The Networked Home as a User-Centric Multimedia System

Ankur Mani Hari Sundaram David Birchfield Gang Qian
Arts Media and Engineering Program
Arizona State University.
e-mail: {ankur.mani, hari.sundaram, david.birchfield, gang.qian}@asu.edu

ABSTRACT

This is a position paper that frames a networked home as a situated, user-centric multimedia system. The problem is important for two reasons – (a) the emergence of high speed networked connections alter media consumption and interaction practices and (b) ordinary consumers currently communicate everyday experiences through limited means (e.g. e-mail attachments). We need new mechanisms for networked creation and consumption of media, as well as new interaction paradigms that will allow us to utilize the full potential of the networked, multimedia environment. We envision an augmented user-context adaptive home that enables the user to rest, reflect, interact and communicate everyday experiences through multimedia.

A key insight is that the *practice* of consumption, communication and interaction with media, across different devices and interaction modalities, affect the user context, and in turn is affected by it. The result is a highly personalized *media practice* for each user. We discuss three focal areas of our current research – (a) models for user context, (b) communication of meaning and (c) situated interaction. Modeling user context is challenging, and we present a novel multimodal context framework. In media communication, we examine research issues in media acquisition, media presentation and networked sharing. Situated multimedia frameworks are physically grounded systems, that require new analytical models, interaction paradigms, and additionally require new real-time concerns. Our framework is promising, and we believe will lead to rich collection of multimedia problems that incorporate networked interaction.

Categories and Subject descriptors

H.5.1 [**Multimedia Information Systems**]: *Artificial, augmented, and virtual realities,* H.5.2 [**User Interfaces**]: *Theory and methods, User-centered design.* I.5.2 [**Design Methodology**]: *Classifier design and evaluation, Feature evaluation and selection, Pattern analysis*

General Terms: Algorithms, Design, Human Factors

Keywords: Networked home, context models , communication, situated systems

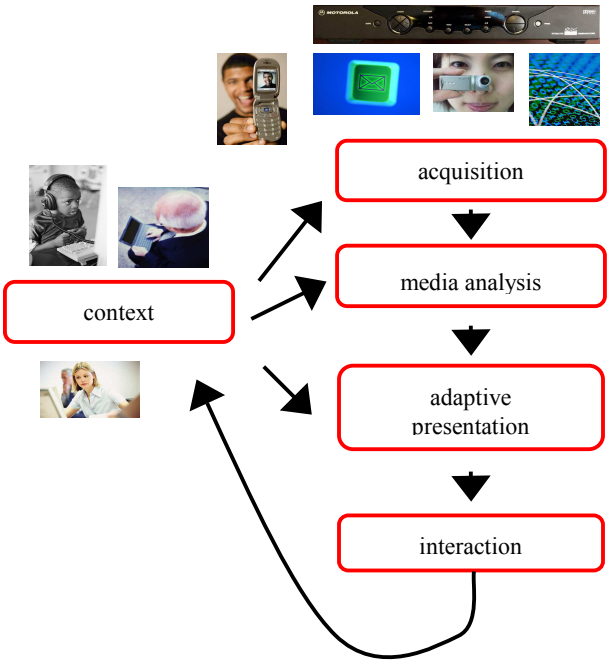

Figure 1: We envision a networked home as a user centric multimedia system. The user context affects the acquisition, analysis and presentation of media. The user interacts with the environment, using novel paradigms, and this in turn affects the user context.

1 INTRODUCTION

This is a position paper on the networked home, framed as a user centric multimedia system. The problem is important – with the advent of high-speed communication networks at home, new exciting applications relating to the communication of human experience through media, have recently begun to emerge. There are problems relating to context [22,31,46], continuous archival [23,36], technology enabled social communication [3,15] and situated systems [39].

The experience of being immersed in an environment that allows high-speed connectivity changes the behavior of the consumer as is well documented in the case of Korea and Japan [10], leading to new mechanisms of communication and interaction. In particular the users typically consume more streaming media (often paid), and alter their everyday habits – for example, more people bring their work home.

We are rapidly acquiring technological artifacts, and communication infrastructures, that allow us to consume, and communicate with media – e-mail, personal digital video recorders, cell phones, e-

mails, web-cams, to name a few. The presence of a high speed network, has not yet led to a significant change in the media experience, even though we are acquiring and communicating media through e-mails, instant messenger programs and cell phones. We are acquiring media at a tremendous rate, yet it is difficult to understand the broader semantic patterns – (a) what are the common interests with a group of friends, (b) who are the people that I listen to? (c) what the media sources that I use the most? Understanding the context in which these communications take place is clearly important in answering these questions.

