Augmenting the Capture and Understanding of Everyday Experiences

Principal Investigator
Gregory D. Abowd

Other Investigators
Christopher Atkeson
Irfan Essa
Blair MacIntyre
Elizabeth Mynatt
Colin Potts
Kishore Ramachandran
William Ribarsky
Spencer Rugaber
Thad Starner

Contents

1 Description of Instrumentation

In this proposal, we are requesting infrastructure to provide embedded capture work surfaces for use in a variety of applications described in Section 5. We are proposing to outfit two different locations for this purpose. The first location will be a classroom and the second will be a meeting room. In the classroom, we will equip each of 40 student seats with a pen-based annotation surface. We have already built a classroom for this purpose (referred to in the rest of the proposal as the prototype Classroom 2000 environment) that contains power and network distribution to each of the 40 seats. In the meeting room, we will provide large electronic whiteboard surfaces, both upright and horizontal, that will provide for group-level interaction. We will also provide 5 individual annotation surfaces, similar to those used in the classroom.

The annotation systems consist of a high-resolution video tablet from Mutoh, driven by a PC/104+ embedded computer that will be attached underneath the student/meeting table. Mutoh is the only company that produces a XGA resolution video tablet currently, and we have obtained an educational quantity discount from the one of the vendor's resellers. We will be assembling the PC/104+ units from parts obtained from Ampro Electronics. In addition, each annotation system will be instrumented with two sensors. The first sensor will be a simple vision system, implemented in software by researchers in this proposal and will require only an inexpensive color QuickCam by Connectix. The second sensor will use an ID tag that users will carry with them to authenticate with our software. This tag will be the Java iButton from Dallas Semiconductor.

We will use two types of large electronic whiteboard surfaces. The first will be an upright rear-projected SmartBoard from Smart Technology. The second will be the Sonic Board from GTCO, which will allow for a variety of orientations.

Full details of the equipment can be found in the budget section of this proposal.

2 Budget Information

Included with this section are full budget details for the equipment requests made to NSF. We are requesting a total of $199,846 of equipment from the NSF, and an additional $67,085 for equipment as matching from Georgia Tech. All equipment is requested in the first year of the proposal.

Funds from this proposal will not cover the cost of installation and operation of the equipment. Details on installation and operational costs are given in Section 8. As discussed later in this proposal, this additional infrastructure will support the ongoing research of at least four efforts that already have received funding for personnel. Two of them (NSF Career and NSF ESS grants) have funding that will extend for the lifetime (three years) of this proposal. From this money, Dr. Abowd is able to hire undergraduate assistants to provide daily operational support for the prototype Classroom 2000. In the future, the College of Computing has agreed to provide funds to cover student assistants for operational support of Classroom 2000. In addition, the College of Computing's Computing and Network Services group provide full-time support for all public laboratory equipment and classroom equipment.

3 Overall Project Summary

The College of Computing at the Georgia Institute of Technology will purchase interactive pen-based computing equipment which will be dedicated to support research in computer and information science and engineering in the area of capture and retrieval of everyday experiences. The equipment will be used for several research projects, including in particular:

- Investigating research issues in ubiquitous computing: The capture, integration, and access problem (NSF CAREER grant; Abowd PI), in which lecture experiences are preserved and made available for later review.
- Automated understanding of captured experience (NSF Experimental Software Systems grant, Abowd, Atkeson, Essa and Ramachandran co-PIs, MacIntyre, Mynatt and Starner collaborators), in which computational perception and machine learning techniques are applied to interpret the semantics behind captured activities in different domains.
- Automated capture of software architectural rationale (DARPA EDCS program Georgia Tech MORALE project; Abowd, Potts and Rugaber co-PIs), in which an interactive whiteboard provides a more natural surface for creating and capturing software design discussion about the high-level organization of complex software systems.
- Command Post of the Future (Army Research Lab grant, Abowd and Ribarsky co-PIs), in which integrated large format interactive surfaces allow for the capture and dissemination of strategic information from a command post to troops.

4 Abstracts of Individual Research Projects

In this section, we provide brief descriptions of four specific projects that will directly benefit from the infrastructure provided by this grant.

4.1 Investigating research issues in ubiquitous computing: The capture, integration, and access problem

Participating faculty: Gregory D. Abowd (NSF CAREER grant) In this

project, we focus on automating the capture of individual and group experiences in order to facilitate access to a richly integrated record of events. We refer to this as the automated capture, integration and access problem for ubiquitous computing. We have developed a prototype environment —Classroom 2000— to examine this problem in an educational context. We have developed tools that automate the capture, integrate and access of the rich interaction of the classroom experience. The additional infrastructure provided by the grant will allow us to investigate more collaborative issues and greater capture, as we will be able to provide capture surfaces to every student in the prototype classroom as well as in other domains (see following projects).

4.2 Automated understanding of captured experience

Participating faculty: Gregory D. Abowd, Christopher Atkeson, Irfan Essa, Kishore Ramachandran (NSF ESS grant); Thad Starner, Blair MacIntyre and Elizabeth Mynatt

The objective of this project is to reduce substantially the human input necessary for creating and accessing large collections of multimedia, particularly multimedia created by capturing what is happening in an environment. The existing software system which will be used as the starting point for this investigation is Classroom 2000, a system designed to capture what happens in classrooms, that will be extended to address other environments, such as meeting rooms, and offices and programmer work spaces. This project explores how the output of captured experiences can be automatically structured, segmented, indexed, and linked.

4.3 Automated capture of software architectural rationale

Participating faculty: Gregory D. Abowd, Colin Potts and Spencer Rugaber (DARPA EDCS project on Mission ORiented Architectural Legacy Evolution — MORALE)

The Software Architecture Analysis Method (SAAM) was developed at the Software Engineering Institute in the mid-90's and has been effective in helping industrial projects to uncover a shared understanding of the high-level organization of large software systems as well as to reveal how that structure is impacted by suggested changes to system requirements. In the MORALE project at Georgia Tech, we are further investigating the potential for a process such as SAAM to capture the architectural rationale of an evolving software project. One of the features of SAAM is that it is a people-oriented process. Yet this creates difficulties in keeping track of the very rich discussions on system organization and change. With the infrastructure provided in this grant, we will investigate how an augmented meeting room environment can support the conventional SAAM process through automated capture, integration and access to the architectural discussions and artifacts produced by a SAAM session.

