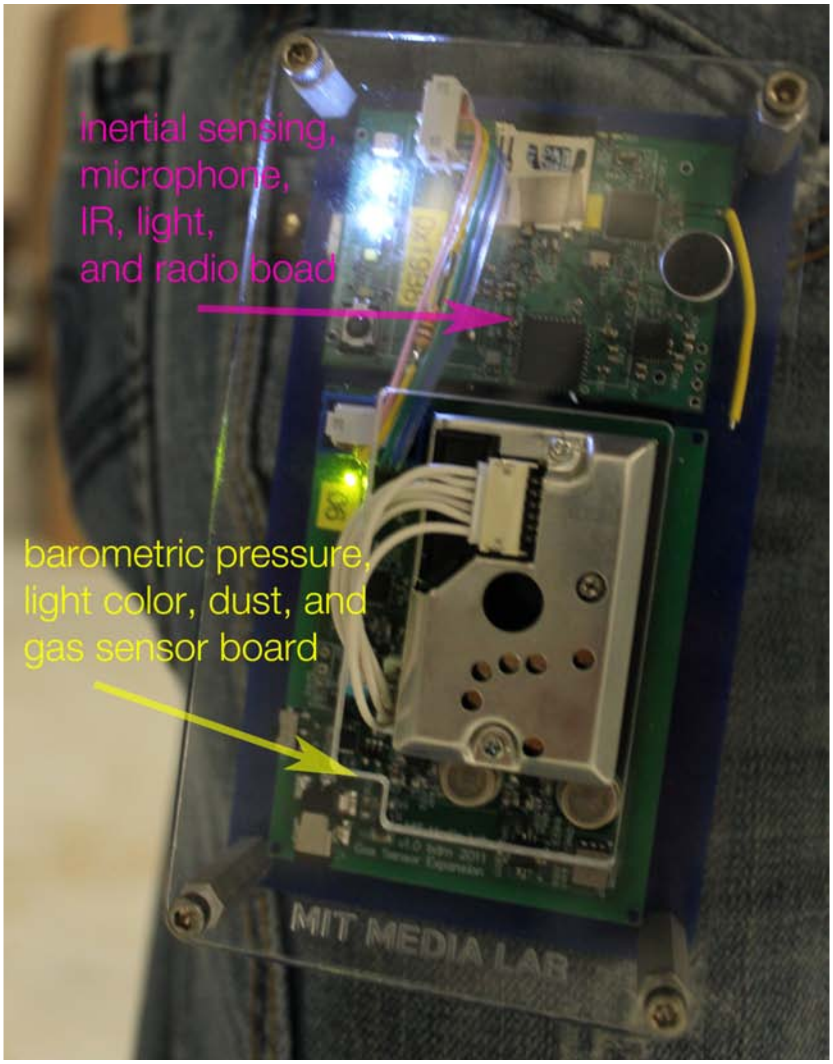


Figure 10: Belt-mounted wearable sensor node, with sensor badge paired to environmental monitoring board.

The mobile node includes a single microphone, the light level and color sensors, the IR sociometric sensor, and the sensor measuring temperature and humidity. It also carries a standard suite of inertial sensors: a 3-axis accelerometer, 2-axis gyroscope, and 3-axis magnetometer. ZigBee radios on both the mobile and infrastructure nodes are used for synchronization and can be used for rough

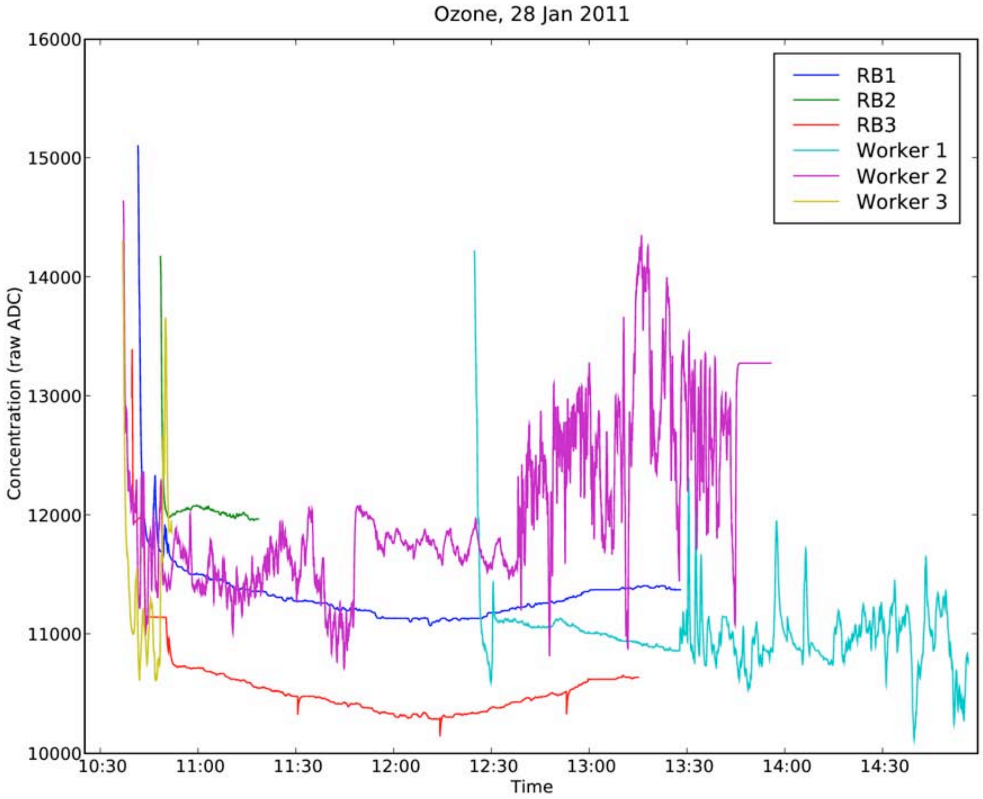


Figure 11: Ozone measurements, reflecting welding activity, collected from 3 workers and 3 base stations on one work day. The sensors are uncalibrated, but reflect relative levels.

localization of the mobile devices to within a rough area of 5-10 meters radius using radio link quality (due to hardware issues, we were unable to monitor the much more effective radio signal strength indicator). In our first test, the radiolocation accuracy suffered from a very challenging environment, facing large metal obstructions nearly everywhere and too low a density of base stations to provide useful location information.

This particular set of sensing modalities was chosen in consultation with construction site safety experts, based on their data about the most common causes of accidents on their sites. Commonly reported accidents include slipping and falling, falling objects, and dangerous chemicals. The gas sensors we included were chosen for the specific construction site targeted for our first deployment, which was to involve steelworkers who would be working at height while cutting and welding very large steel structures. These activities would increase the risk of falling objects while generating generate gases and particulates that could be flammable or harmful if inhaled.

3.3.2 Tracking and Sensor Fusion

In our system, workers wearing belt-mounted mobile sensor devices are monitored and tracked as they go about their normal activities; however, the cluttered and constantly changing environment of a construction site creates a particularly challenging tracking problem, with frequent occlusions, reflections and flashes from welding, and an inconstant background, subject to construction activities and large equipment in motion. Relying purely on radio-based tracking is also not realistic, as base stations must move relatively often to adjust to changing conditions and large metal objects are in constant motion, ruling out careful calibration. Dead-reckoning using wearable inertial sensors is prone to accumulating error, making it intractable. Of all the available sensors, cameras are the most effective external tools for tracking, as long as the worker does not leave the field of view or become significantly occluded; when a track is lost, the camera alone cannot recover it later.

Facing this rapidly changing, cluttered and heavily occluded environment, we opted for a relatively simple vision pipeline: a mean-shift blob tracker [50], operates on the image after a process of frame-differencing, thresholding, morphological operations to remove noise and join disconnected components, and a weighted moving average to smooth the motion. This approach works well to find and track people who move with some degree of frequency, but also finds anything else that moves (thought this problem can be mitigated slightly by a well-chosen search radius and

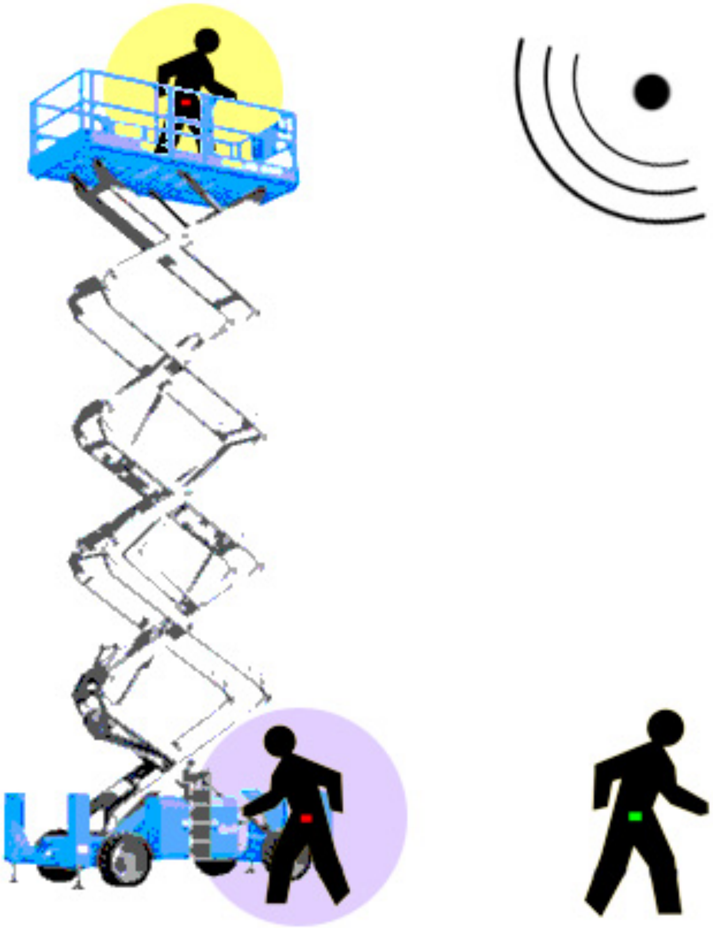


Figure 12: TRUSS system architecture, showing workers outfitted with wearable sensor devices. Risk bubble metaphor reflects parameters of the area around each worker, such as levels of dangerous gases detected by the wearable, and takes into account the tracked positions of other workers, in particular when they are working at height or below others.

blob size threshold). Still, the tracker can not segment multiple people when they occlude each other, causing problems when they separate again and the tracker cannot resolve the path ambiguity.

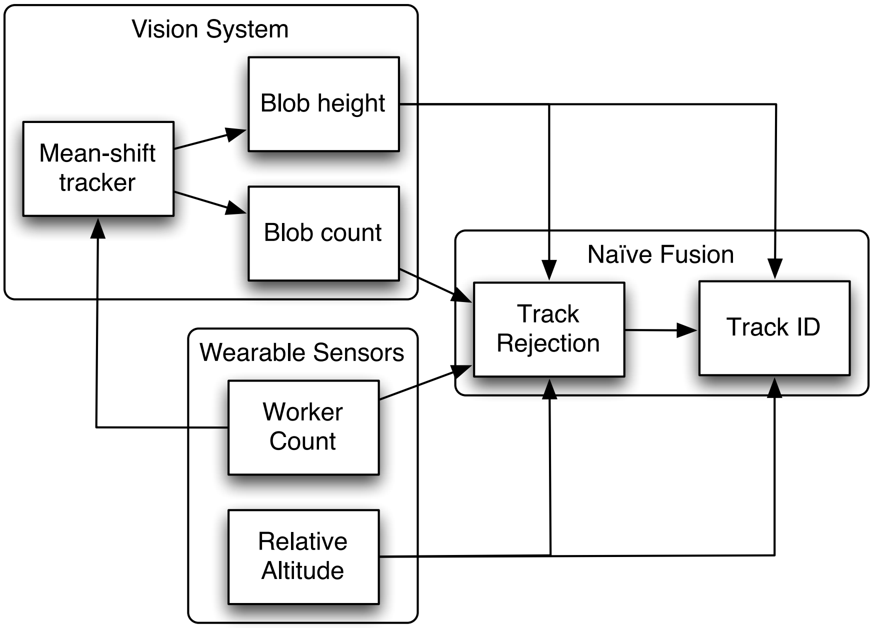


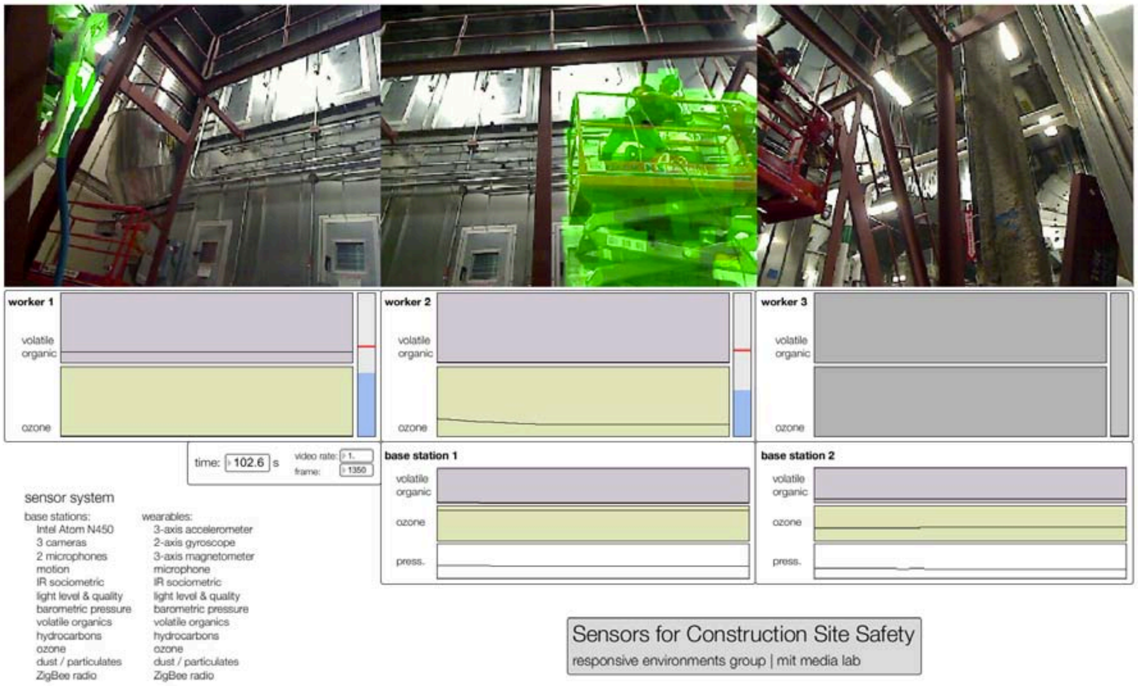
Figure 13: “Naïve” fusion of vision system and wearable pressure sensor, as proof-of-concept.

In [9], Teixeira, et al. use inertial sensors in conjunction with cameras to create correspondences between ambiguous or temporally disparate tracks, developing a distance metric on gait timing events between the signals extracted from the camera and worn accelerometer. However, the cluttered environment of a construction site like the one we were targeting is not conducive to this technique, as there is little observable walking (especially from a sufficient distance to track gait), many continuous occlusions, and long periods of relatively stationary activities (one worker welding in a cluttered area, or two workers on a lift, for example).