Importantly, these communication and interaction mechanisms are *isolated* from each other. Let us assume that we are browsing the web, planning for a vacation to Hawaii. However, the digital video recorder will not record programs related to the vacation, as it is unaware of this context. The recommendations are based on interaction with the cable services alone. The current isolationist paradigm, ignores the actual *practices* of the user, whose context will be influenced by all interactions – she will not isolate the web browsing experience from watching television.

We need new paradigms for sharing experiences over a network. The rapid infrastructural changes while improving the speed and frequency of the communication, have not yet substantially impacted the manner in which this experience is communicated. For example, while we can easily send an e-mail from a cell phone with an attached image, the recipient does not have enough context to understand it – this would require more photos, audio and text. The process of *authoring* and communicating a segment of our everyday experience is hard. Also, new visualization techniques to allow us to browse other network member experiences is needed.

We need new media interaction frameworks. The current mechanisms of interacting with media, still implicitly assume the presence regular input devices such as the keyboard, mouse or remote controls. Currently we also assume that we consume media in the same place – either sitting in front of a television, or the home computer. By rethinking the interaction paradigm in terms of tangible user interfaces [39,65] that allow for manipulation of media through physical objects, touch and gesture, as well ambient displays [40] we are beginning to rethink interaction with media. We now present related work on intelligent homes.

1.1 Related Work

There has been prior work in developing the framework for an intelligent home. Mark Weiser's vision of ubiquitous invisible computing [66], lead to several efforts towards pervasive computing spaces at the office and at home. There have been several projects that deal with the idea of a smart and an intelligent home that continuously monitor human behavior and adapt to the human context. They include Aura [35], Microsoft EasyLiving [21], the Oxygen (Intelligent Room) [18] and the Aware Home project [44].

Aura focuses on the pervasive distraction-free computing and investigates the system and network level adaptation issues [35] required for a context-adaptive pervasive computing environment. There the environment consists of simple contextual information like user location, available bandwidth and other environmental information. However, there is not much of a focus on the user interaction with media.

There has been work at MIT, at several locations. Projects at the MIT Media lab (Smart Rooms), AI lab (Intelligent Room [18]) and the Department of Architecture (House_*n* [38]) focus on providing for a pervasive computing environment at home with multiple vision and speech based sensing technologies. The Intelligent room collects information about the user activity and intention based upon speech and vision based sensors . Semantic networks are used to model the user and environment knowledge [52], and a special system – START is used to answer questions using natural language processing. Finally, adaptive interfaces are constructed based upon Haystack [6].

The House_*n* project at MIT [38], is focused on identifying the design principles for the next generation homes that will bridge the gap between the physical and the digital world. The need of user context-sensitivity is acknowledged in the project and the project implements a unique way of gathering context sensitive sampling of user experiences. However, these projects fail to address the question of the new interactions the user will have with the media, the implications of it, and the methods in which the media experience can be enriched, enhanced, shared and efficiently used in a pervasive computing environment.

The Aware Home project focuses on home environments that will be aware of the user context, however again the context is limited to answering a limited set of questions related to the user such as *where, what* and *who* [44] The system continuously monitors user activities to get more information about the user to build the context as well as help in memory augmentation that can be later viewed and analyzed . However, the analysis of experiences requires answering a set of questions that involve understanding of the user more than location, time and activity.

The rest of this paper is organized as follows. In the next section we present our vision, and goals. In section 3 we present our work on user models, in section 4, we discuss out work on the communication of semantics. In section 5, we present an overview of our research on situated systems and then we present our conclusion in section 7.

2 THE VISION

We envision an augmented user-context adaptive home that enables the user to rest, reflect, interact and communicate everyday experiences through multimedia. We are interested in creating interactive multimedia systems that augment the human experience. A key requirement is that they be user-context adaptive – dynamically changing the system analysis, presentation and the interaction to the user context (profile, media history, cognitive skills, behaviors and knowledge goals). Our initial focus has been on interactive computer-based environments. We are expanding our scope to work on problems relating to knowledge propagation between electronic media (audio / video / text) and traditional media (paper / physical objects). We are also staring work on situated multimedia systems that explore novel knowledge acquisition frameworks for a user in a physical setting, such as an intelligent home. There are three long term focus areas for our research:

1. **The individual:** We are interested in creating a robust user context model for multimodal knowledge acquisition that evolves over time, and that enables multimedia systems to effectively adapt this context. Traditionally, multimedia research has focused on common ontologies, and consensual semantics. We believe that multimodal semantics while influenced by common relations are highly individualistic and whose interpretation is changed by the individual context. We are also interested in relationships between individuals to yield group semantics. We believe that user-adaptation will lead to more successful multimedia applications. The problems that

arise here are: (a) theoretical models of context, (b) acquisition, analysis and interaction with the multimodal data over an individual's life and (c) annotation systems