4.4 Command Post of the Future

Participating faculty: Gregory D. Abowd and William Ribarsky (separately funded by the Army Research Laboratory)

We are working on integrating previous work in capture environments and high-fidelity visualization of geographic information systems. Our goal is to produce a military tactical meeting room capable of annotating 3-dimensional geographic displays and recording tactical discussions associated with the annotated displays for later dissemination. We are aiming to produce a sophisticated environment that allows for natural interaction modes of voice and pen.

5 Detailed Explanation of Proposed Research

The purpose of this proposal is to create a unique testbed for the development and exploration of large-scale applications of ubiquitous computing. We have been building prototypes of ubiquitous computing applications for over three years and have seen a number of common themes emerge across various applications. One of these concerns the development of automated tools to capture live experiences for the purpose of providing access to them later on. One of the potential features of a ubiquitous computing environment is that it could be used to record our everyday experiences and make that record available for later use. Indeed, we spend much time listening to and recording, more or less accurately, the events that surround us, only to have that one important piece of information elude us when we most need it. We can view many of the interactive experiences of our lives as generators of rich multimedia content. A general challenge in ubiquitous computing is to provide automated tools to support the capture, integration and access of this multimedia record. The purpose of this automated support is to have computers do what they do best, record an event, in order to free humans to do what they do best, attend to, synthesize, and understand what is happening around them, all with full confidence that the specific details will be available for later perusal.

The existence of the experimental facility for investigating research issues with the capture and access of everyday experiences will play a major role in developing a new approach to computing. We expect that ubiquitous will play a dominant role in defining how we think about computing in the near future. The new approach represents a shift from thinking about computers as extensions of the desktop to extensions of the environment we live in. Experimental results and associated demonstrations will play a critical role in this paradigm shift. Small-scale implementations, typically limited to a single chair or room, are useful but do not support exploration of large-scale applications.

5.1 Background

At Georgia Tech, the Classroom 2000 project is mainly concerned with capture, integration and access in support of lecture-based education [8, 5, 11, 6]. The many streams of activity in a typical lecture —what is spoken, what is seen, what is written down on a whiteboard and what is shown on public displays— are combined to provide a rich interactive experience that is becoming increasingly more difficult to capture using traditional pen and paper notes. We have also experimented with capture as a historical component for a mobile tour guide, Cyberguide [7]. The Cyberguide prototypes created summaries of when and where a visitor traveled on campus and preserved images taken with a camera or comments made by the visitor that were attached to this temporally- and spatially-indexed travel diary.

Other research teams have used this same notion of capture, integration, and access to facilitate collaborative or personal experiences. Work at Xerox PARC focused on capturing technical meetings to support summarization by a single scribe who was often not well-versed in the subject of the meetings [27, 28]. More work at PARC (the Marquee system [37], together with work at Hewlett-Packard (the Filochat system [38], Apple [13], and MIT's Media Lab [35] demonstrates the utility of personal note-taking with automatic audio enhancement for later review. More recently, the Informedia project at Carnegie-Mellon, initially set up as a digital library project to investigate sophisticated search capabilities on large bodies of video and audio, has been pushing toward the ability to capture real-life experiences and search through them [36].

5.2 Investigating research issues in ubiquitous computing: The capture, integration, and access problem

5.2.1 Explanation of research

Automated capture of live experiences is the subject of the NSF Career Award to Gregory Abowd, (IRI-9703384, Total award: $438,376 (including matching funds), 7/1/1997 – 6/31/2001). We are focused on developing techniques to capture the different multimedia streams that define a rich interactive experience. Once captured, these streams can be related, or integrated, so that information from one stream can cross-index into another stream. Later on, the relationship between streams can be revisited through universally accessible interfaces.

As with most other ubiquitous computing research, our initial forays involved restricted applications. We have limited our research to investigations in the classroom. Specifically, we have built a prototype classroom environment — Classroom 2000— which provides large electronic surfaces for the display of a wide variety of lecture material [8, 5, 11, 6].[1]

The university classroom is a restricted and useful environment to experiment with capture, integration, and access. The Classroom 2000 project has as its goal to help students and teachers function more effectively in the classroom by augmenting a single room with the capabilities to automatically take notes on behalf of the student. The ubiquitous computing interface of the classroom includes large-scale electronic whiteboards, microphones, cameras, networking, computation and mass storage. Its purpose is to allow a lecture to proceed as normal, with the end result being an automatically generated set of notes that allows the student to relive the whole lecture or more efficiently browse and navigate to points of interest. By viewing the classroom experience as a multimedia authoring session, we simplify content generation for educational software.

Over the past two years, Classroom 2000 has been used as a regular component of 40 different undergraduate and graduate courses at Georgia Tech's College of Computing. Versions of the system have been installed and used in other departments at Georgia Tech (Mathematics, Electrical and Computer Engineering), at other universities in Georgia (Georgia State University, Kennesaw State University and Dalton College), and elsewhere (University of California, Berkeley). Reactions from students and teachers have been very positive.

[1]Further information on the Classroom 2000 system can be found at `http://www.cc.gatech.edu/fce/c2000`

Building off of the success of Classroom 2000, we wish to provide another level of capture. The current prototype software, captures all of the activities of the lecture —the prepared slides they may show, the notes they may write on the board, the Web pages they may visit during class. It does not however, provide a way for a student to electronically annotate the lecture and, therefore, make the experience even more personally meaningful. In fact, students in Classroom 2000 still take a significant amount of notes on paper, albeit notes of a different form than they do in other classes. Interviews and questionnaires of students show a strong desire by students to take electronic notes.

Many electronic classrooms, have experimented with putting computers in the hands of every student in the classroom, but these machines were conventional desktop workstations. We do not propose to do that here; rather, we are interested in providing interfaces that are nearly as flexible and natural as paper, with the added advantage of the electronic domain. We want to build an electronic notebook that follows students from classroom to classroom, is available for use in more informal surroundings (such as student common areas), and understands the situations in which it is being used.