However, in the environment of our test and in many similar scenarios, workers are often ascending and descending on lifts and ladders in cramped spaces where there is little opportunity to move radially towards the wall-mounted cameras. This suggests that a fusion between the pressure-based altitude sensor and the camera tracker could recover worker ID after track ambiguities. This fusion would take a small step towards solving the larger correspondence problem by re-identifying workers on occasion, and at the very least, this information could be used to weed out spurious tracks. Of course, this kind of fusion approach fails when non-instrumented workers work closely amongst instrumented ones, or if the workers are at the same height, but the system can subsequently recover. In this scheme, the tracker fuses three pieces of information: the number of workers in the field of view (extracted from the radio signal strength), the altitudes of each worker provided by their wearable pressure sensors, and the estimated height of each image blob.

The approach to this fusion taken in the first deployment, illustrated in Figure 13, is a simple flowchart-based algorithm that performs logical and naïve nearest-neighbor operations on the signals from each sensing pipeline. While this approach is certainly not the optimal one, it is intended as a test of the data correspondence in the prototype system. Under the conditions described above, the system can reject spurious tracks caused by moving equipment and shadows, as well as identify multiple workers in a scene after the tracker's state has been cleared.

In general, the fusion approach seeks to extract state that is shared between the sensor signals. In our tests, detailed in chapter six, systemic and sensing challenges precluded the use of the other wearable sensing modalities in most cases, though there are clear avenues for fusion in future work. This kind of thinking can expand to include fusion between microphones (audio levels) on workers and in the infrastructure (building on [7] and [8]), inertial sensors and cameras (as in [9] and [10]), or new radiolocation strategies that could, together with a vision tracker, effectively solve the correspondence problem outright (see section 7.3)



time: 102.6 s
video rate: 1.
frame: 1350

sensor system

base stations:
Intel Atom N450
3 cameras
2 microphones
motion
IR sociometric
light level & quality
barometric pressure
volatile organics
hydrocarbons
ozone
dust / particulates
ZigBee radio

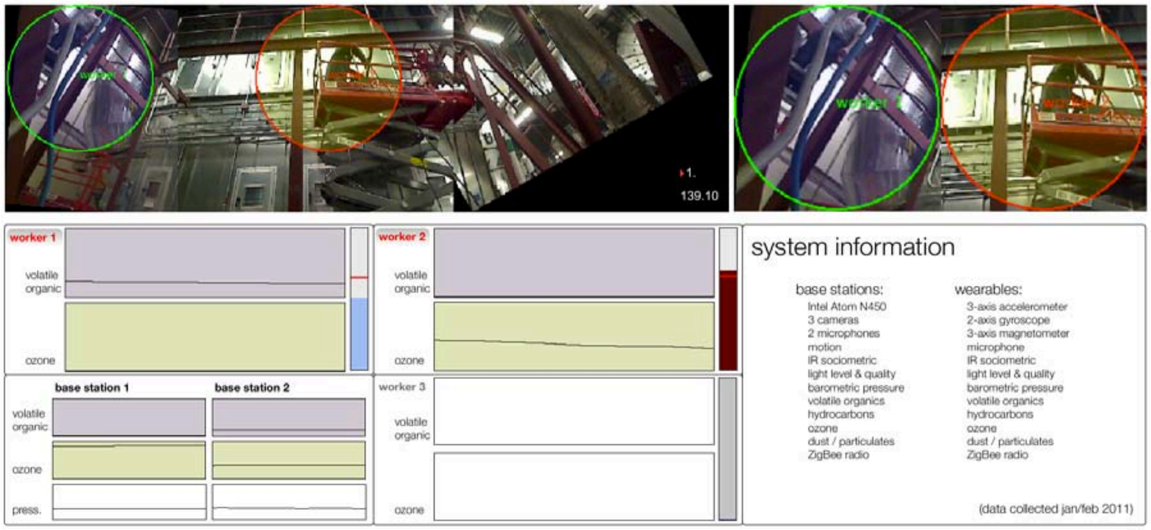
wearables:
3-axis accelerometer
2-axis gyroscope
3-axis magnetometer
microphone
IR sociometric
light level & quality
barometric pressure
volatile organics
hydrocarbons
ozone
dust / particulates
ZigBee radio

Sensors for Construction Site Safety
responsive environments group | mit media lab

Figure 14: First version of the worker safety interface, showing unstitched video and altitude thresholds on each worker; when a worker passes the threshold set on the pressure sensor, the corresponding blob is marked in red.

3.3.3 Interface to Safety Context Through Augmented Video

This section presents a real-time risk assessment interface for remote safety managers that combines sensors from multiple sources to better communicate the changing context around each worker. The software interface augments video of workers on the construction site with information from their wearable sensor nodes, and allows users to set thresholds and priorities on single data streams or combinations of streams. Sketched in Figure 12, this person-centric information architecture imagines a worker safety bubble metaphor, where a sphere of some variable radius encloses a worker's local context, and highlights the intersection of that context with others.



system information

base stations:
Intel Atom N450
3 cameras
2 microphones
motion
IR sociometric
light level & quality
barometric pressure
volatile organics
hydrocarbons
ozone
dust / particulates
ZigBee radio

wearables:
3-axis accelerometer
2-axis gyroscope
3-axis magnetometer
microphone
IR sociometric
light level & quality
barometric pressure
volatile organics
hydrocarbons
ozone
dust / particulates
ZigBee radio

(data collected jan/feb 2011)

Figure 15: Second version of worker safety interface, showing stitched video, altitude threshold, video tinting to reflect levels of dangerous gases, and a “spotlight” view at right, as well as the same graph-based representations of the same data.

This approach is intended to keep the user’s focus on the workers while accounting for the causes of many kinds of common accidents. In many cases, coincident activities and circumstances that might otherwise be relatively safe can together be catastrophic, like a worker passing underneath another worker at height or incendiary gases coming into contact with sparks from a welder. By allowing interactive control over the sensor thresholds, selection criteria, and critical ranges, the visualization can adapt to changing conditions as the expert user sees fit. This thinking positions the interface, like the others in this thesis, in a decision-supporting or visual analytic role with respect to an expert user.

Figure 15 shows the latest revision of the TRUSS user interface to multimodal sensor data and video. The revised interface is designed to prioritize and highlight what its user deems to be the most relevant information while at the same time letting all the data through in the background, providing a broader context to the information it provides. On one side of the interface, all the video is visible—on the other, a salient “spotlight” view of the workers engaged in the riskiest activities. In

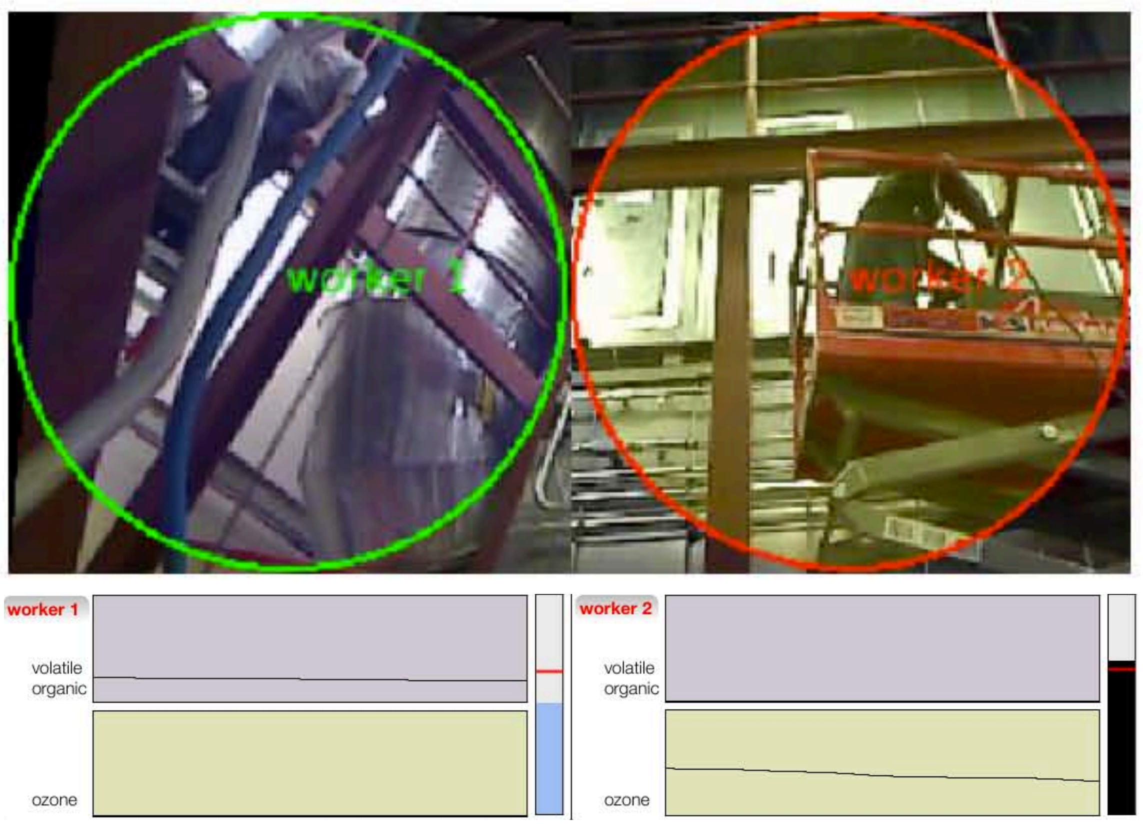


Figure 16: "Spotlight mode" selects and highlights workers based on “risk” as defined dynamically by the expert user; below, graphs of different gas levels with the same coloring as the video tinting, showing levels over the last 45 seconds. The red circle at right indicates that worker 2 has exceeded the threshold set on his wearable altitude sensor.

both views, workers are augmented with differently-tinted circles reflecting both the levels and types of dangerous gases in their vicinities. The outline of each circle reflects whether the enclosed worker has exceeded a threshold set on his pressure-based altitude sensor. Levels are represented by intensity and types by color, matched to the 2-d graphs below. The graphs show recent history, which place the more immediate video augmentation in a temporal context. When the augmented spotlight circles intersect, colors mix, highlighting the extent as well as the nature of the event. This visualization builds on the safety bubble metaphor, treating areas of overlap as particularly worthy of user attention and making that information most salient. In the example given in Figure 17, the elevated level of ozone surrounding the worker on the right indicates, correctly, that he has been welding, while the elevated level of volatile organic gases around the worker at left indicate an activity involving some kind of paint or aerosol. Overlapping bubbles of these kinds would indicate an increased risk of fire or explosion, and trigger an alarm; the radii of the safety bubble can be set dynamically in the interface.

Building on the Flurry installation documented in the last section, the TRUSS interface offers a mode that composites the views from each camera into a single stream. However, rather than compositing based on motion and effectively ignoring the relevant context, the system weights each input by the user's choice of any sensing modality or combination thereof. Depending on the selection, the software might emphasize video of those individuals exposed to the highest levels of ozone, or those most proximate to others working at height. In our first deployment, this feature was not particularly useful because of the relatively small number of workers and cameras in the exercise; further development of this idea is planned for larger deployments.

The combination of open-ended interactivity and salient selection has a number of important motives that figure into the larger thesis. First, as a new interface to these data in a life-or-death application domain, the hands-off approach to the design keeps the expert in control while still seeking to provide as much assistance and guidance through the video augmentation as possible. Second, the interactivity is designed to increase user interest and engagement through play, facilitating exploration of the data by new users. Finally, the multitude of viewing modes of these data provide an opportunity to study how the interface might best fit into a safety officer's workflow. What is the user most interested in seeing? What kinds of events would drive them to action? These questions are preliminarily addressed in interviews with industry professionals, and their answers are documented in chapter six.

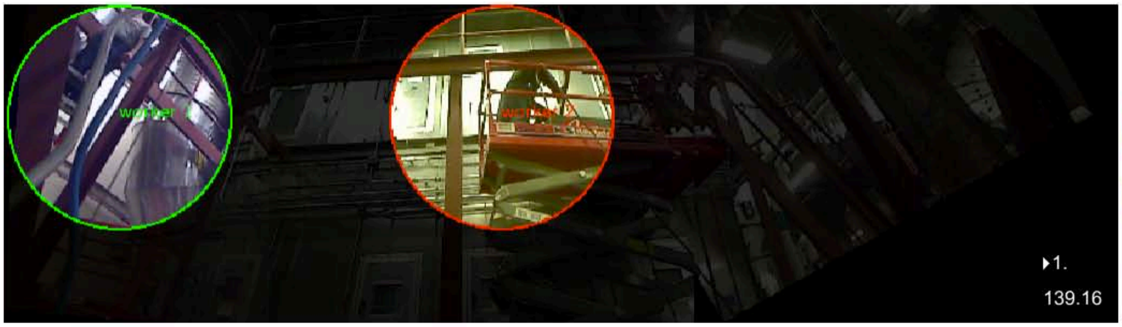


Figure 17: "Spotlight mode" shown as highlighted areas of the larger set of videos, with more obvious tinting and thresholds.

3.4 Video Interfaces to Sensor Network Data

The TRUSS interface presented in the last section combines multiple video sources to form salient views into remote context that are driven by sensors distributed throughout the environment. The system uses basic sensor fusion to identify actors in the video and augment them with information from their wearable sensors. The interface builds on concepts developed in the Flurry installation, seeking to address the latter's shortcomings by using information beyond pure vision to control the visualization, and reaching into context that the camera alone cannot access.

In these interfaces, video serves as a multi-functioning sensor, a microscopic link to the larger remote context, and a canvas for painting information from portals distributed throughout the field of view. The next chapter develops a macroscopic view that leverages a completely different kind of canvas towards a more open-ended user experience of browsing sensor data. Later chapters propose to glue these concepts together to form interfaces that span micro- and macroscopic perspectives through a combination of video and 3-d animation.

4 | Exploring Dense Sensor Networks

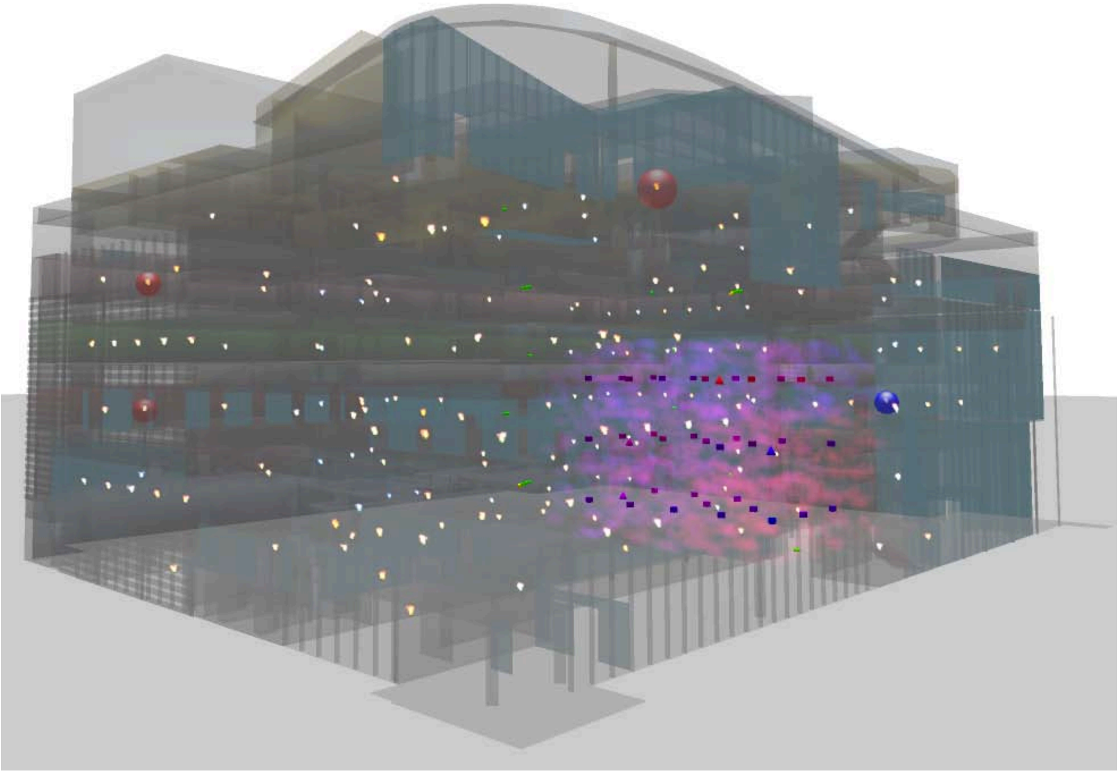


Figure 18: Translucent view of the MIT Media Lab in DoppelLab, with collocated visualizations of real-time and aggregated sensor data from multiple, independent networks.