2. **Communication of semantics:** We are interested in the problem of communication of semantics through multimedia. We are interested in both open-loop as well as closed loop systems (video summary vs. interactive exploration.). A key difference between our goals and traditional pattern recognition approaches is the following: we are interested in making media meaningful / interpretable, rather than in understanding the media (classical AI). The difference is the ontological commitment of a human being in the loop in our framework. The problems that emerge out this focus area include: (a) interactive environments (b) storytelling models, (c) networked exploration and sharing of media experiences and (d) new presentation techniques, (d) parsing the structure of behavior.
3. **Situated multimedia systems:** Situated multimedia systems are physically grounded systems (e.g. an intelligent home). They focus on how meaning can be acquired by a human being through novel acquisition (sensors), modeling, interactions (physical objects / touch), as well as new presentation displays (non-computer screen based feedback – e.g. changes in ambient sound). Since the focus in situatedness, real-time interactions become very important. The problems that lie in this focus area are: (a) interactions / feedback mechanisms, (b) real-time analysis approximations, (c) new representation schemes, (d) systemic analysis for issues like semantic stability, convergence rates.

2.1 Goals

We envision an augmented, interactive, multimodal home experience that adapts to user needs. We believe that the interaction must be as far as possible, un-encumbered, and must encourage exploratory interaction. Our vision for the futuristic multimedia infrastructure must serve the following five needs. *User adaptation*: The home must adapt to the user – users have different cultural interests and educational backgrounds, and differ in their goals. *Exploratory Presentation*: The home must allow the user to explore the media in order to understand better the semantics of the media. *Natural Interaction*: While interaction with a keyboard and a mouse will continue to be a dominant paradigm, we believe problems involving un-encumbered interaction with the home will become increasingly important in the future. In particular, we envision interaction involving hand gestures and manipulation of physical objects. *Social impact*: We would like to promote new forms of interaction amongst a small social network (e.g. family, friends) involving the user. *Reflection:* An intelligent home, must provide users mechanisms that allows them to reflect on the everyday experiences.

Our proposed vision is challenging and we believe that the challenges can be overcome by working across disciplines and incorporating knowledge from each discipline.

2.2 Our framework

Our approach is predicated on a single idea – *the user context affects the consumption and communication of media, across all communication, and interaction devices.*

In our framework (see Figure 2), the networked home is connected to the external world using a high-speed network. The devices inside the house are connected using a wired and wireless (802.11b/g) high

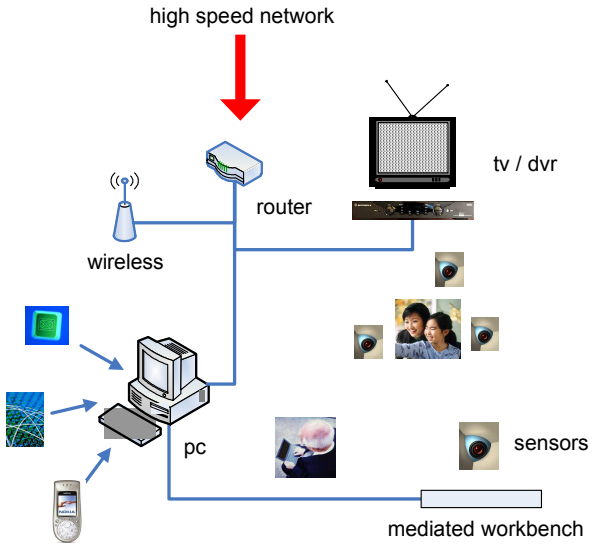

Figure 2: The topology of the devices connected in our framework. The figure shows users in an environment, that allows for multiple interaction, consumption and communication modes.

speed network. Within each home we make several assumptions about the infrastructure:

- We assume the presence of a home computer, with fast components available today (3 GHz processor), and with large main and physical memory (e.g. 512 Mb RAM, 200~300 Gb hard disk capacity). The large capacities are needed for storing information such as video [47].
- The cable box is connected to digital video recorder (DVR), that records media based of each user's context. Different members of the house will have different media recorded for them.
- There are number of cheap video, audio sensors in the house, in particular in areas that require tangible interactions [39].

These assumptions are not unreasonable – the price of disk storage and microprocessor cost is falling exponentially [47]. Hence, we anticipate that we will be able to store as much information as needed in the networked home, on the home computer.

At each interaction point, we make different measurements. At the home computer, we analyze user behavior as she interacts with different messages – e-mails, web pages, creating editing files, in a manner similar to [36]. This will help form the initial user context. The context is communicated to all attached devices (e.g. the DVR, mediated workbench)

At the DVR, we shall analyze the metadata of the media being watched, thus building a user profile. This analysis shall be communicated to the home computer, thus updating the context. The recommendations at the DVR, shall be affected by the global user context. At the mediated workbench, we analyze gesture based interactions with physical objects. The workbench may also contain a projector that augments the interaction. This could be a steerable display [45], easily facilitating user movement, as well a telepresence applications. There is bi-directional communication among the home computer, the DVR and the mediated workbench. Other sensors at home will communicate wirelessly, to the home

computer. Note that in our framework, *all interactions affect the user context*.