Such a student note-taking interface presents many new challenges to the capture problem. There are significantly more streams being captured, meaning that a full-blown system must address the scalability of the multimedia server receiving and timestamping events. Students will be interested in more than one stream at their desk. Not only will they want to have a way to take their own notes (a student stream), they will want to see and additionally annotate the lecturer's notes. Interactive discussions can be facilitated by the exchange of information between teachers and students, so there must be a way to exchange information between all surfaces in the classroom in a way that is non-intrusive. Finally, student note-taking does not end when a lecture is complete. A set of notes can be subject to review and modification at any time after the lecture. Students must be able to easily access their notes after class and be able to modify them. This review is most likely not going to be done with the same physical interface as in the classroom, so the system must provide for multiple modes of input.

Providing students with the ability to take electronic notes will also impacts the kind of notes that they take. In the current Classroom 2000 system, more than 50% of the students report that they change their note-taking practices as the result of having the lecturer's notes automatically available to them. Most of those that report a change state that they take fewer notes and that the character of their notes is more summary than detailed. A small, but significant, portion of students specifically request that the system be improved to allow them to take notes as well. Detailed evaluation of the electronic note-taking practices requires that we provide a full-scale working prototype that students can depend on throughout a course.

We have already created a prototype student note-taking environment — StuPAD— that runs on the equipment requested, and it is shown in Figure 1. In this screenshot, there are two active streams in the system, the lecturer and the student. Students can navigate each stream by interacting with the right

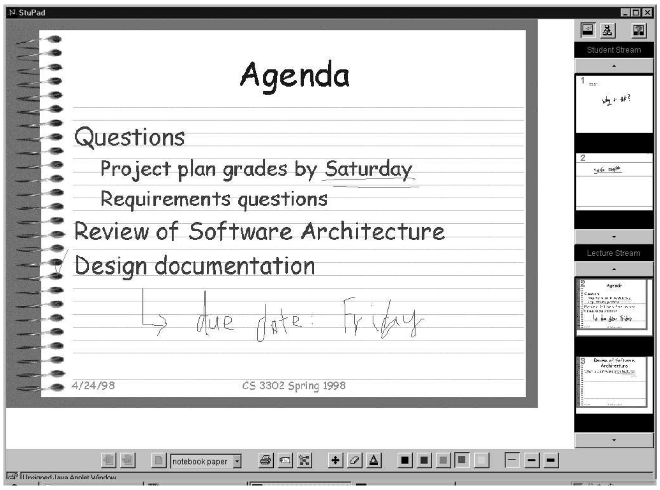

Figure 1: Screenshot of StuPAD, an initial prototype to facilitate electronic note-taking by students in Classroom 2000. Along the right side of the screen, the student can navigate the lecturer's current presentation (lower right) or the student's separate stream of notes (upper right). A page or slide in either stream can be summoned to the main writing panel (left) for further student annotation. Above, the student has selected one of the lecturer's slides and has written a due date upon it.

panel. A particular page or slide in a stream can be put into the student's workspace for additional annotation. In Figure 1, the student has accessed one of the lecturer's slides and has added further notes to it. The infrastructure in this grant will allow us to further experiment with StuPAD in a real classroom environment and gain insight into the proper way to facilitate electronic note-taking.

5.2.2 Equipment rationale

A large portion of the equipment is earmarked for the specific purpose of augmenting the original prototype Classroom 2000 environment. This classroom currently seats 40 students and each seat has power and network accessibility. The equipment from this grant will allow us to build student note-taking surfaces using high-resolution video tablets from Mutoh. We will connect each video tablet an embedded PC/104+ computer attached under the table. Each student surface will be equipped with additional sensors to help us better support the many different students who will use the classroom.

5.3 Automated understanding of captured experience

5.3.1 Explanation of research

In this NSF Experimental Software Systems project, we are exploring large collections of multimedia that are built by automatically capturing what is happening in an environment. Examples of such systems are electronic home movies and automatic note taking in meetings. Major impediments to creating and using such systems is the human input required to index and link each new entry, and the effort required to navigate through the flood of information. The scientific question we seek to answer is what techniques, if any, will allow us to minimize the human input required to index, link, and navigate automatically acquired multimedia collections.

The existing software system we will use as the starting point for this investigation is Classroom 2000, discussed above. Variants of this system have been prototyped for structured meetings as well as informal gatherings. This research explores how we can automatically structure, segment, index, and link the output of a capture system. Now that we have a working system with many users, we would like to move to the next phase of research, which focuses on experimental questions about how we can increase the utility of what is provided by this kind of multimedia capture system. Examples of such questions include:

- What is the pattern of usage by actual users? Which parts of the captured information are used, and for what purpose?
- How can we automatically index, link, and summarize new material?
- What types of data representations, software filters, and search mechanisms best support access to captured experiences?
- How can we use information in streams such as audio and video to assist in providing a more content-based access strategy? Can we automatically recover the focus of attention of the group, and the emphasis and structure of the speaker?

The answers to these questions will guide a major redesign of the system, including the design of novel user interfaces. The current system supports browsing the text, images, and audio and video streams associated with a captured experience. Access to any one stream, such as individual pen strokes in handwritten text, indexes the other streams at the same time. The user can also perform simple keyword searches over phoneme sequences derived from the audio. Example tasks we would like to perform with an enhanced system include:

1. Answering more general queries, including trying to identify the concept the user is searching for rather than the word.
2. Answering cross-media queries, so a query in one media (text, handwriting, speech, ...) can find matches in all media.

3. Automatically segmenting an experience, to match the emphasis and structure generated by the speaker.

4. Automatically linking captured material to other material such as textbooks, briefing materials, or web pages.

5. Automatically creating navigation bars, maps, and other graphical material to assist browsing.

6. Automatically finding keywords or making summaries or abstracts, to aid human navigation of large amounts of material.

This work raises major research questions to be answered in the areas of machine learning and computational perception, but at the heart of this proposal is an experimental ethic.

New abilities we plan to implement in Classroom 2000 include

- answering more general queries, including trying to identify the concept the user is searching for rather than the word,
- automatically segmenting material, and
- automatically linking captured material to other material such as textbooks, briefing materials, or web pages.