4.1 Introduction and Motivation

The examples presented thus far have introduced sensor networks paired with user interfaces that together collect and communicate contextual information through augmented video. These interfaces are designed to enable new users to explore and make use of the data, but they are also highly specific to the application and environment that they support, and their system architectures are deeply tied to specific networks. In the event that a new data stream should become useful for providing additional context, the systems would need to be redesigned to accommodate the new information. In short, these interfaces may well be effective and intuitive in their specific applications, but they do not scale or generalize easily.

The video interfaces presented in the last chapter have the advantage that their relationship to some physical space is clear and concise; what you see is what you get. But this feature can also be limiting; that is, even if the sensor-driven augmentation can extend the visible range of the camera, the interface is always constrained to some fixed angles of view. In addition, the user has no way of relating the static view to the larger environment of the camera. This problem is exacerbated as the amount of information increases, leaving little room in the 2-d plane of the video in which to layer much more information.

Motivated by increasingly dense deployments of disparate sensor networks and a distinct lack of interface tools that can truly span across these siloed data streams, this chapter develops DoppelLab, a scalable, open-ended platform for creating collocated representations of general sensor network data. This platform supports an immersive, 3-d virtual space in which users can explore real-time sensor data and inferences of all types, as well as long-term, aggregated information and knowledge. DoppelLab aims to drive new interfaces to physical world actuation and control by providing tools for rapid parsing, visualization, and application prototyping that can take advantage of the platform's fully horizontal relationship to otherwise independent sensor networks.

Intuitively and scalably collocating representations of dense sensor network data presents a significant design challenge. This chapter elucidates these challenges, and draws on visualization design strategies to help solve them. In addition this chapter details the back-end server and database framework that supports collection of data from multiple, disparate networks and generic, scalable storage,

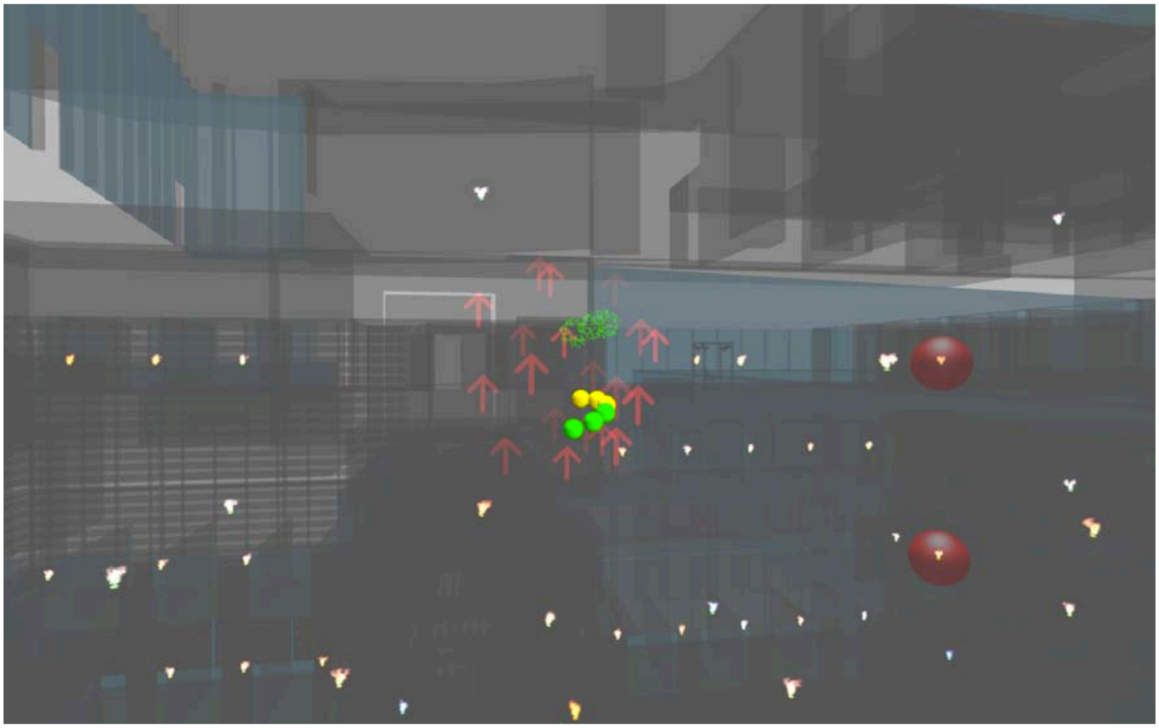


Figure 19: An unusually high level of social activity is detected on the fifth floor, by the cafe, and it appears to be increasing quickly. There is usually less going on between 10am and 11am on a Tuesday. Meanwhile, in the background, two server closets are overheating.

providing a means for aggregation and analytics that cut across individual sensor networks. More broadly, this chapter lays the groundwork for intuitive, exploratory modes of user interaction with sensor network data. By enabling users to move fluidly between convergent data streams and the higher-level inferences that fuse them together, this work aims to develop interfaces that foster better understanding of sensor-driven context and the data that produce it.

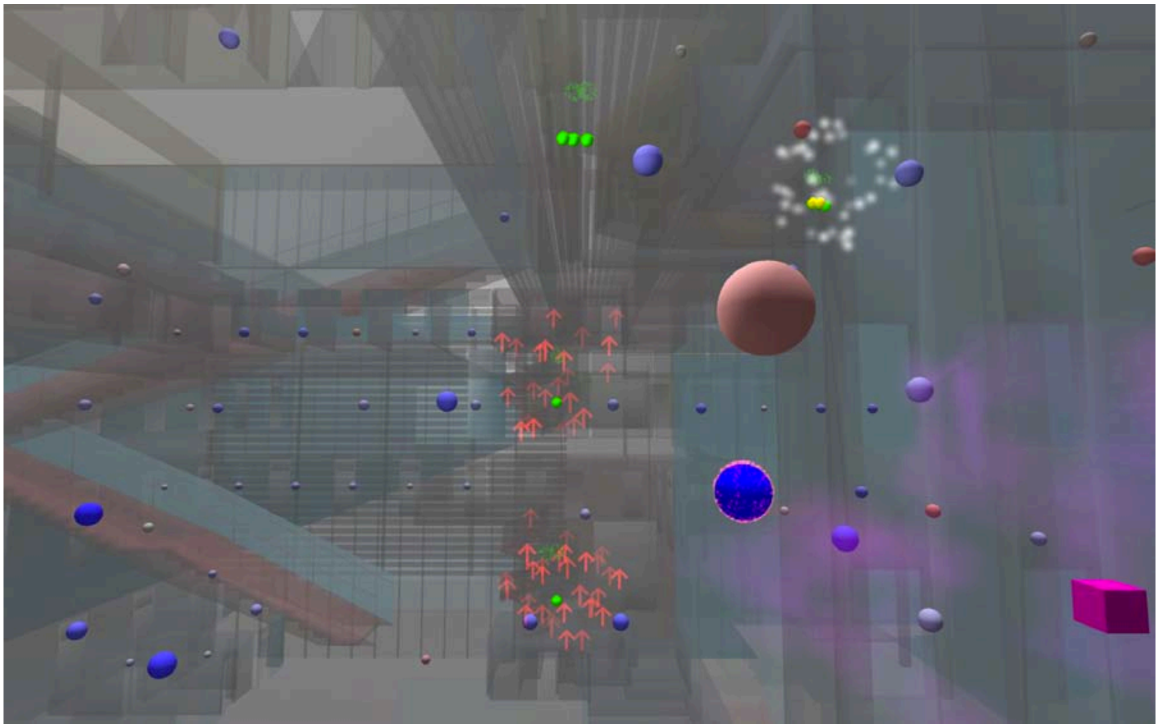


Figure 20: Inference of levels of social activity show two areas where activity is increasing (red arrows) and one area where activity is higher than normal but remaining constant (ghostly apparitions above right).

4.1.1 Sensor Data in Three Dimensions

Moving to 3-d increases the complexity of the representation but adds a great deal of flexibility and visual bandwidth, as well as possibilities for interaction through first person perspective and exploration. Although 3-d visualizations can support a great deal more information, it is not immediately clear how best to organize these diverse data streams in the 3-d space for presentation to users; naturally, different applications benefit from different solutions. However, we propose that organizing data by the space from which they originate makes for a largely intuitive and general platform from which to make both broad and specific queries about the activities, systems, and relationships in a complex, sensor-rich, human-actuated environment. This system of organization emphasizes the relationship between people and their physical environments, and imagines distributed sensors as densely distributed atomic portals into that relationship. Organizing these portals by their physical-world arrangement supports queries that hinge on people and space:

- How does the temperature in one room relate to the one next to it, or one across the building, where the sun sets at this time?
- Where is social activity occurring?
- Where are people most excited about what they are doing?
- Do any or all of these data streams correlate in some way? Across people? Across spaces?

This adherence to the physical space also provides a means for scaling visualizations against the architecture itself, where the importance of a visual cue is always contained by the walls of the model, while at the same time the model itself becomes a medium for painting information. While projections and transformations to other spaces can produce more efficient representations of data for many kinds of queries, the physical space serves as a consistent and relatable starting point. Discussion with facilities industry experts has revealed the need for building occupants and managers to share some common frame of reference; in this work, the 3-d space serves in that role.

4.2 DoppelLab

DoppelLab is an immersive, cross-reality virtual environment that serves as an active repository of the multimodal sensor data produced by a building and its inhabitants. Built on the Unity3D game engine, the platform transforms standard architectural models from computer-aided design (CAD) into browsing environments for real-time sensor data visualizations and animations, organizing the

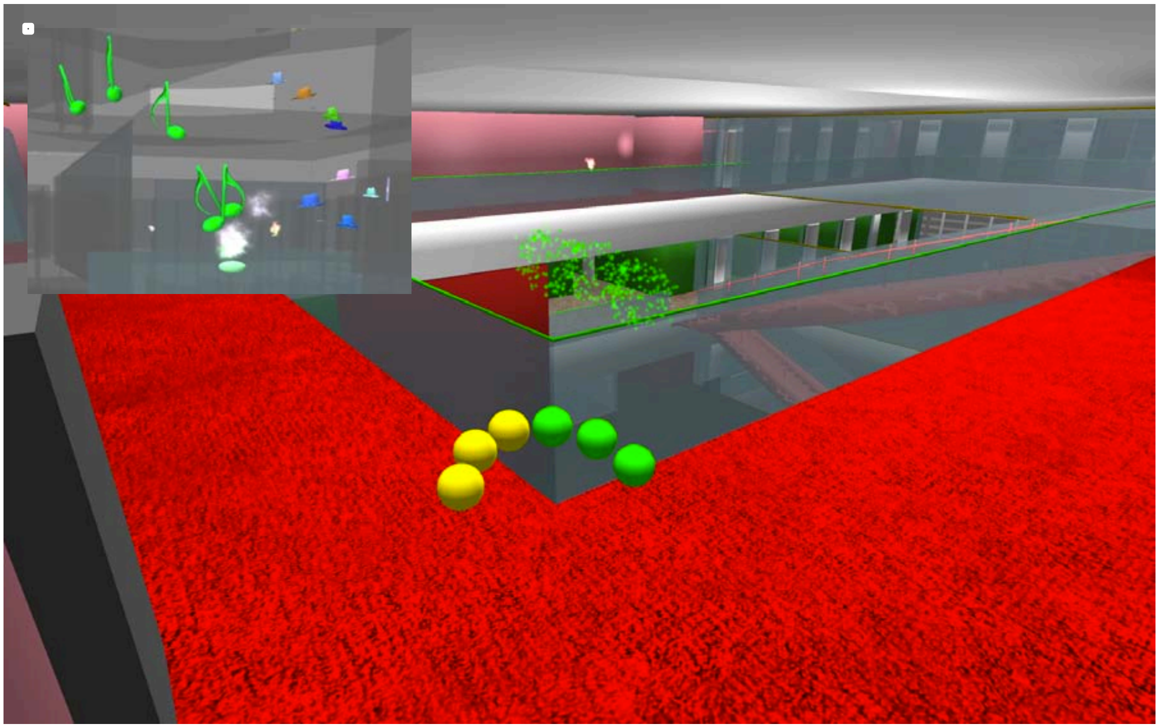


Figure 21: Typical levels of audio and motion from 4:00 PM-5:00 PM compared to the real-time levels. The particle-system representation of aggregate data materializes above the real-time one only when the user approaches. Inset: another representation of the same quantities, formed by accumulating musical notes showing audio levels and flying bowler hats with cigars showing motion.

representations by the physical space from which the sensors originate. DoppelLab lets users walk through and interact with these representations within a building, or fly out of the building to see it buzzing with layers of activity and information. The platform leverages physics and lighting engines, interactive animations, and other game engine affordances such as distance culling to structure and encourage exploration of the data. At present, DoppelLab supports 3-d, spatialized audio streams from real-time sources, with speech obfuscation performed at the node-level for privacy protection. In an parallel research effort, we are beginning to mix these live streams with sonifications of the data that will make the user experience more immersive [51].

4.2.1 Towards Visual Analytics in a 3-d Virtual Environment

In DoppelLab, visual representations of data take the form of metaphorical animations that express absolute and relative magnitudes, rates of change, and higher-level inferences that relate data across time, space, and sensing modalities. Visualizations take place within the walls of the building, corresponding to the physical locations of the sensors, and the walls can be toggled into translucence to expose the relationships that span across the building.

The animations are easily adapted or swapped for different applications; a simple development process is central to our system design, allowing developers to quickly prototype their own visualizations and customize existing ones. The inset in Figure 21 shows one example of this modularity, an alternate, more whimsical representation of audio levels and motion. The application makes a number of animations "drag-and-drop" ready so that developers need not repeatedly reinvent existing designs, and to produce a certain level of consistency in the user experience. The goal is that DoppelLab support a rapid development-to-visualization-to-development cycle, whereby visualizations suggest relationships and correspondences to users, and the environment enables on-the-spot prototyping of new applications. One example of this process, designed for building facilities managers, reveals anomalies in the thermostat data—specifically, large deviations from the local set point. This idea came about because a visualization of absolute temperatures showed what appeared to be strongly correlated and unusually high temperatures in a set of adjacent rooms; significantly, the result exposed a previously unknown fault in the building HVAC system that was later corrected as a result.

We draw on Edward Tufte's seminal works to identify a number of visual principles that structure our interface design [52], [53]; these principles are discussed here as a set of design goals that address some of the challenges of dense, multimodal visualization. As mentioned above, our adherence to the physical space provides a means for *self-representing scales* that normalize the size (and relative significance) of any visual cue against the architecture. Moreover, parameterization of the properties of the 3-d model turns the model into a useful chart in itself, where the walls can take on qualities of the activities they contain.