We have thus far described our vision, and explained the infrastructural framework of the house. Over the next three sections, we describe our current research in achieving the three long term focus areas – (a) modeling the user context, (b) enabling new communication paradigms, and (c) creating a situated multimedia system.

3 CONTEXT

In this section, we shall present our research on creating user context models [22]. We begin with the challenges, and then give a brief overview of the user context model framework.

3.1 Challenges

An important aspect of the system is the adaptation to the user's needs and the environmental settings, in short the context. The Merriam-Webster [1] defines context as "*the interrelated conditions in which something exists or occurs*." Context is very important in a multimedia system since perception and interpretation of any media is relative to the context in which it is consumed. Secondly actions are also user context dependent.

Prior work on context includes work in the Natural Language Understanding [24], and ubiquitous computing [29]. In pervasive multimedia systems a broader notion of *multimodal context* is required. The three parallel phases of the context model depend upon the set of questions are *context acquisition, context representation* and *context evolution*. Research on the architecture of context aware systems have focused on a modular approach to context modeling that attempts to maintain a demarcation in these phases [30]. Based upon the set of questions and the phases the context can be classified into three types as follows.

Environmental context: The environmental context answers a set of questions related to the user environment such as, *"where", "when", "what"* and similar questions [30]. *User context:* The user context is related to the meaning and user understanding. It answers the questions such as, *"who", "what are the skills", "what are the interests", "what are the goals", "what is the understanding"* among others. Hence, it is concerned with semantic inter-relationships between concepts, which can be arbitrary. *Application context:* The application context answers the questions such as, "what situation and state of the application".

3.2 User Context Model

We address the issues related to context in our work on context models [22]. The relationships between the concepts are arbitrary. Our previous work on context models has successfully been applied to different problems [22,57]. The formal model is defined using a semantic-net – a graph $G = <V,E,W>$ where the nodes $v_i \in V$ represent the concepts, the edges $e_{ij} \in E$ represent the type of relationship (semantic, spatio-temporal, feature-level) between the nodes i and j and $w_{ij} \in$ W, specifies the strength of the relationship between the two nodes. The notion of a concept is multimodal. Thus, a concept node is associated with a specific instance that could be an image, video, an audio segment, text or other media or a combination of one or more of these. We defined the context be the union of semantic-nets:

$$C = \bigcup_{i=1}^{k} G_i \qquad <1>$$

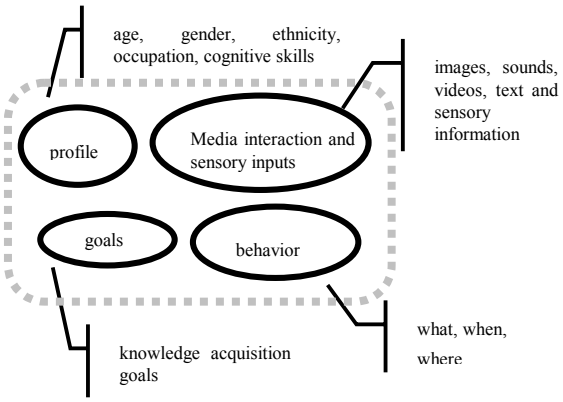

Figure 3: The user context model has four components – (a) profile (b) media history (c) behavior and (d) goals

where C is the context, k is the total number of semantic-nets and G_i is the i^{th} semantic-net. We also discussed (a) the composition and construction of user context (b) the relationship between media and the user context and (c) the context evolution during the user's interaction with the environment.

As depicted in Figure 3, the user context is comprised of concept nets composed from (a) the initial user profile (stating the user's interests, background etc.), (b) user interaction with the media and the system and the sensory information about the user (c) user behavior (d) user's goals. This information is used to make concept nets that together form the user context. The relationships between the concepts like the concept cover and the concept distances were also introduced in the paper. Finally a memory model was introduced to model the temporal evolution of context. The work gives a novel method of modeling multimodal user context with information coming from different sensors. In Figure 6, we show how the context evolves over time, with user interaction .

The context model introduced in is multimodal. However; the cross-modality relationships and the relationships between the concepts represented by media such as images and audio are only statistical and very elementary. More sophisticated and arbitrary relationships are required for multimodal sensory fusion and context modeling.

4 COMMUNICATION

In this section we discuss our steps to create new networked communication paradigms. In the next three sections, we shall discuss the acquisition of media, how media is presented to the user and finally new paradigms of networked communication.

4.1 Media Acquisition

In this section we shall present our framework for media acquisition and annotation. We shall describe in detail the mechanisms by which media is acquired and annotated. We shall defer discussion of the analysis sensory data till a later section.