All of these abilities require methods to identify what concepts the user is interested in or are referenced in a segment of text. We plan to use some "off the shelf" methods and some methods we have developed for matching concepts to text to provide *concept-based* access. Information retrieval by semantic category will require capabilities that go far beyond those available from keyword retrieval systems such as AltaVista. The retrieval of natural language documents requires the construction of intermediate representations which are more closely related to document content than a simple list of keywords. We intend to learn these representations directly from large corpora using statistical techniques. The discussion of this section will assume that recognized text is available. Later we will discuss how to integrate information retrieval and speech and handwriting recognition processes more effectively.

The problem of representing text to enable machine manipulation or understanding has been studied in several fields. Information Retrieval considers the problem of clustering similar documents or finding documents given a set of query terms [32]. Natural Language Processing includes several levels of analysis of text, from parsing or tagging sentences to summarizing or extracting relevant sections of text [12]. Naturally, we would like to apply techniques from natural language understanding to this problem. Unfortunately, the sheer size of the data to be represented makes this difficult. We wish to process millions of documents, each of which may contain any of up to 100,000 different words. Current natural language systems require significant computation and would have difficulty processing this much data. Agent-based approaches have been used in web browsing [23, 30], web shopping [14], newsgroup reading [18, 34], interface assistance [39, 33], information retrieval [21], and email management [26, 22, 29].

We will use a statistical approach to language understanding that is based on matching vocabularies. A vocabulary for a segment of text is a measure of the relative frequencies of different words in the text. There are many variants of how exactly to represent a vocabulary. For now let us assume a vocabulary is a sorted list of the words in the text, with each word paired with its relative frequency. A simple measure of the similarity of two such vectors is the inner product of the relative frequencies: $S(\mathbf{x}, \mathbf{y}) = \sum_{w \in W} \mathbf{x}_w \mathbf{y}_w$ where W is the set of words in either $\mathbf{x}$ or $\mathbf{y}$ and $\mathbf{x}_w$ and $\mathbf{y}_w$ are the number of times w appears in the corresponding text divided by the total number of words in that text. $\mathbf{x}_w$ and $\mathbf{y}_w$ are zero if the word w is not in the corresponding text. A large $S(\mathbf{x}, \mathbf{y})$ means the vocabularies are similar, and $S(\mathbf{x}, \mathbf{y})$ close to 0 means the vocabularies are unrelated. In practice, suffixes are removed and counts are weighted by their natural frequency in forming the word vectors.

The key insight in many statistical approaches to language is that "concepts" can be represented by vocabularies. Using this definition of concept enables concepts to emerge from the data, and allows us to avoid hand created concepts or semantic features. Pieces of text which have similar vocabularies are probably about related topics, and are probably relevant to each other. This information can be used to automatically make links between two pieces of text. The vocabulary that surrounds a word can be used to disambiguate its meaning if the word is used in a keyword search. Changes in the vocabulary used can be used to segment a piece of text. If a sliding window is used to select a text segment (100 consecutive words, for example) the corresponding vocabulary vectors can be computed. When the differences between the vocabulary vectors of adjacent windows are large then there is probably a change in topic.

Much of our content will be from actual speech rather than text. Using speech as input will have a large effect on all our algorithms, which must deal with incomplete and ungrammatical utterances and incorrect outputs from the speech recognition module. We are using and plan to expand our use of several commercial speech recognition products, including the HTK toolkit from Entropic [16] and Naturally Speaking from Dragon Systems [15]. Neither system is perfect, but we have been able to get usable results in pilot studies. One experimental question we would like to address is how good the speech recognition has to be to support functions like search. We suspect that the demands of searching for relevant sections of an audio stream are much less than the requirements for generating a readable transcript. In the Classroom 2000 system the output of the speech recognition module is a sequence of time-stamped tokens.

One of our research goals is to use a wide range of measurements to determine the focus of individual and group attention, the speaker's emphasis, and natural subdivisions of captured experiences. This section focuses on using sensory cues captured from the environment to extract useful semantic information about what is happening in the scene. This form of semantic information can aid in automatic understanding and annotation of the captured data streams. More specifically, our interest is in observing, measuring, and tracking the activities of the people in our environments from the captured video and audio streams, and from that extract a greater semantic content to assist in automated annotation, leading to content-based access. Towards this end we are interested in building systems that can a) find and identify people in the environment b) track gaze and body orientation, c) recognize facial expressions, gestures and body language, d) process speech at both the verbal and nonverbal levels to recognize the words and capture the affect, intensity, and style of the speakers.

5.3.2 Equipment rationale

In the NSF ESS grant, we are able to purchase some video equipment infrastructure to better perceive the environment. However, we are not yet able to add enough sensing of the classroom to know who is where. That is why in this proposal we have explicitly asked for equipment to instrument each of the student seats with simple sensors —a color camera to help determine when a student is present and where they are focusing their attention. We have also requested a large number of inexpensive Java iButtons to enable students to quickly and reliably identify themselves to a note-taking station without need of a keyboard.

In addition, we will explore applications outside the classroom in this project and so we have requested infrastructure to equip a special-purpose meeting room that will be able to capture activity on large-scale upright and horizontal surface. For upright electronic whiteboards, we use the rear-projected SmartBoard from Smart Technology that we have successfully used in the lab and in Classroom 2000. For horizontal surfaces, such as might be used by an architect, designer or programmer, we will use the SonicBoard by GTCO. For collaborative meetings, we will provide capture stations similar to those used in the classroom.

5.4 Automated capture of software architectural rationale

5.4.1 Explanation of research

Software evolution is a costly and time consuming software development activity. Effective system evolution requires understanding both the way that an existing system accomplishes its tasks and the mission-oriented rationale for any changes that feed its evolution. The MORALE project at Georgia Tech addresses the problem of designing and evolving complex software systems. The MORALE acronym summarizes its goals:

- Mission ORiented: We want the legacy system enhancement process to be driven by the mission to be accomplished rather than by purely technical criteria.
- Architectural: The most time consuming and costly alterations to software are those that distort architecture, by which we mean its structure and behavior. We want to provide a mechanism for predicting the impact of architectural changes so that the risks of making those changes can be ascertained early in the evolution process.