In DoppelLab, a large and fast growing set of data sources poses a visual and interface design challenge. As such, we seek visual representations that engage with users to reveal information and expose functionality in response to their exploration through the virtual space, making the density of information in a representation a function of the user's virtual proximity to it. This notion of *macro and micro design* makes the platform scale to large numbers of data sources and analytical visualizations without overwhelming the user at the macro level. In this vein, we have defined design archetypes for a number of different categories of information. Figure 21 shows particles hovering above a representation of audio and motion as an altered copy; these particles reflect the aggregated amount of those sensor values for the particular time and day of the week, and only materialize when the user is close enough to the representation to indicate interest. We have developed four of these animation archetypes thus far:

- Objects composed of smaller objects, or particles that resolve only when a user is proximate—e.g. a sphere made up of small arrow particles that form a vector field.

- Objects with changing or morphing shapes—e.g. a cube that transforms into an upwards- or downwards-facing cone to indicate sign and rate of change.

- Objects or animations that that share form but differ in material and/or makeup—e.g. a solid object and its translucent (glass) counterpart.

- Translucent objects that can take on properties like color and shape, but that contain additional representational objects within them

Also with an eye towards scalability, we aim for visualizations that make use of every graphical element to show data; these *multi-functioning elements* consolidate conceptually related sensing modalities (i.e. coincidental audio and motion) into compact forms, using multiple properties of each object or animation. This strategy associates the modalities, and reduces the *chart junk*, while still representing all the independent data streams.

4.2.2 Visualizations in DoppelLab

In its current form, DoppelLab renders data from an increasingly large and dense set of building-wide distributed sensors at the MIT Media Lab complex and one researcher's instrumented living space. To date, we have incorporated hundreds of building-wide distributed sensors at the MIT Media Lab and elsewhere. Currently DoppelLab is rendering visualizations of data from a number of densely-distributed sources in the MIT Media Lab complex, as well as one researcher’s instrumented living space.

A dense network of 45 temperature and humidity sensors is suspended in a large atrium and represented in DoppelLab by nodes whose color represents temperature, where red is hotter and blue is colder (a consistently applied metaphor). A node’s shape reflects the sign of its rate of change, at a user-definable timescale; upwards-facing cones reflect increasing temperature, and downward-facing decreasing. These nodes are surrounded by fog-like clouds of particles that track levels of humidity, where denser, redder fog indicates high levels, and bluer, sparser fog the opposite.

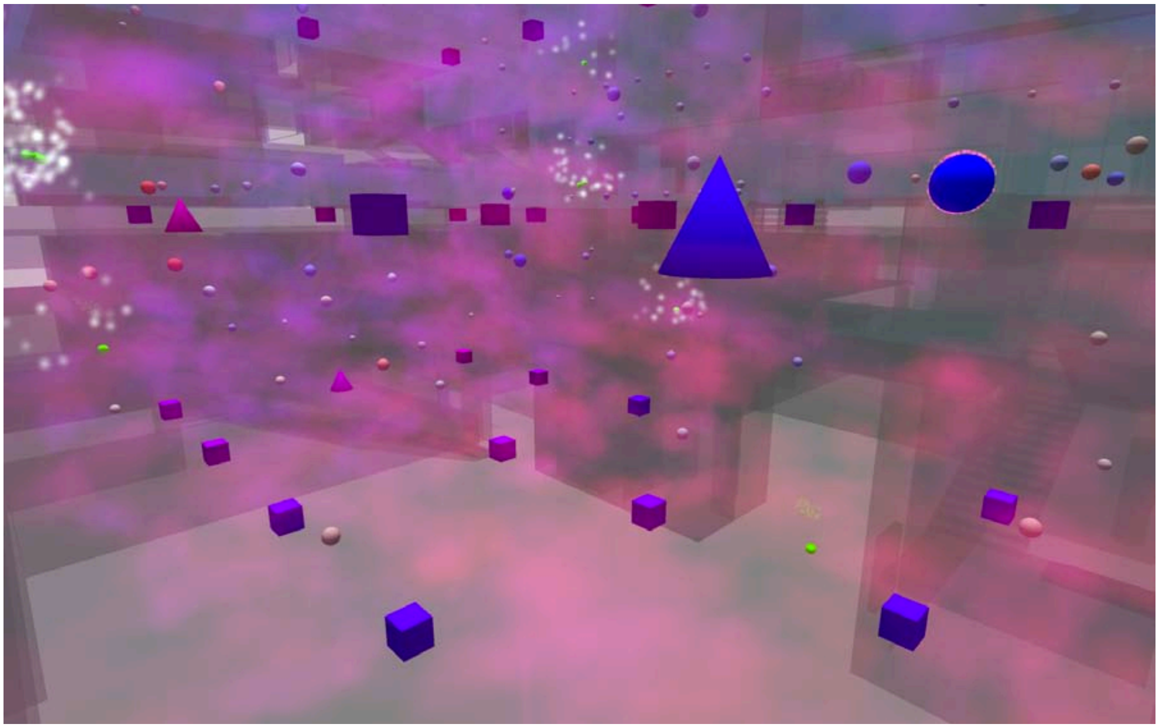


Figure 22: A dense sensor network monitors temperature and humidity in a large atrium. Redder shapes reflect hotter temperatures and redder, denser fog higher humidity. The shapes themselves indicate whether the rate of change is positive, negative, or zero.

DoppelLab supports multiple visual interpretations of the same data that can be toggled using keyboard modifier keys, like *shift* and *alt*. Shown in Figure 23, a system of several hundred thermostats at the Media Lab is represented alternately by flame animations whose colors reflect absolute temperature, and by spheres whose color and size reflect the local deviation of the temperature from the set point; both visualizations highlight anomalously large differences between the two. The ability to quickly switch between these representations allows users to explore how the rooms relate to each other in terms of energy flow (i.e. whether hotter rooms heat their neighbors) and HVAC system performance (i.e. whether poorly performing nodes are spatially or systemically proximate). The two snapshots in Figure 23 shows a set of 3 server closets that are stacked directly above one another; the top two are overheating (indicated by large, red spheres), while the

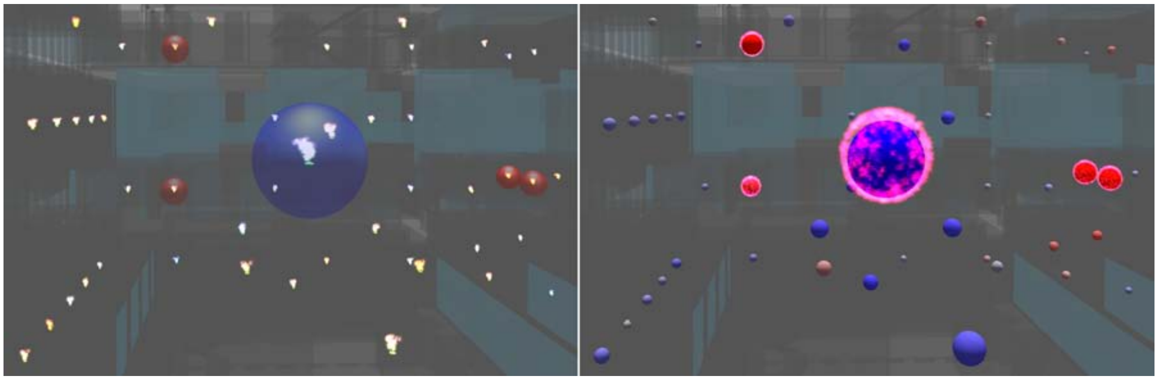


Figure 23: Thermostats are represented by colored flames (left). An anomaly between the set point and measured temperature is indicated by a pulsating sphere whose color reflects whether the temperature is above or below the set point. An alternate visualization (right) shows the magnitude and sign of the discrepancy as the sizes and colors of the spheres; those that exceed user-defined thresholds are highlighted by pink clouds.

bottom one is properly cooled (to a level below the building-wide set point). The image on the left shows the first-floor server room to be colder than all its neighbors, while the image on the right shows the HVAC system to be performing perfectly in that space.

DoppelLab connects to a network of distributed sensor nodes carrying motion, temperature, humidity and light sensors, as well as stereo microphones. Some of these nodes are linked to daughter boards with barometric pressure, light color, ozone, volatile organic, and particulate sensors, for monitoring machine shop activity (detailed in section 3.3.1). Coincident audio levels and motion are represented as a set of spheres that accumulate into circles as audio levels increase (resembling a traditional level meter), and undulate as motion increases, combining the two into an indicator of social activity. When a user approaches the visualization, the typical sound and motion levels for the current hour and current day of the week appear for comparison as a ghostly copy of the original. As shown in Figure 24, when the activity level over the preceding 10 minutes exceeds the

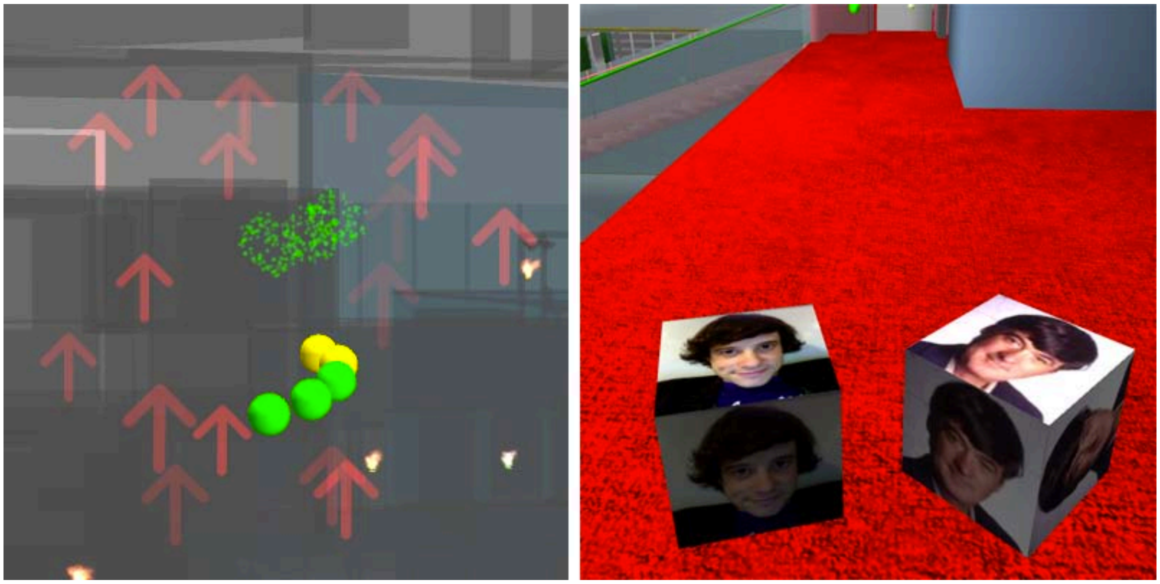


Figure 24: Social activity visualizations: at left, an audio level-meter ripples with the amount of coincident motion, while the typical levels float overhead; the arrows reflect an unusually high level of social activity that is trending upward. At right, RFID tags are detected and faces are shown.

typical value by some margin, the visualization spawns a cloud of arrows whose density and orientation reflects the trending social hot-spot.

DoppelLab makes use of a network of RFID readers distributed throughout the Media Lab, showing the faces of people carrying passive tags as they move through the building (Figure 24). In addition, Twitter messages appear in DoppelLab as they are broadcast, localized to individuals' virtual offices using an institutional directory. The position of a ping pong ball on a sensor-equipped table is shown as a continuous game of 2-d pong (Figure 25). Finally, DoppelLab incorporates external factors, like the position of the sun and the weather, under the assumption that it may impact the HVAC and other systems.

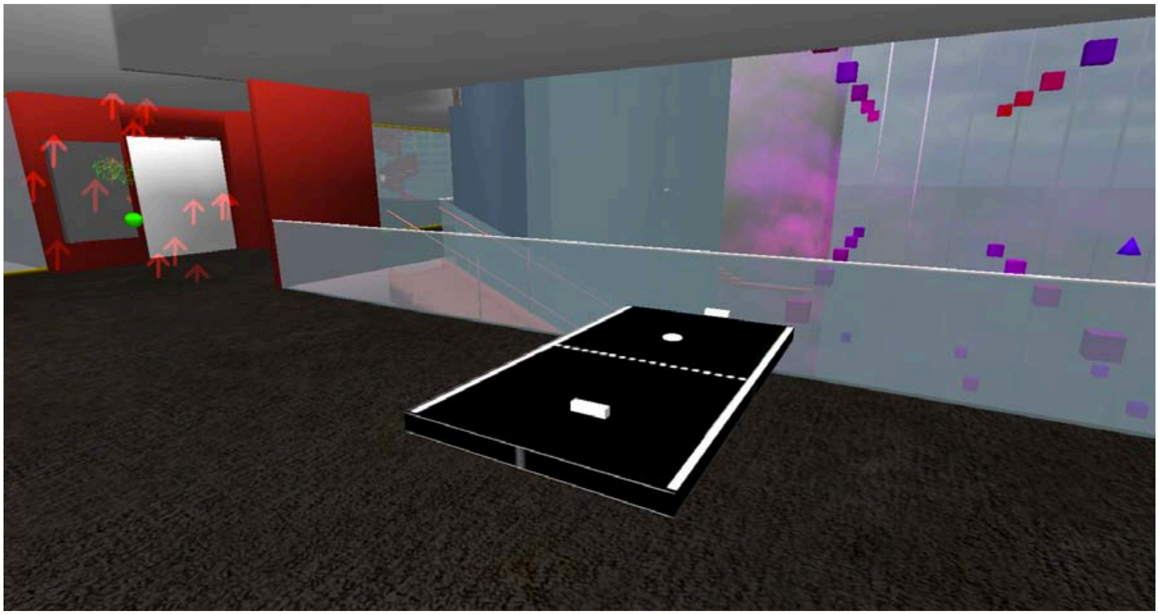


Figure 25: The position of a ball is collected from a sensor-enabled ping pong table {Ishii:1999vh}, delayed by one sample, and shown as a continuous game of 2-d pong.

4.2.3 Client Implementation

This work is catalyzed by significant advances in consumer 3-d graphics and new game engine development tools that make possible efficient rendering of complex, dynamic scenes and animations. The DoppelLab client is built on the Unity3D game engine, which makes use of Mono, the open-source version of Microsoft's .NET Framework. Development is primarily done by scripting in C# and UnityScript, a JavaScript-based language. The client can be compiled and run as a standalone binary, or embedded in a web browser. In the web browser, the client can communicate with the surrounding page elements using JavaScript, enabling additional, web-based 2-d interface elements that complement the 3-d DoppelLab view. We have begun to use this functionality to provide interface instructions that adapt to the user’s context, as shown in Figure 27.