A networked home, with multiple users will acquire media of interest in many ways (ref. Figure 4) – (a) directly acquired by user (digital photos, music, web downloads), (b) network push (cable broadcast networks, friends publishing media over a private network), and (c) the capture of sensory information about the user interaction with the system. The acquisition is done by taking the user context into account.

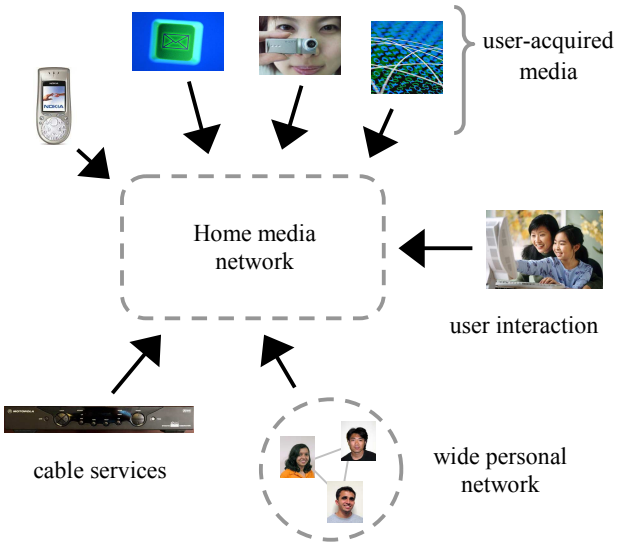

Figure 4: Data in a home media network is acquired from several sources: (a) media due to user from e-mail, web, camera and cell-phone, (b) network (cable, wide personal network) and (c) sensory information about user interaction.

In our proposed framework, the user adds media to the environment through digital camera uploads, cell phones, and media downloaded via e-mail or the web. We expect that the primary location of these media will be the personal computer. Secondly, the personal video recorder will store media based on user preferences that is acquired from cable service providers using regular subscription services. We also expect that the media from the wide personal network (friends and family not co-resident, but connected via broadband) to be present in the environment. The networked media can be viewable across devices at home, in particular will be accessible via the home media recorder and the personal computer.

The user interaction is captured through sensors in the environment. This capture is important since we want to augment the traditional interaction (keyboard / mouse / TV remote) with the home media network with a gesture based interaction. We anticipate that a networked home will contain cheap cameras, microphones, as well as proximity sensors. It is not unreasonable to expect that in networked home with users having a medical condition (e.g. irregular heart-beat) will contain sensors that will help monitor and if needed disseminate this information to the doctor in a real-time basis.

The acquisition process and in particular the media acquisition for presentation is influenced by the context. As for example, if the user behavior on the personal computer shows that user is looking forward to buy a house, the system may find and acquire media related to houses (need to discuss more about this).

4.1.1 Incentive based annotation

We have been working on the problem of creating annotation systems, that give users incentives to annotate [58]. The goal of our work is to create novel semi-automated, intelligent annotation algorithms that bridge manual methods for annotation and fully automatic techniques. There has been prior work in semi-automatic image annotation using relevance feedback [25]. While there are rich mathematical models used, we believe that there are two shortcomings of the current work:

- **Experience**: None of the systems focus on the end-user experience. There is very little return on the enormous time invested by the users to annotate media. Currently annotations only enhance the search capability and *not* the presentations. *A annotation system that provides insight will create incentive in the user to enter richer annotations.*
- **Semantics**: Current approaches to annotate images [25] essentially treat words as symbols regardless of their semantic relationships with other words, which is no different than any normal image feature. The lexical meaning of the keywords/annotations is not exploited.

In this work we address issues relating to both semantics and the end user experience. We establish the semantic inter-relationships amongst the annotations by using WordNet [50]. We attempt to map the annotation problem as one of an experiential system [42,60] – the key idea being that the user will gain insight about the media in relation to her context, thus providing a strong incentive for the user to annotate the media.

The annotation procedure is as follows. Our annotation system uses a combination of low-level as well as WordNet distances to propagate semantics. As the user begins to annotate the images, the system creates positive example (or negative examples) image sets for the associated WordNet meanings. These are then propagated to the entire database, using low-level features as well as WordNet distances. The system then determines the image that is least likely to have been annotated correctly and presents the image to the user for relevance feedback.

The system also attempts to provide insight by presenting knowledge sources to the user. This is done using context-aware hyper-mediation. In our approach, we use Google to automatically generate hyperlinks. This is done by taking into account the user profile, the semantics of the media and the semantic relationship between the media item and the user profile.

4.1.2 Media analysis

In this section we shall summarize our current work in the analysis of media that are being consumed by the user, as well as the media transmitted or communicated to other users.