- Legacy Evolution: We are concerned with the evolution of legacy systems. We want to provide a cost effective way of analyzing existing software, and once analyzed, extracting those parts of it which can be used in the new version.

MORALE addresses these goals by integrating several different technologies: reverse engineering to extract and visualize architectural information[17]; requirements analysis to structure and understand an organization's changing mission-oriented goals[31]; and finally software architecture impact analysis to determine how requirements changes affect the high-level organization of a software system.

The work proposed here focuses on a particular process in support of the latter technique. Specifically, we make use of the Software Architecture Analysis Method (SAAM) developed at the Software Engineering Institute in the mid-90's ([10, 19, 20]. SAAM is a scenario-based method that can help extract how changing requirements will impact an already existing software system. In this context, we can understand the entire SAAM process as a design rationale technique that is centered on software architecture. We therefore refer to SAAM as an architectural rationale technique.

The SAAM process is a very people-oriented technique, and as such, has advantages and disadvantages. The main advantage is that it is a practical and useful technique for extracting a common understanding of the high-level software organization of a complex system. This is particularly important in that software architecture is rarely properly documented for the benefit of future designers. The other important advantage of SAAM is that it makes explicit how intended changes will impact the existing system, oftentimes helping an organization avoid undesirable development efforts. One disadvantage of SAAM is that it assumes that designers are able to describe the existing organization of the software, even when the original architects are not available. Secondly, SAAM requires additional time and effort to produce a coherent written summarization of the results of the analysis.

One goal of the MORALE project is to address the first disadvantage of SAAM by integrating with visualization techniques in reverse engineering to uncover the architecture of an existing system. In this research, we are concerned with how ubiquitous computing technology might address the second disadvantage.

A typical SAAM session is a live event involving discussions by 3-10 designers, managers and facilitators that is centered around drawings of the architecture on a public display. By converting the public display to an electronic whiteboard surface and recording the discussion with digital streaming media technology, we can automatically capture the SAAM sessions and then provide the ability to salvage summary information afterwards. By allowing the ubiquitous computing infrastructure to do what it does best, record and capture activity, we can free the humans to do what they do best, synthesize and understand activity in a technical discussion.

We have built an initial prototpye environment to investigate how automated capture tools can support an architectural rationale technique such as SAAM. The prototype, called SAAMPad, is a meeting room environment centered around a large electronic whiteboard. This prototype environment has been shaped to support naturally defining a software architecture and capturing scenario-based impact analyses during a SAAM session.

As a capture problem, this application presents new problems. The stream of activity is not linear as one would get from a lecture or a meeting. The SAAM analysis focuses on a drawing of a high-level organization of a system, typically in the form of a box-and-line diagram. This diagram is hierarchical, so the discussion consists of creating the architecture in levels and continually popping in and out of levels of the hierarchy. A discussion of evolution of an existing architecture consists of revisiting the same initial architecture of the system and discussing how it must be changed to meet new needs. Several such discussions quickly result in a very cluttered set of changes over time, so there must be ways to organize and visualize the collection of rationale discussions in such a way that it continues to inform future design activities.

5.4.2 Equipment rationale

The special purpose meeting room will be used to further develop the prototype SAAM capture environment. In particular, the large-screen electronic whiteboards will be of use immediately. The horizontal GTCO interactive surface will allow us to experiment with tracking discussions on real paper artifacts as well.

5.5 Command Post of the Future (CPOF)

5.5.1 Explanation of research

In this research, sponsored in part by the Army Research Laboratory at Georgia Tech, we are addressing a military environment, the command post, as the subject of capture and intelligent retrieval. Elements of this research include:

- Development of multimodal interaction, interpretation, and capture techniques. Multimodal interaction and capture techniques will go beyond text-based methods to multimodal user inputs and a range of media. User access will be wireless, unconstrained, and ubiquitous with recognition of gestures, expressions, and roles for multiple persons simultaneously.

- A location (4D) for everything with accurate place, time, and semantical information. These structures will be efficiently accessed and adapted and will support real-time display with accurate insertion of objects across multiple visual spaces.

- 4D smart rooms with histories and predictions. The system will keep track of alternate histories with branches for different predictions. The user will be able to travel between these different histories and see the results visually.

- Visualizations based on the knowledge gleaned from captured resources with both accurate, high resolution display and abstracted display of key elements, their behaviors and their histories. The system will permit communication and interactive switching between the two display methods.

- Highlight reels based on 2D, 3D, and 4D "cartoons" derived using visual abstractions, predictions, and needs assessments. These will be fast, automatic, and on the spot constructions that will distill several hours of activity into highlights that will tell the story of that activity quickly. They will go well beyond just the capture of video frames, audio, and graphical images. (See the Visualization part of this section for more details.)

We are exploring the hypothesis that machine capture and understanding of command post activities can support more effective planning and decision making. We will pursue an integrated approach to the capture and understanding of human activities, including speech, handwriting, drawing, and gesture. What drives the integration is the need to make all of these captured human activities electronically accessible in a unified way. Our research goal is to develop algorithms to provide concept-based access to captured material, and to also guide access using information about the decision and planning process and the original focus of attention.

Our approach to information management hinges on developing:

- Methods to unobtrusively capture integration, planning, and decision making activities.

- Algorithms to provide concept-based access to online knowledge sources. Standard speech, handwriting, and gesture recognition algorithms will be adapted to work well in this context, and information retrieval algorithms will be adapted to make use of the current level of performance of these recognition algorithms.

- Presentation methods and navigation algorithms to handle the flood of captured data and using the knowledge derived from captured resources to select and refine the visualizations.

- Tools and interfaces to enable captured resources to assist collaboration across space and time.

Answering queries, automatically structuring captured events, automatically generating presentations, and automatically summarizing presentations and events require methods to identify concepts of user interest or those referenced in captured material.