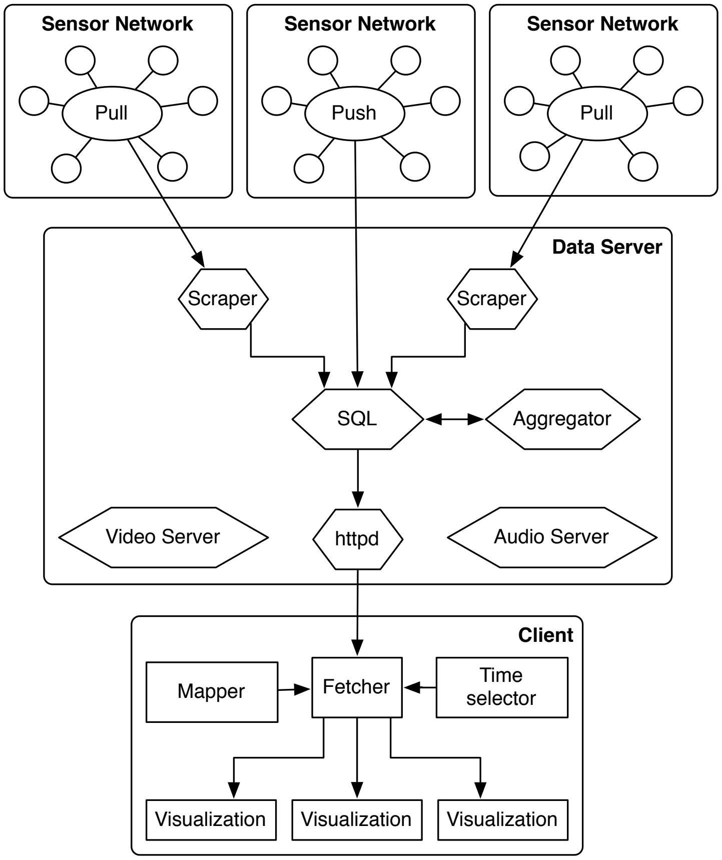
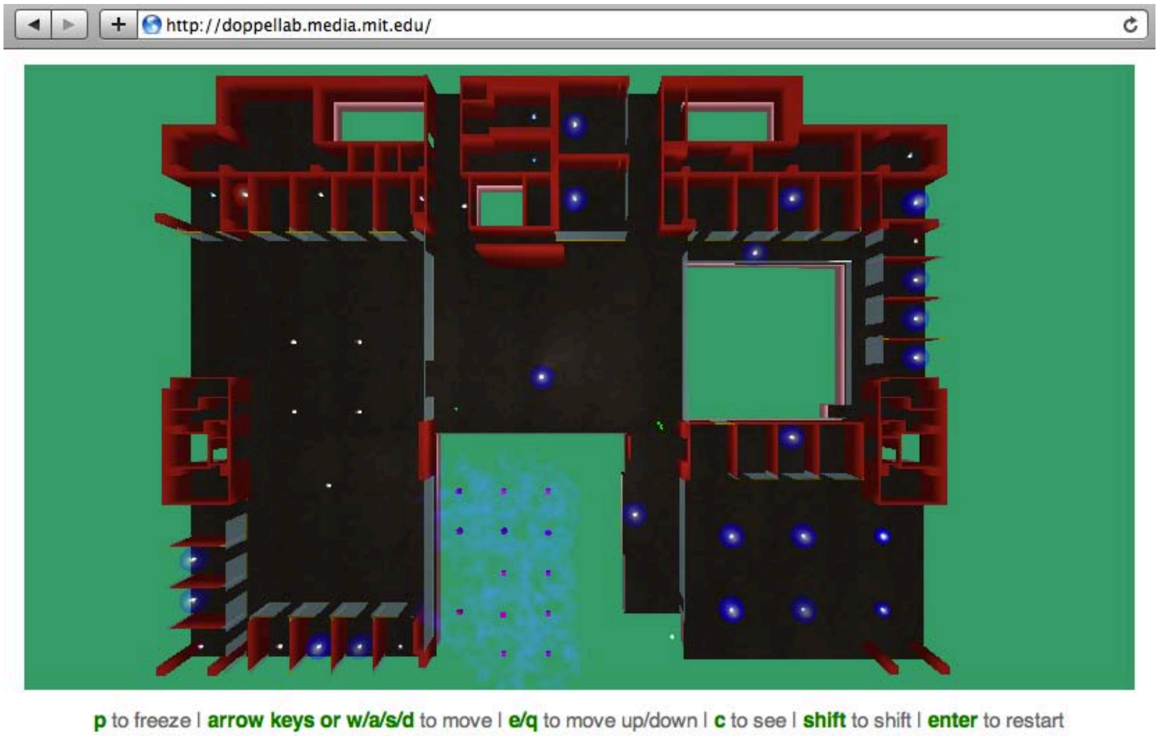


Figure 26: DoppelLab system diagram, showing its hierarchical structure and modular design for a streamlined development process.

As shown in Figure 26, the system is organized hierarchically and designed for modularity; on the client, centralized scripts deal with server communication and parsing, and pass messages to lower-level scripts that manage visualizations of each network. Centrally, the update loop periodically polls the data server, and the server responds with an XML file containing the requested data. If no data is found for any given sensor, the corresponding animation is disabled. Lower-level scripts that manage visualizations and animations use the data to control properties of objects in the environment. These parameters are visualization dependent, and include object properties like color, shape and size, or in the case of particle systems, systemic properties like emission rate and lifetime.



p to freeze | **arrow keys or w/a/s/d** to move | **e/q** to move up/down | **c** to see | **shift** to shift | **enter** to restart

Figure 27: Top-view (floor plan view) in DoppelLab. Also shown: application is embedded in a web browser, with identical functionality. In the web-based view, HTML instructions below the window change depending on user context, communicating with the application through JavaScript.

The client provides a facility for relating the local coordinate system to the building floor-plan, which scripts make use of to push data to the right places. The building coordinate system can also be converted to an approximation of the geographic latitude and longitude (given two fixed anchor points). This allows DoppelLab to interface with global positioning and other systems that use that standard.

Users interact with the client by controlling a set of cameras that can move throughout the virtual space, much like a first-person perspective video game. Cameras can be placed anywhere, and can be made fixed or mobile. In its current form, DoppelLab makes two types of views available: the main, first-person perspective depicted in many of the figures above, and a top-view that resembles a

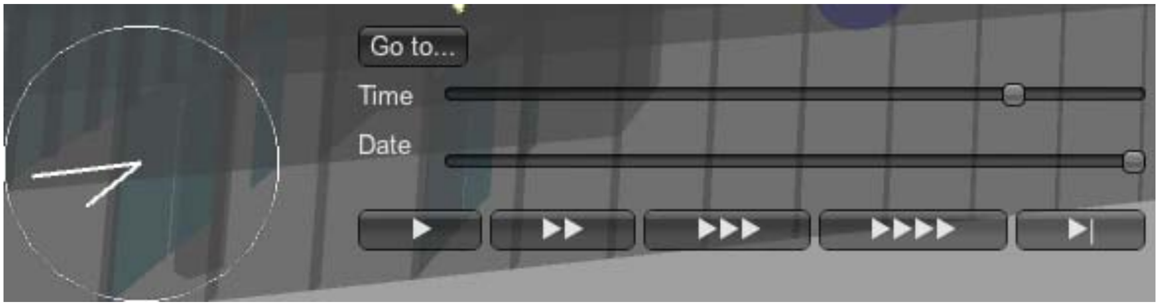


Figure 28: Interface for exploring historical sensor data; the top slider controls the time over a 24 hour period, and the bottom slider sets the date, going back as far as data is available. The buttons set the rate of playback, and the clock reflects the current simulation time and rate.

floor plan, shown in Figure 27. The top view shows the same visualizations and animations as the main camera, but within an isolated slice of the building.

In addition to streaming in real-time, data are also stored, aggregated, and analyzed on the server, enabling users to speed through weeks or months of data in minutes. We created two preliminary interfaces to this functionality on the client side, supporting a simple (exploration) mode and an advanced mode. The exploration mode, shown in Figure 28, is modeled on video editing timelines, where the top slider sets the system time within a 24 hour range, and the bottom slider moves that range across a span of days and months. Data can be played back at 4 different time scales—real-time, 1 minute per second, 10 minutes per second, and 1 hour per second. The sliders advance with the time, together with an analog clock. The advanced mode supports input of specific times into a text box.

4.2.4 Server and Database Architecture

The data used by DoppelLab is generated by a large number of independent networks. To provide the cross-network aggregation and analytics necessary for the client, the DoppelLab platform architecture includes a server that brings together data from these independent networks, storing it in an SQL database. This provides the client access to both current and historical sensor data from many different sensor networks in a generic and accessible format.

Initially, DoppelLab was designed to follow a fully distributed data collection model, where each client would independently connect to all the sources it required in an asynchronous, threaded fashion. This architecture avoids replicating data already stored for the purposes of the independent networks, and requires no single dedicated server. However, there are many problems with this approach. First, it requires that requests for data be made from many different sources at once, resulting in client performance that depends on the operating state of many independent networks servers. But more importantly, distributed data collection makes any depth of analytics on aggregated data difficult or impossible. Moving to a dedicated server model, we developed a database schema that could scalably accommodate, analyze, and serve large quantities of sensor data.

The DoppelLab database distinguishes between two types of sensor data, *samples* and *events*. *Samples* are numeric values produced by a sensor at regular intervals, and are stored as a triplet of a unique sensor ID, a timestamp, and the numeric value. *Events* occur irregularly, and contain larger amounts of data, generally in the form of strings. These data are stored as time-stamped blobs of JSON combined with event type, balancing drive space considerations with data accessibility. Currently, events include the appearance of a tag (with an associated username) at an RFID reader, or a message posted to a Twitter feed.

Metadata about each sensor is also stored in the database, including the sensor's type and location. Sensors are organized into groups, generally by their physical connection on a single node, facilitating higher-level queries. A thermostat, for example, has both a temperature sensor and a set-point, and a node in a sensor network might have a collection of various sensors. We are also beginning to implement facilities for sensors with changing locations, where the location of a sensor is itself a sensed value. The locations of mobile sensor groups are stored in a separate table with group IDs, transformation matrices, and time-stamps. This allows for efficient queries for sensor group locations that can be made separately from queries for the individual samples. Finally, a facility for relating people to sensor groups and person-specific sources, like Twitter or Facebook, enables analytics and representations that relate people, location, and sensor data. The people table also allows the system to keep track of usage and preferences when users are logged in, relating them to physical locations and relevant representations.

Scripts on the server periodically request the data from each source and store the values in a generic format in the database. Simultaneously, the server computes hourly averages, which are cached in a separate samples table with the same structure as the original. This aggregation permits fast execution of queries over longer timescales, such as “What is the average reading on this sensor at 3 PM on Mondays?”, or “What days of the week tend to have the highest average motion activity at noon?” Ad hoc queries for aggregate data can be made as well, though queries for averages or other analytics over a longer period than 10 minutes can take some time, given the volume of data in the database. This means that where real-time, fast frame-rate visualization is concerned, the client is limited to queries that perform computations over a window of up to 10 minutes and any longer than 1 hour (using the hour-aggregated samples table), a compromise that seems to account for any of the relevant time windows we have encountered. Within those time windows, however, the computation is fast enough to provide the client with sample-integrating playback speeds between 1x and 600x, and anything greater than or equal to 3600x.

The DoppelLab client accesses the data stored on the server and requests computation via HTTP requests, as described in the last section.

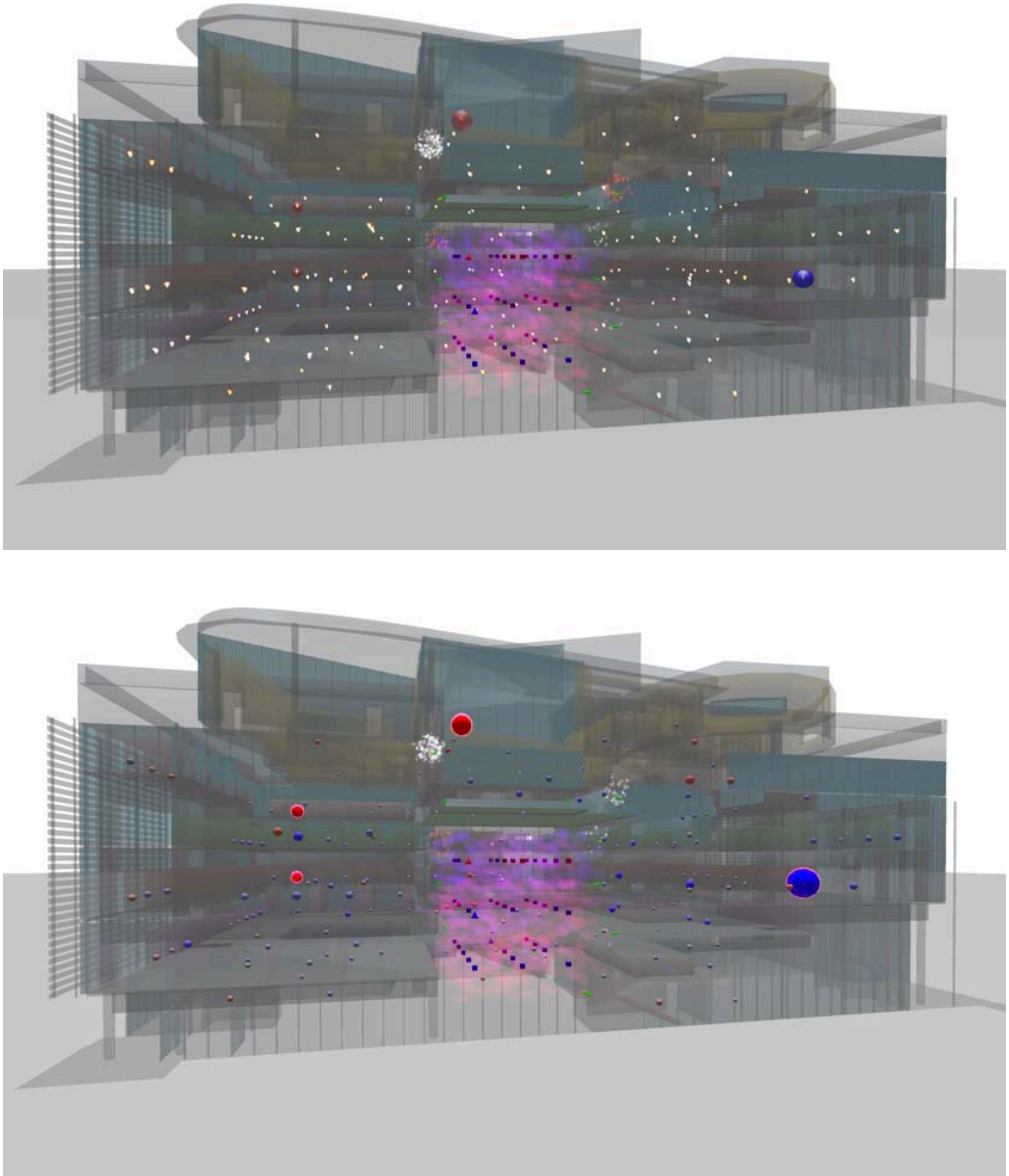


Figure 29: Top: the un-shifted view – absolute thermostat temperatures with faults highlighted. Bottom: the “shifted” view – difference from set-point as relative sphere size.

4.3 Applications

Deeper motivations notwithstanding, DoppelLab has several applications near its present form that have come out of the development process and increasing exposure to a variety of people and industries. The next section discusses an application design affordance provided by DoppelLab that might adapt the depth of information in a representation to a user's role. In addition, two specific applications, gaming and building information management, are briefly discussed here. The facilities management application is discussed further in chapter six, with preliminary industry assessment.

4.3.1 Applying Data Visualization in DoppelLab

Designers of applications for personal data visualization face a number of trade-offs setting the level of detail in the visual representation provided to a user. Two of these are of particular concern here. First, a trade-off between saliency and completeness balances a need to provide information relevant to a targeted query or geared towards a specific application (HVAC management, for example) with the goal of communicating the broader context. Second, many applications designed for non-expert users choose ease-of-use through simplified abstraction at the cost of immersion through expertise.