We are attempting to create a framework, that analyses messages to and from the user to answer high-level semantic questions. For example, if the user is engaged in activity to buy a house, then we would be interested in asking several related questions – (a) which friends were my influences, (b) which information sources do I trust, and (c) what were my main concerns during this time.

We choose to analyze messages, as they form the primary mechanism of communication. A message is a packet of information that comprises one or more entities (text, audio, video, sensory data). Examples include – e-mails, recorded audio / video among others. Messages always have an author and the recipients could be self, another user (or group), or to a network (sensor, web). We are proposing a simple hierarchical communication model, where social communications (conversations, monologues, broadcasts, web browsing, creating and editing of electronic documents) are conducted though and exchange of messages (emails, chat transcripts, web page contents, documents, speech, video broadcasts etc.) that comprise entities (text, images, audio and video). This is summarized in Figure 5. An entity or message that is connected to

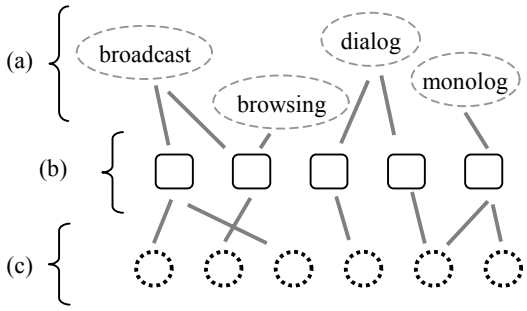

Figure 5: Message model – (a) different social communication modes, (b) messages that are used in the communication and (c) specific multimodal entities that comprise the message.

more than one element at a hierarchical level above, implies that the entity or message is being re-used.

4.2 Media presentation

In this section we shall discus the problem of communicating media to the user. We first begin with the challenges associated with media communication and, then discuss our context adaptive scheme. We conclude by presenting ideas on ambient displays.

4.2.1 Challenges

Traditional computer-based mechanisms of communicating media to the user, that support exploration and interaction, are data and task driven. Image based storyboards offer a key-frame based non-linear navigation of the video data [64]. In [28], the authors explored different visualization schemes – semantic relationships, maps and time lines. A video skim [27] is a short audio-visual clip that summarizes the original video data.

Our review of current work indicates that while there are examples of specific interactive exhibits (e.g. [59]) , there are no information design tools available that would use the *same tools and mechanisms across multiple contexts* (i.e. the generalizations to other spaces / criteria is not easily possible.). There is also no principled user-centric, context-aware strategy for content creation to address issues such as how and what multimedia data needs to be shown, and how are these are to adapt to the user context. We now describe our approach to the problem.

4.2.2 Novel visualization schemes

We are conducting research on novel, context aware multimodal presentation schemes that adapt in real-time to user actions.

The presentation problem can be broken down into three sub-problems: (a) media selection (b) media presentation and (c) synthesis. We plan on developing *a joint* optimization framework that determines the correct media, presentation method and the synthesis simultaneously. For all the problems, we plan on expanding on recent work on summarization (in the film domain) and presentation [61] that investigated media presentation as an utility maximization problem. There, the presentation method was fixed (a video skim), but the media elements and the synthesis was done dynamically based on a utility framework. The synthesis was constrained by syntactical and structural properties of the media.

We have prototyped our ideas in an interactive context adaptive environment [22] that allows the users to explore concepts in geography. The framework is multimodal, and dynamic – the user-actions and needs affect the presentation.

In [22], we developed a novel multimodal learning environment involving concepts in geography, for active learners. Active learners are people who prefer to learn by doing and experiencing rather than by reflecting. The problem is important because traditional pedagogical techniques focus on reflective learners, however a significant number of children are known to be active learners. We investigated the relationship between comprehension of a multi-modal concept to its representational complexity, and we modeled the dynamics of the interaction to maximize coherence. Our presentation environment incorporates graphic design principles, when synthesizing the environment.

The environment comprised a of three maps (urban / sub-urban / rural), that could be interactively explored based on geographical location. Each visualization is uses a collage of images and text, that are interactively revealed (as a slideshow) with user interaction. Since each concept is multimodal, the audio and the visual elements need to be synchronized in time, as well as rendered for an appropriate time for the content to be understood.

The purpose of slideshow sequence is to emphasize some content as being associated, as well as ensure that some content is presented in such a manner as to maximize its potential to be comprehended clearly and quickly. The design challenges – text and images that are mutually dependant should be positioned in close proximity to one another in order to offer a clear spatial indication of what information should be associated together and what should not. Attention to this sort of detail is of particular importance when content is arranged in a collaged (i.e. overlapped) manner, since it may be quite easy to accidentally pair information of which such immediate association is not intended.

We also used sonic collages for conveying information about geographical concepts and population density, and increase their effectiveness by ensuring that: (a) the selection of audio samples and their presentation is informed by the context, (b) we use auditory information in addition to connotation to shape sonic environments and exaggerate important concepts and (c) the sonic environment is constantly evolving and transforming in real time as a user explores the virtual space. The sonic environment transforms as a user navigates from one virtual location to another. This is summarized in Figure 6.