We are using a wide a wide range of measurements to determine the focus of individual and group attention, the speaker's emphasis, and natural subdivisions of captured experiences. Sensory cues captured from the environment will be used to extract useful semantic information about what is happening in the scene. This form of semantic information can aid in automatic understanding and annotation of captured datastreams. More specifically, we will observe, measure, and track the activities of people in our environments from captured video and audio streams, and, from that, extract semantic content to assist in automated annotation, leading to concept-based access, as described above. Towards this end, we will build systems that can:

- find and identify people in the environment;
- track gaze and body orientation;
- recognize facial expressions, gestures and body language; and
- process speech at both the verbal and nonverbal levels to recognize the words and capture the affect, emphasis, and style of the speakers.

Knowledge of the group focus of attention is usually lost in captured experiences. The Classroom 2000 system, discussed earlier, captures several large projected displays simultaneously (currently up to 4). The display that was last modified is assumed to be the current focus of attention, and is more prominently featured when the captured experience is replayed. However, this simple assumption is often wrong, as the speaker may indicate a change of focus verbally, by gesture or change of body configuration, or just by moving his or her eyes. If such changes are captured, it would be possible to have a much more informative playback. Emphasis and segmentation are also typically lost during capture. Speakers use variations in volume, pauses, stereotypical hand gestures, and stereotypical utterances to indicate an especially important phrase or part, or a change of topic. This information is useful in automatically making summaries or indicating sections of a presentation.

It is therefore essential to focus on capturing, processing, and automatically annotating these communication modes (speech, facial expressions, and hand gestures) from captured experiences. Though they are essential to interpersonal interaction, only recently have there been attempts to study them within the context of human-machine interactions and automatic annotations [1, 2, 3, 4]. We will explore new methods for machine perception, understanding, and interpretation of human activity in a scene or environment. As a testbed, we are using instrumented spaces, such as a special-purpose meeting room treated as a military command post with cameras, microphones, and recording facilities for off-line analysis of audio and video signals. The extracted semantic information will enhance and direct the decision-making and data collection processes. The semantics and extracted multimodal clips will also be essential in developing of the narratives and highlight reels.

To meet our goal of moving transparently and quickly among the many display, analysis, and input modes of the CPOF environment, we must have data structures and communications interfaces that speak a common language.

It is natural to discuss these things in the visualization section, because the display must fuse these data from multiple sources and because it will often use high-level representations and abstractions that in themselves are part of the common language. In addition the visualizations provide a path for real-time system responsiveness that must be enforced throughout the CPOF system.

All objects in the visualizations and in all other CPOF system components will have a particular time, place, and (optionally) author. Enforcement of this important concept will mean accurate placement of objects (including vehicles, buildings, images, annotations, icons, etc.) among any of the displays and within the spatial or temporal organization of any dataset. For example, annotations created on a 2D wall display will appear at the appropriate place in the 4D terrain navigation system. The CPOF and its occupants will have well-defined locations with respect to the rest of the world, and the CPOF staff will be able to reach into other coordinate systems to make selections or insertions. All this will require multiple, perhaps moving, coordinate systems with transformations mapping from one to the other. Time will occupy a special, parametric status in the coordinate structure. Everything will have a time stamp to determine time order. In particular, time ordering will be used to create stories describing command post actions, annotations, and decisions. All place and time updates throughout the CPOF system will be organized to meet time budgets, ensuring interactivity.

The visualization system will be built on the successful and widely used VGIS system, which two of the co-PIs developed in conjunction with ARL.[24, 25].[2] The very efficient hierarchical structure, fast paging, and scalable data organization [9] will be extended to handle all spatially organized data within the CPOF system. The data structures will be enhanced to handle data of any size and fast data insertion techniques will be built to handle spatial data of any type. However, the CPOF system will have significant capabilities well beyond those of VGIS. For one thing, it will handle all data, not just geospatial data, in its hierarchical structures. New fast traversal techniques coupled to these hierarchies will permit priority preparation and paging of these data so that they arrive when needed. Multiple image blending will be possible so that, for example, maps and geographic images can be overlaid accurately on terrain with the transparency of each image layer controlled interactively. The user will be able to either zoom into or jump to high resolution areas (even in very large terrain datasets) without loss of navigational speed.

5.5.2 Equipment rationale

The infrastructure proposed here for the meeting room environment will be used to complement laboratory equipment already used by Abowd and Ribarsky to simulate a command post environment. Specifically, the upright SmartBoard electronic whiteboards and the video tablet capture stations will be used to support the roles of individuals in the command post. Our existing VGIS software

[2] See `http://www.cc.gatech.edu/gvu/datavis/research/vgis.html` for details.

will be run on high-end workstations, not listed in this proposal and displayed on the various capture surfaces.

6 Vitae for Investigators

7 Current and Pending Support for Investigators

8 Maintenance and Operation Plans

This proposal will be providing equipment that will run in up to three different locations within the College of Computing. The classroom equipment (video tablet capture stations) will be hosted in Room 102 of the College of Computing. This classroom receives maintenance and operational service from three sources. Since it was originally built with funding from the Office of Information Technology on campus, they provide routine network and audio-visual maintenance support. Operational needs for this classroom room are significant, since it is the desire of the researchers to provide a seamless operation to students and lecturers who use the services of Classroom 2000. The operational budget for the classroom is estimated at $20,000 a year. Funding for this operational work comes from Abowd's NSF CAREER grant, grants from industry (Mitsubish Electric Research Lab, the Mobility Foundation and the Corporation for National Research Initiatives), and from the College of Computing and is used to pay undergraduates to support the technology in the room. This funding will include the cost of assembling and installing the video capture stations.

The meeting room facilities will be hosted both in the main College of Computing building, in an as yet undetermined conference room, as well as in Abowd's Future Computing Lab in the Centennial Research Building on campus. Dr. Abowd will pay graduate students to help install and maintain the equipment for these two locations and funding for those students comes from the NSF CAREER grant and NSF ESS grant. In addition, Abowd, Atkeson and Essa have applied for a Major Research Instrumentation grant from NSF that will help to pay for full-time staff to maintain and operate the equipment in the special-purpose rooms, estimated at $50,000 a year.

In addition, the Computing and Network Services group in the College of Computing dedicate a full time equivalent staff member to maintain equipment in classrooms and laboratories. This cost is estimated at $50,000 a year.