Consolidating a suite of energy-monitoring sensors into the health of a virtual flower or the mood of a smiley-face [54], for example, may suit a first-time or casual user, but does not necessarily facilitate a deeper engagement with the data. This trade-off becomes particularly interesting when the application seeks to move beyond the encouragement of a user's awareness of some context (their energy consumption relative to some shared target, for example) towards the formation of a hobby specific to the detailed monitoring and management of that context. Both could be called persuasive interfaces, but the latter requires a higher level of expertise on the part of the user. That expertise can in turn propel a longer-term interest in exploring and tracking the data. DoppelLab provides a means for engaging with users in this way, tracking usage over time and adapting the representation to match a user's level of engagement.

4.3.2 Gaming

DoppelLab can serve as a platform for building games that leverage its relationship to densely distributed sensor networks for cross-reality, where physical and virtual players collaborate to achieve tasks. Past research has explored this domain, in work that has tended to take the form of either small scale, tabletop display-based games, like Drew Harry's Stiff People's League [55], or city-wide, multiplayer games like Pac Manhattan [56]. Benford, et al. explore this space in [57]; generally, these games have taken place in large, outdoor environments, using GPS, mobile devices, and WiFi.

Combining densely distributed sensors and simple scripting of sense-able goals, DoppelLab enables indoor, multiplayer game scenarios on a larger scale than the tabletop interfaces and a smaller, denser scale than the GPS-based games, taking advantage of a much wider variety of physical world input modalities. A DoppelLab game would find users through mobile phones, distributed digital signage, and web browsers, adapting its representation of goals and progress to the medium. To explore these possibilities, we built a prototype cross-reality, task-oriented game using DoppelLab, in which virtual players see animated representations of simple, real-world tasks that they then communicate to their physical teammates. These tasks are built around sensor nodes, and include moving objects from one part of the building to another and making sound in specific areas.

Whimsical animations, like a carnival-style “high striker,” are used to show progress through each task. Successful completion unlocks capabilities, like “x-ray vision” in the virtual world, leading to subsequent levels. More broadly, this application seeks cooperation between remote users (who have access to rich sensor data streams but cannot act on them) and local users (who may not have all the information available, but can physically actuate the environment). This problem extends far beyond gaming, to critical applications like construction site safety and others.

4.3.3 Building Facilities and Information Management

The spatial arrangement of data in DoppelLab mimics physical infrastructure, making it a natural candidate for the application of facilities and building information management. In this application, experts monitor the real-time state of complex building systems that interact with inhabitants, like HVAC, lighting, and networking, looking for obvious faults as well as broader systemic inefficiencies. DoppelLab was developed with this application in mind; an early result was the exposure of a previously unknown fault in the building HVAC system. Rapid prototyping of inference rules and accompanying visual representations provide an ideal means for fault-detection using the platform.

Beyond inference rules, the layering of synchronized data from multiple building systems can expose relationships and inefficiencies in these systems; for example, the lighting system in the Media Lab tracks occupancy, data that clearly shows building inhabitants remaining in place long after the HVAC system has shut down for the night. Similarly, a dense network of temperature and humidity sensors used to test fluid dynamics models of large atrium spaces may have relevance to HVAC control, or vice versa. Only in the layering of these disparate sources can the extent of these relationships become clear.

5 | Interfaces to Sensor-driven Context

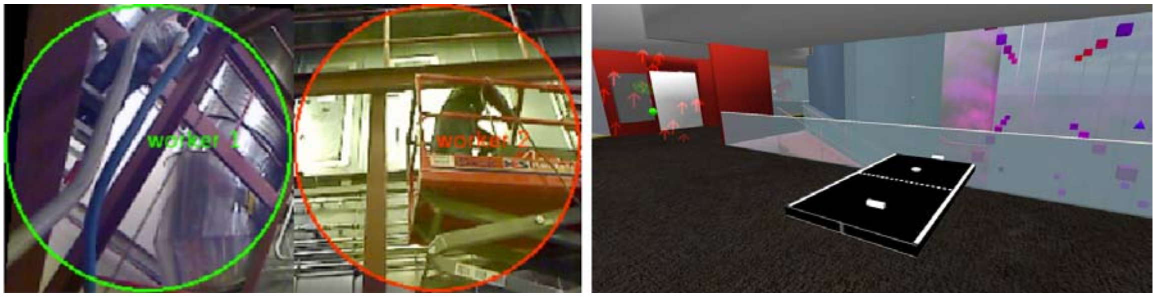


Figure 30: TRUSS and DoppelLab: what do these interfaces have in common?

This chapter condenses the body of work presented the last chapters into a set of features and design criteria, for comparison and insight. On the back-end, these systems stand to gain from a shared infrastructure that manages storage and analytics, while providing for synchronized, low-latency data sharing. On the client side, the interfaces take very different forms; this chapter extracts shared design parameters and seeks conceptual common ground. Generalizing from the specific examples, this chapter conceives of a framework for designing interfaces to sensor network data that support and encourage open-ended exploration though interactive visualization.

Figure 31 summarizes a set of interface design parameters and affordances distilled from the projects presented in earlier chapters, cataloging each interface's modes of interaction, treatment of perspective, range of a user's perception (referred to as aura [58]), modularity of representation, and relationship to physical sensing, as well as the roles of users and their relationships to remote context.

	Flurry	TRUSS	DoppelLab
users	subjects	remote experts	subjects & experts
interface	natural, incorporative	2-d buttons & sliders	2-d & 3-d 1st person
perspective	no natural perspective	mixed 1st person	1st person
aura [58]	limited, field of view	extended by sensors	extended, interactive
representations	fixed	fixed, interactive	modular
modalities	video	video, fixed sensing	open-ended
sensing	unimodal	multimodal	multimodal

Figure 31: Table of interface parameters

Flurry presents a natural and incorporative interface, where a user's exploration happens in exchange for participation as a subject. This creates a positive feedback loop, where user interest drives content and engagement, and vice versa. In this way, Flurry excels as a tool for driving new interest in the sensor network, demystifying the video network's privacy-threatening behavior by blending content and functionality; Flurry's medium is very much the message [1].

Figure 30 shows screenshots from the TRUSS interface and DoppelLab for comparison. At left, the figure highlights a feature of the TRUSS system that extracts workers from their surrounding context and collapses the physical space between them to form a new view, while simultaneously augmenting the resultant video with information from workers' wearable sensor nodes. In the absence of the sensor data, it would be impossible to identify a worker's context. This process is designed to highlight user-selected axes of contextual information, such as each worker's altitude or surrounding gas concentrations, and through open-ended interaction enable exploration of these specific parameters in the new, de-contextualized space.

On the right, Figure 30 shows a screenshot from DoppelLab, depicting the third floor atrium of the Media Lab. DoppelLab transforms a model of the physical building into a virtual world devoid of context, and then selectively inserts layers of information from the real world. As these layers populate the model, a unique view of the activity in the building begins to form, where distributed sensors act as atomic portals into the systems and activities in the physical space. Taken together, these portals tell a story; in the figure, for example, a group of people appears to be forming in the atrium around a fast-paced game of ping pong. In the background, a network of temperature and humidity sensors passively sample the environment in which this activity is unfolding.

By *selectively re-contextualizing* these de-contextualized scenes with information from sensors, both interfaces produce hyper-mediated perspectives that guide users through the narrative dots constituted by the sensor data; in the absence of any other context, these dots form stories like the one taking shape above. Both interfaces map these fixed parameters onto virtual *canvases* that are removed from but still tied to some physical world setting. These depictions do not purport to be copies of the world, but instead form explicitly mediated, impressionistic portrayals in response to user input.

For TRUSS, interactivity in the interface means that expert users can explore the limited axes of context and choose multimodal combinations to compose a view that makes their choices salient. For DoppelLab, interactivity functions in a much more open-ended way, where narratives unfold over time, through exploration. Playing through a day in a minute reveals students' late-morning arrivals and late-night departures, as well as regular socializing that tends to occur in the late afternoon; these events happen against a backdrop of a rising and setting sun that significantly heats the atrium in the morning, and an HVAC system that turns on long before people arrive and turns off long before they leave for the night. The visualization reveals faults in these systems, and makes visible their impacts on building inhabitants.

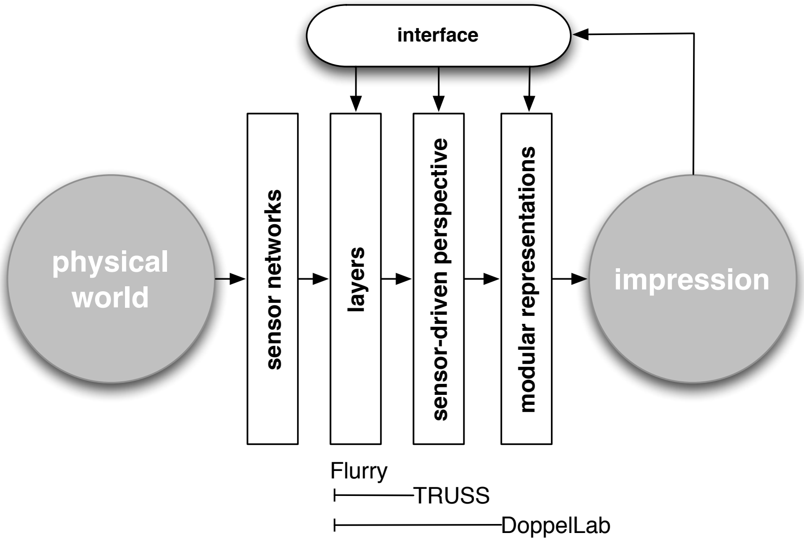


Figure 32: Sampling the world to form a re-contextualized impression in an interactive interface.

The modularity of representation in DoppelLab adds a third layer of interactive control, where the impressions themselves can be swapped and adapted. Alternate representations can be expressions of whimsy, like accumulating musical notes or bowler hats with cigars (Figure 21); but they can also help explore the data, and facilitate new, creative ways of seeing the world.

In these interfaces, sensor fusion happens not only in the background, at the sensor data level, but also on a perceptual level, at the network scale—in the perspectival layering of multi-sensor, multi-network data streams on the client. While this fusion is not algorithmic, and no inference is given, it is visually analytic, and provides the grounds for understanding the relationships between data streams. This is a critical step in the analytic cycle; once these relationships become visually apparent, the environment provides a means for building the statistical data fusion, and for visually representing its results.

6 | User Testing and Observation

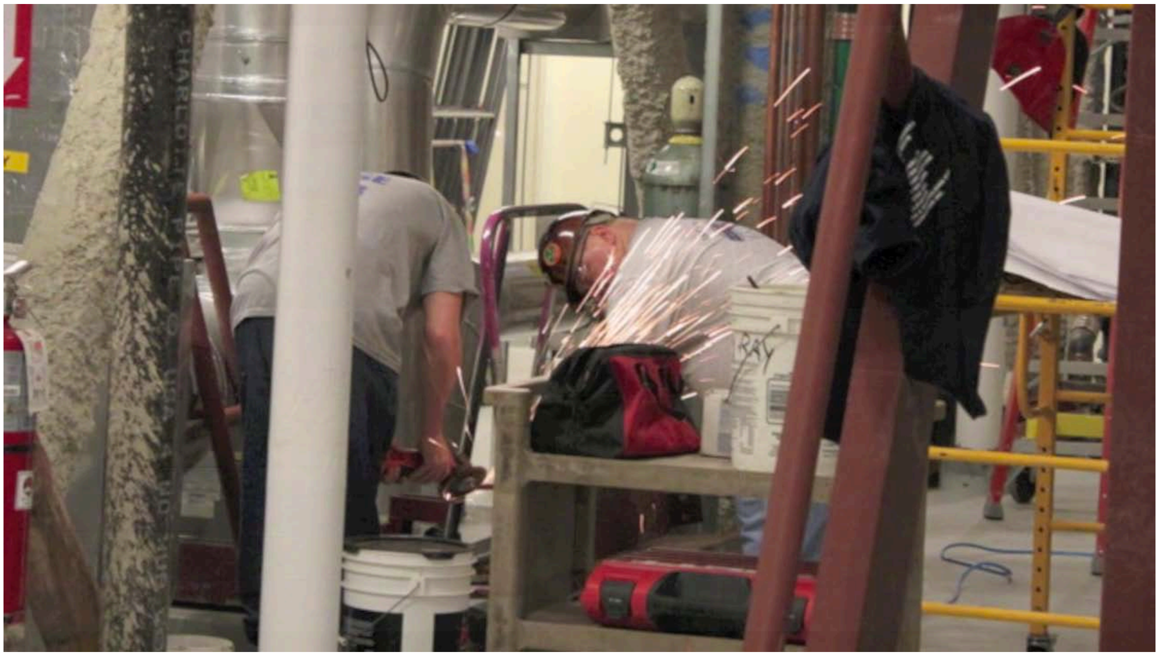


Figure 33: Workers wearing sensors while participating in a user study during construction.

User testing of the work presented in this thesis was performed in a number of ways, depending on the application. The TRUSS system was evaluated through a real-world system deployment and human subjects testing of the hardware and software for initial data collection (depicted in Figure 33), as well as through interviews with construction industry experts about the results. Evaluation of DoppelLab was done through several workshops with novice users as well as interviews with building facilities managers. DoppelLab has also been scrutinized through exposure to industrial research partners, who have taken an interest in its development and provided critical feedback related to their application spaces. Further evaluation of both systems is planned, with more details in chapter seven.

6.1 TRUSS for Safety

We deployed a prototype of the TRUSS system described in Chapter 3 for data collection from workers on an active construction site. For a period of approximately two weeks, sensor base stations and video devices were installed in the construction area and three steelworkers were instrumented with wearable sensors. During this time, the workers erected several large catwalk structures surrounding a set of 2-story air handlers on the penthouse floor of a building under construction. The workers' primary activities included arc welding, cutting steel, and carrying heavy material, all while using ladders, lifts, and cranes for rigging heavy steel frames and platforms.

The system components were deployed early each morning and collected at the end of each day for analysis; on several occasions a base station was left for several consecutive days, having been mounted at height by the workers themselves. To simplify deployment, the base stations were mounted magnetically to the air handlers, allowing the workers to move them with little effort as the work area shifted day to day. The wearable sensors, also distributed and collected daily, were each assigned to a specific worker at the start of the user study. The embedded computer and magnetically-mounted camera system were set up to be triggered and controlled wirelessly, and enabled remote access from offsite, LAN infrastructure permitting.

We encountered several major problems during the deployment, both systemic and environmental. Embedded software instability caused occasional system hang-ups. In addition, an unexpected amount of heat radiating from the gas sensors caused noise on the barometric pressure sensor. The environmental challenges included extreme, bit-scrambling electric fields caused by welding, as well as clouds of errant conductive metal filings falling into circuit vents and causing shorts. In order to allow air to flow freely into the gas sensors, the boards were left mostly uncovered, resulting in at least one catastrophic short due to worker sweat (that worker later received a better-sealed device). Taken together, these problems resulted in extremely spotty data; the number of variables in the system was high enough that some module was failing nearly all the time. For example, on a day when the wearable sensors functioned perfectly, metal filings shorted the camera node. With little experience designing hardware and software to stand up to the challenging environment of a real construction site, some of these failures were inevitable. In the end, we collected enough data to prove concepts and establish correspondence between activities and measurements, but not enough to build robust models. The next version of the hardware, discussed in chapter seven, addresses many of these problems. Further deployments with this new hardware are imminent.

There was some concern that workers would view the sensors and cameras with suspicion, as tools of surveillance, and resist our study. We were clear, under the terms of our human subjects committee application, that workers had absolutely no obligation to participate, and could terminate the study at any time, and were surprised to find workers not only agreeable, but extremely receptive to the research and supportive of its aims. The workers reported concern for their own safety, and a hope that real-time sensing could help them better understand and respond to their context in real time. Indeed, it was a construction manager who expressed the most concern for this issue, but characterized it as a problem limited to the small number of workers who would already be looking for ways to thwart the system by taking unsafe shortcuts. We did not encounter this attitude in our study.