4.2.3 Ambient soundscapes

In this section we describe our recent work and vision for the use of sound-based ambient displays [40,65] in the home. These are presentation devices that allow the communication of information

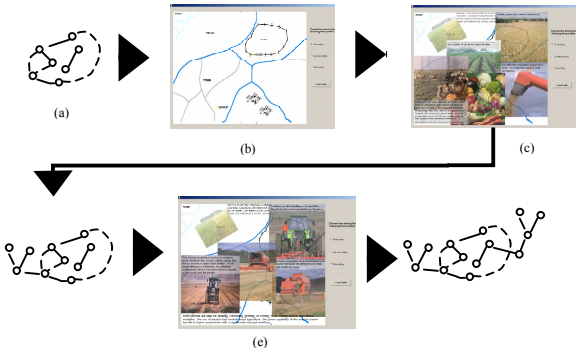

Figure 6: The graphs ((a), (c) , (f)) represent the context for the same user at different times.

through slight changes in lighting and sound that utilize mechanisms that are latent in human peripheral awareness without demanding a user's full attention.

We specifically focus on the problem of creating context aware soundscapes that will foster reflection and contemplation in the home, and go beyond the problem of sonifying information such as stock-market quotes (e.g. [40]). Through the introduction of imaginative sound collages and soundscapes, that are informed by the environmental and the user's context, we can both transform the home environment and employ it as a sophisticated multimedia communication hub.

Soundscapes and sound collages are capable of creating a strong sense of time, place, and reference. This has been documented in analytical work from musicians [67], theorists [9,16], and ethnomusicologists [34,37]. Soundscape collages from the real world have been captured and studied [56], and have been artificially synthesized for artistic purposes [62]. In [51] the authors propose an approach to the automated generation of soundscapes in their work on Audio Aura. There they dynamically generate sonic seascapes to convey information in specific applications such as email alerts and ambient displays. However, their generative mechanism is hindered by a small palette of sound sources and a narrow focus on information display. In our recent work [11,13,12], we have developed generative model for the automated synthesis of sonic environments and music. Although these models have yielded successful musical results, they do not focus on the home. We propose to extend these models to generate sonic environments and ambient displays that are specifically geared for presentation in the home.

The introduction of broadband in the home provides the opportunity to deliver high quality audio to users on the local network and to communicate audio between users across the internet. Much attention has been focused on the sharing of pre-produced, commercial music between users over the internet [5]. Although such sharing of data is facilitated by the broadband home, our framework is focused on experiential construction through generative mechanism that will synthesize sounds that are recorded from a user's everyday experiences inside and outside the home.

We are developing models that will not simply replay these sonic events, but, informed by the user and environmental context, will present these media elements in the home in such a way that will promote reflection and contemplation. Furthermore, through the use of broadband communications, we propose to remediate a user's experiences in the homes of friends or family as a means of communicating ideas and events not simply through narrative or descriptive methods, but rather, through experiential construction.

For example, if a user takes a vacation to Hawaii, she might record the sound of waterfalls, the ocean, the forest, boats in a bay, and the laughter of a loved one with children. In our recent work [8] we describe practical methods and tools for media documentation of everyday experiences. As in our previous models, these collected soundfiles would be added to a database of media objects in the home and annotated by the user. Later, either at the user's request or in response to the environmental context, these sounds can be used to immerse the home occupants in the sounds of the vacation. This soundscape is not simply a sequential playback of the recorded sounds, but instead, the generative model constructs an imaginative and rich soundscape that broadly and imaginatively communicates the experience of the vacation using the collected media. Additionally, using the capabilities of broadband networking,

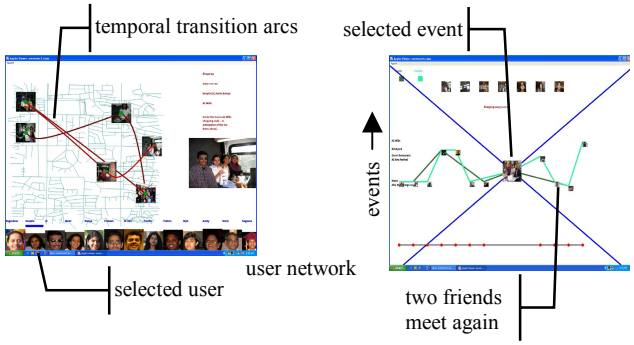

Figure 7: (a) Spatio-Temporal evolution of digital media in our system. (b) An event cone. This is a snapshot of all the events associated with the selected users.

friends of the vacationers can experience a similar ambient display in their home that is adapted to suit their own experiences and contexts.