9 Facilities Description

The College of Computing has a variety of general-purpose and specialized computing capabilities (some of which are jointly administered with other campus units) all of which are available for use by the PIs and their students and post-docs. In particular, the following specialized facilities are relevant for the proposed research:

- In January 1997, a prototype classroom was created for both teaching and research in the College of Computing. This classroom has specialized presentation and capture facilities suited for the Classroom 2000 project. Specifically, there is a 67-inch diagonal rear-screen projected Liveboard and two additional ceiling-mounted LCD projectors from Proxima Corporation. There are high-quality ceiling microphones and a wireless lapel microphone for recording of audio in the room. Ceiling-mounted cameras provide several video feeds. All audio/video feeds are coordinated by a mixer that records on both analog and digital media. The periphery of the room contains 15 teacher/student workstations (Macintosh/PC/SGI/Solaris). Each of 40 students seats is provided with dedicated power and Ethernet through a raised floor. In addition, the College of Computing provides two other Liveboards for research and education purposes.

- The GVU Center and College of Computing opened an auxiliary lab, th Future Computing Lab, in September 1997 that contains a 3000 square foot facility with 50 workstations (Macintosh, PC, SGI, Solaris) and specialized equipment for robotics, ubiquitous computing and virtual environments. This specialized equipment consists of a number of LCD projectors and electronic whiteboards, that will be directly used in this project.

- a cluster of 16 Sun UltraSPARC processors and 5 UltraSPARC dual-processors utilizing Myrinet (8 systems), Dolphin (4 systems), and ATM interconnects (all systems),

- a cluster of 16 Pentium quad-processors interconnected by Myrinet and 100 Megabit Ethernet,

- an SGI Origin 2000 quad-processor, and an SGI Power Challenge with 12 processors, (these two are expected to be upgraded to a 16 processor SGI Origin 2000),

- a cluster of 16 Silicon Graphics R4400 processors utilizing an ATM interconnect,

- SGI, Sun, and Intel video servers, and high-end SGI, Sun, and Intel multimedia workstations.

In addition to the above, the following campus-wide high-performance machines are available for experimentation.

- an IBM SP-2
- a 20-processor SGI Origin 2000 Multiprocessor

All of these facilities are linked by a dedicated high-performance network utilizing ATM.
The college also has the following computing capabilities:

- 40 Sun and SGI systems used as file and compute servers, 10 of which are quad-processor machines; more than 350 workstation class machines from Sun, Silicon Graphics, Intel, Apple, Hewlett-Packard, and IBM; and over 150 X-terminals,
- the Graphics, Visualization, and Usability (GVU) Center houses a variety of graphics and multimedia equipment, including high-performance systems from Silicon Graphics, Sun, Hewlett-Packard, Digital, Intel, and Apple, as well as extensive video and audio facilities for recording and editing,
- a Scientific Visualization Laboratory with additional equipment from Digital and Silicon Graphics jointly operated by the GVU Center and the Institute's Office of Information Technology (OIT),
- the Broadband Telecommunications Center (BTC), have several labs equipped with leading edge computing, communications, and test equipment. These include the Hybrid Fiber/Coax (HFC), Asynchronous Transfer Mode (ATM), Home Information Infrastructure (HII), Protocols, Wireless Technologies, and Video Sources Labs,
- other specialized laboratories supporting research in databases, robotics, open systems, software engineering, computer networking and communications systems, future computing environments, educational technology, and artificial intelligence.

All of the College's facilities are linked via local area networks that can provide a choice of switched 10 Mbps Ethernet, switched or shared 100 Mbps Ethernet, 155 Mbps ATM, or 622 Mbps ATM connections to most locations, including offices, labs, and classrooms. The College's network employs a high-performance OC12C ATM (622 Mbps) and FDDI (100 Mbps) backbone with connectivity to the campus ATM network and to the global Internet beyond via redundant OC3C ATM (155 Mbps) links. The primary campus Internet connection is provided by a direct FDDI (100 Mbps) link to the service provider's Atlanta switching center. An OC3C (155 Mbps) ATM connection to the NSF vBNS (very high performance Backbone Network Service) research network has recently been brought into service. Georgia Tech is also leading efforts to establish a southern regional gigabit network as part of Internet2.

Additional computing facilities are provided to the Georgia Tech campus by the Office of Information Technology, including eight public-access clusters of Apple, IBM, Dell, and Sun workstations, a Sun SPARCcenter 2000 with 12 superscalar RISC processors, various mainframes, and public-access interactive terminals.

Previous NSF equipment

The cluster of 16 UltraSPARC single and dual processor machines and the cluster of quad-Pentium machines was obtained from a NSF Infrastructure grant received in 1996.

Abowd, Atkeson and Essa have recently applied for a NSF Major Research Instrumentation grant.

References

[1] *FG Workshop: Proceedings of the International Workshop on Automatic Face and Gesture Recognition.* IEEE Computer Society Press, June 1995.

[2] *Proceedings of the International Workshop on Automatic Face and Gesture Recognition.* IEEE Computer Society Press, October 1996.

[3] *NSF/DARPA Workshop on Perception of Action*, 1997. Position papers available from http://www-white.media.mit.edu/~bobick/downloads.

[4] *Workshop on Perceptual User Interfaces*, Banff, Canada, October 1997. In conjunction with UIST'97.

[5] Gregory D. Abowd, Christopher G. Atkeson, Jason Brotherton, Tommy Enqvist, Paul Gulley, and Johan LeMon. Investigating the capture, integration and access problem of ubiquitous computing in an educational setting. In *Proceedings of the 1998 conference on Human Factors in Computing Systems — CHI'98*, pages 440–447, May 1998.

[6] Gregory D. Abowd, Christopher G. Atkeson, Ami Feinstein, Cindy Hmelo, Rob Kooper, Sue Long, Nick Sawhney, and Mikiya Tan. Teaching and learning as multimedia authoring: The classroom 2000 project. In *Proceedings of the ACM Conference on Multimedia — Multimedia'96*, pages 187–198, 1996.

[7] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. Cyberguide: A mobile context-aware tour guide. *ACM Wireless Networks*, 3:421–433, 1997.