After the deployment, a construction industry expert was shown the TRUSS interface, and expressed several important concerns and recommendations, as well as interest in further study to be conducted on new sites. He reported that worker behavior and state of mind are major risk factors that we are not presently considering, giving the example of a worker who may be frustrated by heavy traffic on their morning commute, resulting in reckless behavior later in the day. Still, he found our notion of a personal safety bubble compelling and useful, and suggested that we form similar bubbles around objects and machinery. He noted that the interface makes some axes of otherwise invisible context clear to users, and could be useful for training. He also noted that 80% of safety management is prevention and 20% field control, and expressed that TRUSS could fit well in both. Finally, he asked if there was some way he could see the whole site in a macroscopic way that would expose faults and other points of interest for further exploration through a more detailed interface like TRUSS. Shown DoppelLab, he strongly suggested that the two interfaces be integrated.

6.2 DoppelLab

User testing of DoppelLab has come about through several channels, including several visualization development workshop and brainstorming sessions, interviews with building facilities managers, and critical feedback from a number of engaged industrial partners.

6.2.1 Workshops

Several workshops were run with programming novices from various industries, during which simple sensor-driven animations were developed live with creative input from the audience; some participants followed along with the development process on their laptops, while others watched. The workshop leveraged several of

DoppelLab’s facilities for rapid prototyping of visualizations, including drag-and-drop animations and script frameworks. The resulting visualization mapped audio levels from a sensor node in the room to the energy of a bouncing ball in DoppelLab. Users were able to follow along and understand the modules for scraping and parsing data, designing a visual representation in the game engine, and mapping the data to some parameter of the animation. Afterwards, participants were split into groups to brainstorm new visualizations and applications specific to their industries. Topics and suggestions included health care, towards management of hospitals and assisted living facilities, sociometric analysis for office productivity assessment, and the control and visualization of data center traffic.

6.2.2 Building Facilities and Information Management

Another channel for evaluation has come through interviews with a number of professional building facilities managers. The most consistent feedback in that space regards DoppelLab’s 3-d, architecturally-linked organization of data. Managers found DoppelLab particularly useful for understanding how HVAC faults propagate through adjacent rooms, and for discovering whether inhabitants’ complaints relate to repairable faults or individual comfort preferences. Feedback from data center managers reflects similar concerns; DoppelLab’s presentation of data on the building model facilitates investigation into the spatial propagation of faults like network congestion or overheating.

One result of the interview process was the discovery that building facilities managers are generally much less less concerned with inhabitant comfort than with system faults. One manager suggested that DoppelLab could be used to prove to uncomfortable building inhabitants that the systems are perfectly functional. This leads to an avenue for future work, discussed in the next chapter, where

DoppelLab could be used to mediate between building inhabitants and managers; the parties would share data to track both subjective comfort (through user reporting) and system state (through set point and actual temperature), to better understand how these quantities might relate to each other.

6.2.3 DoppelLab Game

A test of the DoppelLab game described in section 4.3.1 was run with guests during a Media Lab event, but the game failed to generate much user interest. The tasks were contrived, with little reward for either team member. More importantly, the game relied on active communication by phone between players, resulting in awkward and irritating gameplay for both parties. Future work will investigate other means communication between participants, through mobile phones, distributed digital signage, and web browsers, adapting its representation of goals and progress to the medium.

7 | Ongoing and Future Work

The projects presented in this thesis are continuing, and new work is already underway. This chapter presents avenues for near-term development, as well as broader plans for future work. In the long-term, we seek to integrate technologies and concepts from the disparate projects in this thesis into new platforms and interfaces that combine multi-perspective video, 3-d animation and visualization. Plans for future work are treated separately for each project in the subsequent sections, and holistically in the last section of this chapter.

7.1 TRUSS for Safety

As discussed in the last chapter, a combination of hardware and software shortcomings and environmental challenges caused major problems with the data collection in the first deployment of the TRUSS system. Further deployments are planned using a new wearable sensor node, currently in active development. Because of the challenges we faced in the initial deployment, we have opted to perform further tests in the more controlled environment of a lab machine shop, before moving back to a real construction site next year. We are planning a user study that will use the new hardware to monitor pre-determined tasks involving welding, machining, water-jet cutting, and laser-cutting, all of which emit gases and particulates and involve risks to shop users. The controlled environment will allow us to watch the behavior of the sensors closely as we vary the conditions.

The new hardware in development addresses the challenges faced by the first-generation system. The safety daughter board will be re-used, but new badges improve on the older hardware with a radio that better supports RSSI localization and a new IMU section, as well as analog circuitry to perform envelope following and peak detection on the microphone. The barometric pressure sensor has been moved from the daughter board to the new badge, as it had been unpredictably drifting with heat from the nearby gas sensors, causing problems with the measurements. Finally, the new badge uses an 8-bit microcontroller that will simplify software development, compared to the 32-bit processor used by its predecessor. The package is much more compact (about the size of a pager), making it more comfortable for users to wear and less prone to falling from workers' belts. A small, simple radio base station has been developed that will enable dense instrumentation of the space with minimal effort, towards much better RSSI-based location services.

A major issue in the first deployment was the difficulty we faced working with the video and network bridge computers that we brought onsite. The nodes required ventilation, and shorted a number of times due to metal shavings falling into the open vents. To address this problem, we have begun working with very small form-factor Intel Atom-based embedded nodes that are significantly more robust to challenging environments than the system we used before. The node contains non-volatile flash memory (instead of a spinning drive), and is much better sealed from the environment. The system will be simpler to use and less intrusive to workers, as the new video nodes can be more tightly integrated with camera modules for rapid deployment and collection.

The machine shop targeted for our next deployment is also equipped with an existing networked camera system, used by shop managers to monitor occupant activity during off-hours. This network vastly streamlines our video deployment, which will not require as many video nodes to be brought onsite. The new system will use a tiered architecture, like the one shown in Figure 35, to manage the large number of streams. In addition, the fixed video infrastructure in the new space has far more extensive coverage from more useful fields of view than we were able to achieve in ad hoc deployment. On the user interface side, the increased video coverage provides an opportunity to test and improve on many of the ideas seeded by the proof-of-concept interface, including the video mixing application, which was not used in the first deployment.

More reliable data will also facilitate further sensor fusion research and testing. Real-time camera-IMU fusion is planned for the shop deployment, and may build on the work in [10]. Further down the line, integration of the TRUSS system with UWB radios that support precise time-of-flight ranging will close the correspondence loop, providing much more reliable video augmentation and enabling distributed appearance modeling on a large scale.

We will be deploying and testing the new hardware imminently. Moreover, the expert evaluation described in the last chapter represents the beginning of much larger plans to work with construction industry experts in new real-world deployments that will realize real-time sensing and workflow integration on active, contracted building sites at MIT.

7.2 DoppelLab

There are a number of research avenues being pursued for DoppelLab in the near-term, as well as a set of longer-term goals. The application is evolving quickly as we add sensors and features, as well as new buildings and environments, including more of our own instrumented living spaces. We are actively incorporating data from the new wearable devices and sensors described in the last section, including the actual TRUSS deployment, as well as building-wide network traffic levels and lighting state, which will provide much finer-grained information about individual activity and group behavior. On the client side, we are developing better facilities for the multi-user (multiplayer) use case, geared toward shared annotation of information and space, as well as joint exploration. The web-embedded instance of the client enables mixed 2-d and 3-d interface elements, which will allow for much more intuitive interaction and efficient representation that combines the macroscopic view that DoppelLab provides with more detailed analyses and data histories.

DoppelLab's architecturally-linked arrangement of representations provides an intuitive platform for making queries about people and their environments, but there are other 3-d spaces that would reveal hidden relationships between data along different axes. One avenue for research involves the development of new spatial arrangements that can be fluidly animated through transformations from the physically-linked starting point. Examples of this thinking include a 2-d slice of the building that would spread time out along the third axis, or a collapsed grid of representations that form small-multiples for comparison. Transformations would involve continuous motion from one arrangement to another, connecting the architecture to the newly created space, and providing a means for relating inferences across spatial representations.

Currently, DoppelLab's modular system simplifies the design and development process for visualizations and sonifications, as well as new back-end analytics. An extension of the efforts to facilitate rapid deployment of new interfaces is the design of a sensor data parsing and visualization markup language that will codify and further streamline this process. Already, a port of the scrapers, database management and analytics modules to the Python-based, SQL-wrapping toolkit SQLAlchemy [59] has streamlined the process for adding new sensor networks on the back-end. On the client side, code in development enables the generation of new data streams and visualizations in a highly structured manner. This work is leading to a new graphical user interface to development in DoppelLab which will turn the process partly into a point and click wizard, with automatically-generated code that can then be edited to fit users' needs, or in response to visual analytic discoveries.

Finally, further deployments and evaluations of DoppelLab are in the pipeline. The application is currently running 24/7 in a remote, corporate setting, as a real-time window into building systems and activity. More such deployments are planned, which are certain to generate application-specific feedback. We are in discussion with MIT facilities managers to test DoppelLab in their operations division, where building systems are centrally monitored and work dispatched.

7.3 The Disappearing Act

Driven by significant advances in computer vision, imagers are increasingly used as general sensors for everything from gesture recognition in human-computer interaction to assisted living. As such, privacy in pervasive camera networks has become a real concern, especially as such systems begin their first commercial deployments. Even outside the sphere of futuristic, sensor-rich ubiquitous media systems like [11], the unintended or incidental capture and electronic transmission of images and video of passers-by to social networks pose a major challenge to individual privacy, motivating new research in that space. While technologies like automated facial blurring, now standard for large commercial image databases like Google Street View, begin to address some of these issues, face blurring falls far short of privacy protection in semi-private environments like offices. The state of the art in individually configurable dynamic privacy in physical sensor networks, articulated in [46], is opt-out and catch-all, meaning that users who want privacy must carry a registered physical tag that enforces a total network blackout for everyone in its vicinity.

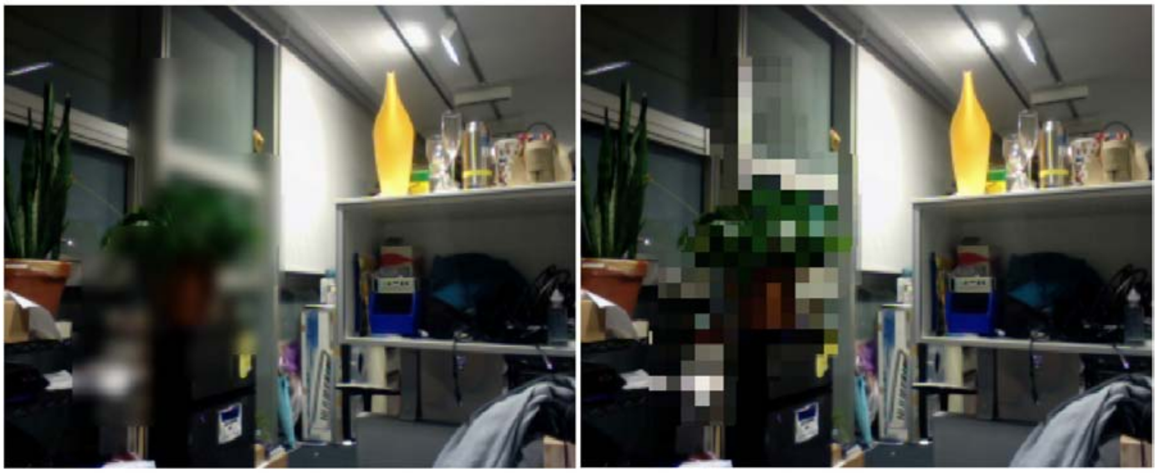


Figure 34: Two examples of privacy-preserving indications of presence using a lower-resolution static estimate of the background mixed with live video to mask a figure. On the left, down-sampling, Gaussian blurring, and interpolating back to the original resolution; on the right, down-sampling and up-scaling, with no blur.

We are developing a system that fuses fine-grained radio-location services with distributed video cameras, enabling users carrying radio tags to dynamically disappear and reappear in video streams, or disappear entirely by carrying no tag. This conceptualization points to an opt-in video privacy system, in which users explicitly authorize the transmission of their likeness by carrying a physical object, and everyone else is invisible; the system replaces their pixels with estimates of those of the background they are presently occluding. This filtering process takes place on the low-power camera node, preventing the transmission of users' images to further protect privacy.

An initial deployment of the system will take the form of an installation, called The Disappearing Act, that prototypes the user interactions and visual representations. On the back-end, the Disappearing Act fuses a commercial UWB-based radio-location system, called Ubisense [60], and a set of distributed video and computer vision nodes that are also tagged with radios, matching estimated object positions and sizes to compute correspondence. This camera-radiolocation fusion enables a new user interface to a mediated context, building on the Flurry and TRUSS systems, and serves as an early exploration into systems that use fine-grained radio-location to solve the video correspondence problem, and the implications of this development.

Reduced spatiotemporal resolution of the completed gaps may be acceptable, or even desirable as a visual marker of ghosts, as shown in Figure 34, where individuals are replaced by a lower-resolution or blurred version of the background. This kind of representation attempts to balance viewers' engagement and subjects' privacy concerns, providing some connection to the local context while still obscuring identity. Practically, this approach produces a more compelling overall video result when the background is changing. More broadly, this work also explores how the level of context might be parameterized to convey just as much information as can be reconstructed (where current algorithms might produce incoherent artifacts in the absence of sufficient prior information). Existing computer vision algorithms for video completion tend to succeed in finding an optimal solution or completely fail to do so, with little middle ground. One goal of this work is the exploration of the space between success and failure in rendering video through the use of selective blurring and an objective of information-conveyance over optimality of the completion.

The Disappearing Act is will be tested imminently within a single lab, with plans for a much larger scale deployment when building-scale location systems come online.

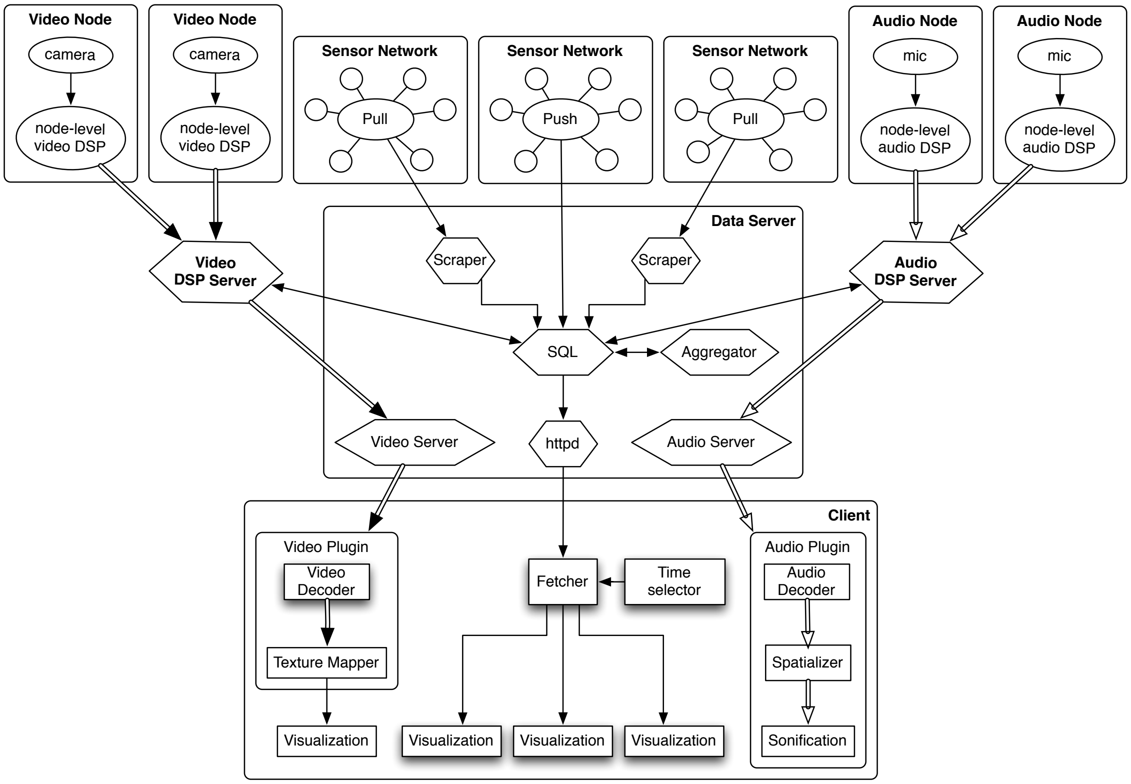


Figure 35: Fusing sensor data, audio, and video for an immersive interactive interface that would facilitate macro/micro scale exploration.