[8] Gregory D. Abowd, Jason Brotherton, and Janak Bhalodia. Automated capture, integration, and visualization of multiple media streams. In *Proceedings of the 1998 conference on IEEE Multimedia and Computing Systems*, 1998.

[9] Douglass Davis adn T.Y. Jiang, William Ribarsky, Larry Hodges, and Nick Faust. Intent, perception and out-of-ccore visualization applied to terrain. In *IEEE Visualization'98*, 1998. to appear.

[10] Len Bass, Paul Clements, and Rick Kazman. *Software Architecture in Practice.* Addison-Wesley, 1998.

[11] Jason A. Brotherton and Gregory D. Abowd. Rooms take note: Room takes notes! In *Proceedings of the 1998 Spring AAAI Symposium on Intelligent Environments*, 1998.

[12] Claire Cardie. Emperical methods in information extraction. *AI Magazine*, 1998. to appear.

[13] Leo Degen, Richard Mander, and Gitta Salomon. Working with audio: Integrating personal tape recorders and desktop computers. In *Proceedings of ACM CHI'92 Conference*, pages 413–418, May 1992.

[14] B. Doorenbos, O. Etzioni, and D. Weld. A scalable comparison-shopping agent for the world-wide web. In *Proceedings of the 1997 International Conference on Autonomous Agents*, February 1997.

[15] DragonDictate and Dragon Naturally Speaking, Speech Recognition products by Dragon Systems Inc. `www.dragonsystems.com`.

[16] HTK, ESPS, WAVES+, GrapHvite and other Speech and HMM related products by Entropics Research Laboratories Inc. `www.entropic.com`.

[17] Dean Jerding and Spencer Rugaber. Using visualization for architectural localization and extraction. In *Proceedings of the Fourth Working Conference on Reverse Engineering*, October 1997.

[18] Tomonari Kamba, Krishna Bharat, and Michael Albers. The krakatoa chronicle: An interactive, personalized newspaper on the web. In *Proceedings of the Fourth International World-Wide Web Conference*, Boston, MA, December 1995.

[19] Rick Kazman, Gregory Abowd, Len Bass, and Paul Clements. Scenario-based analysis of software architecture. *IEEE Software*, 13(6):47–56, November 1996.

[20] Rick Kazman, Len Bass, Gregory D. Abowd, and S. Michael Webb. SAAM: A method for analyzing the properties of software architectures. In *The proceedings of the International Conference on Software Engineering — ICSE-16*, pages 81–90. IEEE Computer Society Press, 1994.

[21] Daphne Koller and Yoav Shoham. Information agents: A new challenge for ai. *IEEE Expert*, pages 8–10, June 1996.

[22] Yazdi Lashkari, Max Metral, and Pattie Maes. Collaborative interface agents. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume 1, Seattle, WA, August 1994. AAAI Press.

[23] Henry Lieberman. Letizia: An agent that assists web browsing. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Montreal, August 1995.

[24] Peter Lindstrom, David Koller, William Ribarsky, and Nick Faust. Real-time continuous level of detail rendering of height fields. In *SIGGRAPH'96*, pages 109–118, 1996.

[25] Peter Lindstrom, David Koller, William Ribarsky, Larry Hodges, and Nick Faust. An integrated global GIS and visual simulation system. *Transactions on Visualization and Graphics*, 1998. submitted for review.

[26] P. Maes and R. Kozierok. Learning interface agents. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 459–465, Washington, D.C., July 1993. MIT Press.

[27] Scott Minneman, Steve Harrison, B. Janseen, G. Kurtenbach, Tom Moran, Ian Smith, and B. van Melle. A confederation of tools for capturing and accessing collaborative activity. In *Proceedings of the ACM Conference on Multimedia — Multimedia'95*, November 1995.

[28] Thomas P. Moran, Leysia Palen, Steve Harrison, Patick Chiu, Don Kimber, Scott Minneman, William van Melle, and Polle Zelweger. "I'll Get That Of the Audio": A case study of salvaging multimedia meeting records. In *Proceedings of ACM CHI'97 Conference*, pages 202–209, March 1997.

[29] Terry R. Payne and Peter Edwards. Learn: An investigation of learning issues in a mail agent interface. *Applied Artificial Intelligence*, 11:1–32, 1997.

[30] Michael Pazzani, Daniel Billsus, and Jack Muramatsu. Syskill & webert: Identifying interesting web sites. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Portland, OR, August 1996. American Association for Artificial Intelligence.

[31] Colin Potts, K. Takahashi, and Annie Anton. Inquiry-based requirements analysis. *IEEE Software*, 11(2):21–32, 1994.

[32] Gerard Salton. *Automatic Text Processing.* Addison-Wesley, Reading, MA, 1989.

[33] Jeffrey C. Schlimmer and Leonard A. Hermens. Software agents: Completing patterns and constructing user interfaces. *Journal of Artificial Intelligence Research*, 1:61–89, 1993.

[34] B. Sheth and P. Maes. Evolving agents for personalized information filtering. In *Ninth Conference on Artificial Intelligence for Applications*, pages 345–352, Orlanda, FL, March 1993. IEEE Computer Society Press.

[35] Lisa J. Stifelman. Augmenting real-world objects: A paper-based audio notebook. In *Proceedings of ACM CHI'96 Conference*, pages 199–200, April 1996. Short paper.

[36] Carnegie Mellon University. Informedia: Researching digital video libraries. http://www.informedia.cs.cmu.edu/html/enter.html.

[37] Karon Weber and Alex Poon. Marquee: A tool for real-time video logging. In *Proceedings of ACM CHI'94 Conference*, pages 58–64, April 1994.

[38] Steve Whittaker, Patrick Hyland, and Myrtle Wiley. Filochat: Handwritten notes provide access to recorded conversations. In *Proceedings of ACM CHI'94 Conference*, pages 271–277, April 1994.

[39] Andy Wood, Anind Dey, and Gregory D. Abowd. Cyberdesk: Automated integration of desktop and network services. In *Proceedings of the 1997 Conference on Human Factors in Computing Systems (CHI '97)*, 1997.