7.4 Fusing Perspectives

Conceptual common ground notwithstanding, the interfaces presented thus far fall into two distinct categories—augmented video and 3-d animation. The first is quite effective for communicating information through a relatively microscopic, fixed perspective, even collapsing that perspective further to form salient views. The latter provides a view into the events unfolding throughout a large space; while DoppelLab's first-person perspective enables close examination, evaluation with users confirms that the medium in its current form encourages distant, macroscopic assessment of state.

At the same time, we have developed mechanisms in DoppelLab for zoomable interface-like interactions [19] that would enable microscopic investigation into remote context, triggered by the suggestions in the macroscopic view. The next step in this research is towards a fusion of the affordances of the interfaces presented in this thesis, towards an interactive, immersive environment that facilitates micro and macro scale exploration of dense sensor data, audio, and video.

Figure 35 shows a system diagram for the front- and back-end combination of such diverse data streams, both for real-time, algorithmic data fusion and multimodal visualization. This work builds on and integrates past research like [31], which projects video into a 3-d virtual environment, as well as [61] and [62], which bring audio sources into the virtual space and call for low-latency, direct-manipulation interfaces in sonification, respectively.

Work already underway enables 3-d spatialization of privacy-protected audio streams within DoppelLab [51]. Future work will map video from a Media Lab machine shop onto virtual objects, incorporating the TRUSS for Safety interface into the DoppelLab framework. Depth cameras like the Microsoft Kinect will enable closer correspondence between radiolocation devices and cameras, as well as the direct mapping of video from physical surfaces to virtual planes. The tiered system architecture shown in Figure 35 facilitates both, for more immersive and intuitive user interfaces to these data.

8 | Looking Ahead

McLuhan critically and cryptically assesses the cultural impact of “instant speed” brought on by electricity [1]; in the electrical era, he suggest, from pure observation and by temporal association, a light switch causes light. Cognition is contextually situated; while ubiquitous sensor networks have brought instantaneous, densely distributed portals into the physical world, these portals operate on highly specific axes of context. Interfaces to sensor network data present the world along such axes, as if they form complete narratives outside of the physical context. How are these sensors and interfaces situating our cognition of the phenomena they observe?

This question motivates new thinking in cross-reality, where the virtual world need not take the form of pervasively shared representation, but rather looks different depending on one’s role in it. The shared thread becomes the underlying data, from whatever source will provide them, and the representation becomes fluid. By this logic, building facilities managers need not see the same representations as building inhabitants, if there are more effective ways to look at the data for that application. Some kind of representation of the same data is available to both, and the transformation from one space or representation to another reflects each party’s divergent priorities.

A user’s selection of modalities towards some impression explicitly situates their understanding of this sensor-driven context. If the virtual world maps only light intensity and switch sensors, the instantaneous, sensor-driven representation implies causality between lights and switches, like McLuhan suggests. If the virtual world includes a representation of the electrical current flowing through the wire, the story changes.

In this cross-reality, the representation of a thermostat reading available to a facilities manager is composed not only of a temperature and a set-point, but also a user-reported level of comfort gleaned from their wearable device, and the current occupancy of the space; an anomaly or fault in this system can be interactively defined to weight any combination of these sensed parameters, and take any number of visual or aural forms. These new cross-reality interfaces zoom from a macroscopic view of the spatial and systemic relationships in a building to a microscopic visual analysis of the behavior of a single sensor across a long period of time. The density of information in the representation matches the user's level of engagement, as well as the medium through which the engagement is made; the interface encourages user interest by revealing information interactively and responding to direct manipulation. This work imagines a transformative moment in ubiquitous computing, where applications built atop distributed sensor streams connect people to the rapidly emerging internet of real-time data.

References

[1] M. McLuhan, *Understanding Media: The Extensions of Man*. New American Library, 1964.

[2] T. Teixeira, G. Dublon, and A. Savvides, “A Survey of Human Sensing: Methods for Detecting Presence, Count, Location, Track and Identity,” *ACM Computing Surveys*, 2010.

[3] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, 2004.

[4] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893.

[5] K. Marzullo, “Tolerating failures of continuous-valued sensors,” *ACM Transactions on Computer Systems*, vol. 8, no. 4, pp. 284–304, Nov. 1990.

[6] Lin Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” in *International IEEE Symposium on Information Processing in Sensor Networks (IPSN 2005)*, 2005, pp. 63–70.

[7] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, and T. Starner, “Recognizing workshop activity using body worn microphones and accelerometers,” *IEEE Pervasive Computing*, 2004.

[8] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, “Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, 2006.

[9] T. Teixeira, D. Jung, G. Dublon, and A. Savvides, “PEM-ID: Identifying people by gait-matching using cameras and wearable accelerometers,” in *IEEE International Conference on Distributed Smart Cameras*, 2009, pp. 1–8.

[10] T. Teixeira, D. Jung, and A. Savvides, “Tasking networked CCTV cameras and mobile phones to identify and localize multiple people,” in *Proceedings of the 12th ACM international conference on Ubiquitous computing (Ubicomp '10)*, 2010.

[11] M. Laibowitz, N.-W. Gong, and J. A. Paradiso, “Multimedia Content Creation using Societal-Scale Ubiquitous Camera Networks and Human-Centric Wearable Sensing,” in *Proceedings of the international conference on Multimedia (MM '10)*, New York, New York, USA, 2010, p. 571.

[12] J. A. Paradiso and J. A. Landay, “Guest Editors' Introduction: Cross-Reality Environments,” *IEEE Pervasive Computing*, vol. 8, no. 3, pp. 14–15, 2009.

[13] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE Transactions on Information Systems*, 1994.

[14] J. Lifton and J. A. Paradiso, “Dual Reality: Merging the Real and Virtual,” *Facets of Virtual Environments*, 2010.

[15] J. Lifton, M. Laibowitz, D. Harry, N.-W. Gong, M. Mittal, and J. A. Paradiso, “Metaphor and Manifestation—Cross-reality with Ubiquitous Sensor/Actuator Networks,” *IEEE Pervasive Computing*, vol. 8, no. 3, Sep. 2009.

[16] M. Laibowitz, N.-W. Gong, and J. A. Paradiso, “Wearable Sensing for Dynamic Management of Dense Ubiquitous Media,” in *Proceedings of the IEEE Body Sensor Networks Conference (BSN '09)*, 2009.

[17] D. F. Reilly, H. Rouzati, A. Wu, J. Y. Hwang, and J. Brudvik, “TwinSpace: an Infrastructure for Cross-Reality Team Spaces,” in *Proceedings of the 23nd annual ACM symposium on User Interface Software and Technology (UIST '10)*, 2010.

[18] C. Ware, *Information Visualization, Second Edition: Perception for Design*, 2nd ed. Morgan Kaufmann, 2004, p. 486.

[19] A. Woodruff, J. Landay, and M. Stonebreaker, “Constant information density in zoomable

interfaces," in *Proceedings of the working conference on Advanced Visual Interfaces (AVI '98)*, 1998.

[20] B. Bederson, J. Hollan, K. Perlin, and J. Meyer, "Pad++: A zoomable graphical sketchpad for exploring alternate interface physics," *Journal of Visual Languages and Computing*, 1996.

[21] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Computer Graphics and Applications*, vol. 26, no. 1, pp. 10–13, Jan. 2006.

[22] J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization," in *IEEE Transactions on Visualization and Computer Graphics*, 2007, vol. 13, no. 6, pp. 1–8.

[23] M. Lewis and J. Jacobson, "Game engines in scientific research," *Communications of the ACM*, vol. 45, Jan. 2002.

[24] G. Brown-Simmons, F. Kuester, C. Knox, S. Yamaoka, and D. Repasky, "Kepesian Visualization," in *Proceedings of Connectivity, the Tenth Biennial Arts and Technology Symposium*, New London, CT, 2006, pp. 25–36.

[25] F. Kuester, G. Brown-Simmons, C. Knox, and S. Yamaoka, "Earth and Planetary System Science Game Engine," *Transactions on Edutainment II*, vol. 203, pp. 203–218, 2009.

[26] B. Kot, B. Wuensche, and J. Grundy, "Information visualisation utilising 3D computer game engines," *Proceedings of the ACM SIGCHI conference on Computer-human interaction (SIGCHI '05)*, 2005.

[27] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*, 1st ed. A K Peters, Ltd., 2005.

[28] S. Feiner, B. Macintyre, and D. Seligmann, "Knowledge-based augmented reality," *Communications of the ACM*, vol. 36, no. 7, pp. 53–62, Jul. 1993.

[29] S. Mann and J. Fung, "VideoOrbits on Eye Tap devices for deliberately Diminished Reality or altering the visual perception of rigid planar patches of a real world scene," in *Proceedings of the International Symposium on Mixed Reality (ISMR '01)*, 2001.

[30] J. Herling and W. Broll, "Advanced self-contained object removal for realizing real-time Diminished Reality in unconstrained environments," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR '10)*, 2010.

[31] I. O. Sebe, J. Hu, S. You, and U. Neumann, "3D Video Surveillance with Augmented Virtual Environments," in *ACM SIGMM international workshop on Video surveillance (IWVS '03*, New York, New York, USA, 2003, pp. 107–112.

[32] P. C. McLean, "Structured video coding," *S.M. Dissertation, Massachusetts Institute of Technology, Media Arts and Sciences*, 1991.

[33] E. Elliott, "Multiple Views of Digital Video," *Technical Report, MIT Media Laboratory Interactive Cinema Group*, Mar. 1992.

[34] L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *Proceedings of the first ACM international conference on Multimedia*, New York, 1993.

[35] C. D. Correa and K.-L. Ma, "Dynamic video narratives," *ACM SIGGRAPH 2010*, 2010.

[36] J. Lifton, M. Mittal, M. Lapinski, and J. A. Paradiso, "Tricorder: A mobile sensor network browser," in *Proceedings of the ACM SIGCHI conference on Human Computer Interaction*, 2007.

[37] M. Mittal, "Ubicorder: A Mobile Interface to Sensor Networks," *S.M. Dissertation, Massachusetts Institute of Technology, Media Arts and Sciences*, 2008.

[38] M. Mittal and J. A. Paradiso, "Ubicorder: A Mobile Device for Situated Interactions with Sensor Networks," *IEEE Sensors Journal*, 2011.

[39] T. Sohn and A. Dey, "iCAP: An Informal Tool for Interactive Prototyping of Context-Aware Applications," in *Extended Abstracts on Human factors in Computing Systems (CHI '03)*, New York, New York, USA, 2003, pp. 974–975.

[40] A. Bamis and A. Savvides, "STFL: a spatio temporal filtering language with applications in assisted living," in *Proceedings of the ACM 2nd International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '09)*, 2009.

[41] "Google PowerMeter - Save Energy. Save Money. Make a Difference.," *Google PowerMeter*. [Online]. Available: http://www.google.com/powermeter/about/. [Accessed: 29-Jul.-2011].

[42] "Conserve Energy, Save Money - Microsoft Hohm," *Microsoft Hohm*. [Online]. Available: http://www.microsoft-hohm.com/. [Accessed: 29-Jul.-2011].

[43] *Phillips DirectLife*, *Phillips DirectLife*. [Online]. Available: http://www.directlife.philips.com. [Accessed: 01-Aug.-2011].

[44] *Microsoft Hohm*, *Microsoft Hohm*. [Online]. Available: http://www.microsoft-hohm.com. [Accessed: 01-Aug.-2011].

[45] J. W. Davis and A. F. Bobick, "The representation and recognition of action using temporal templates," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97)*, 1997.

[46] N.-W. Gong, M. Laibowitz, and J. A. Paradiso, "Dynamic privacy management in pervasive sensor networks," in *Ambient Intelligence*, 2010.

[47] P. Angove and B. O'Flynn, "Air-quality Monitoring for Pervasive Health," *IEEE Pervasive Computing*, vol. 9, no. 4, pp. 48–50, 2010.

[48] M. Laibowitz and J. A. Paradiso, "The UbER-Badge, a versatile platform at the juncture between wearable and social computing," *Advances in Pervasive Computing*, 2004.

[49] W. Brunette, J. Lester, and A. Rea, "Some sensor network elements for ubiquitous computing," *4th International Symposium on Information Processing in Sensor Networks*, 2005.

[50] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *CVPR 2000*, 2000, vol. 2, pp. 142–149.

[51] G. Dublon et al., "DoppelLab: Tools for Exploring and Harnessing Multimodal Sensor Network Data," in *Proceedings of the international IEEE Sensors Conference*, 2011, pp. 1–4.

[52] E. R. Tufte, *Envisioning Information*. Graphics Press, 1990, p. 126.

[53] E. R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. Graphics Press, 2001, p. 200.

[54] *Intel Corp. White Paper - Energy Efficiency*, *Intel Corp. White Paper - Energy Efficiency*. [Online]. Available: http://www.intel.com. [Accessed: 02-Aug.-2011].

[55] E. Naone, "Virtual and Real Play Combined," *MIT Technology Review*, 12-Sep.-2007.

[56] "Pac Manhattan," *pacmanhattan.com*, 2004. [Online]. Available: http://pacmanhattan.com/. [Accessed: Aug.-2011].

[57] S. Benford, C. Magerkurth, and P. Ljungstrand, "Bridging the physical and digital in pervasive gaming," *Communications of the ACM*, vol. 48, no. 3, pp. 54–57, Mar. 2005.

[58] M. Fernström and E. Brazil, "Sonic Browsing: an auditory tool for multimedia asset management," *Proceedings of the International Conference on Auditory Display (ICAD '01)*, 2001.

[59] *SQL Alchemy: The Python SQL Toolkit and Object Relational Mapper*, *SQL Alchemy: The Python SQL Toolkit and Object Relational Mapper*. [Online]. Available: http://www.sqlalchemy.org. [Accessed: 12-Aug.-2011].

[60] Ubisense, *Ubisense Real-time Location System*. Cambridge, UK: http://www.ubisense.net, 2011.

[61] R. Bargar, I. Choi, S. Das, and C. Goudeseune, "Model based interactive sound for an immersive virtual environment," in *Proceedings of the International Computer Music Conference (ICMC '94)*, 1994.

[62] A. Hunt and T. Hermann, "The importance of interaction in sonification," *10th Meeting of the International Conference on Auditory Display (ICAD '04)*, Jul. 2004.