Extensions of our previous work will focus on two crucial aspects of sound collage creation, *selection and presentation of sounds, and the filtering and enrichment of those sounds for affect*. In the proposed work, relevant sound samples are derived from the sounds that the user heard during the day, and are stored in a database. We propose to systematically explore the influence of soundscape attributes such as density, amplitude, the diversity of sounds, and sonic familiarity for users. Once the basic form and material of the sound collage has been constructed, we propose to explore auditory effects such as reverberation, and frequency based filtering, to orient these soundscapes in different times and spaces that will be meaningful for individual users and homes.

4.3 Networked sharing

A key part of our vision is to be able to communicate and share media experiences within members of a social network. We have been working on the problem of communicating everyday media experiences amongst members of a disconnected network [7,8] – the goal is to allow a network of users to explore the activities of a set of friends. This is an emerging problem in several contexts: (a) sharing of media is important in online social-networks such as Friendster [3], where a set of friends share a set of multimedia experiences. (b) In the continuous archival of personal media [36], and an multimodal event browser is key to efficiently navigating such a large event set, and help provide insight.

In this work we develop a simple multimodal event model. We use the dictionary [1] definition of an event: *something that happens*. In our framework, events have the following properties associated with them: name, location, time, media elements (set of images, sounds, text), as well as participants. There are two novel aspects to our model – (a) a set of viewpoints is attached to each event and (b) the semantic relationship between event is dependent on user context. We also develop an event-centric user context model, that adapts to user behavior.

We develop an task-driven event exploration environment – (a) visualize what a particular user has done after (or prior) a specific event (*event conditioning*), (b) visualize the event sequence that lead to two users to meet (*event support*) and finally (c) for a given user, determine *interesting events*.

The environment has three exploration environments – (a) spatio-temporal evolution, (b) event-cones, and (c) viewpoint-centric

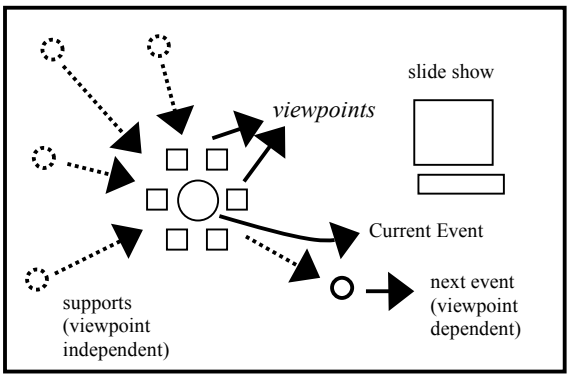

Figure 8: Viewpoint based exploration: Every event has a set of supporting events and one succeeding event, according to the currently selected viewpoint. The user can dynamically change the viewpoint, thus changing the slideshow associated with the viewpoint, as well as the future event.

evolution. Our *spatio-temporal visualization* combines space and time information. We dynamically generate maps using online geographic data available in XML format [2], on which events related to a selected user, unfold over time and space. For each event, it will also create a slideshow of images and text. *Event cones* are snapshot of all the events associated with the selected users. The visualization is important since it allows a non-linear exploration of events, much like our natural thought process. The *viewpoint centric* visualization (Figure 8) allows the user to brows the event space by following the individual viewpoint through the event space. Our user studies indicates that the exploratory environment is very well received.

4.3.1 Telepresence

Telepresence is a key application for the networked home. Given the popularity of webcams, the high-definition telepresence will have a important societal impact. We present a brief overview, since we are not currently active in this area, but networked communication of presence (e.g. [32]) is an area of interest.

In [41], Jain mentions two important challenges that need to be tackled – (a) formation of real, dynamic world model, and (b) and rendering a multisensory environment in real-time. Contemporary work on immersive telepresence includes 'Office of the future' project [55]. The project is based on a unified application of computer vision and computer graphics to bring to the user real-time immersive displays for normal office work and telepresence using projectors and cameras. The project focuses on the problems of acquisition of the remote scene, modeling of the world, tracking of user's eye positions, rendering with respect to the user's views and the stereo presentation for the user. The project demonstrates adaptation of the display with the user context environmental conditions . However, the user context is limited to the eye gaze of the user.

Another work by Pingali et. al [53] focuses on steerable projected displays that move the display along with the user. The authors discuss a system with multiple cameras that capture the user movement and activity and multiple steerable projectors that present the display on the real-life objects close to the user. Multiple issues like the distortions due to the shape and color of the projected surface, object on which the display is projected, the obliqueness of the projection and the user obstruction in the projection path are discussed in their work. Pingali et. al. also discuss the concept of vision based interfaces [53].

A humanistic critique [49] of telepresence is that it attempts to communicate meaning by employing literal communication – exactly communicating the world. More semantically efficient communication may be possible with subsampling, and creating new representations of the world.

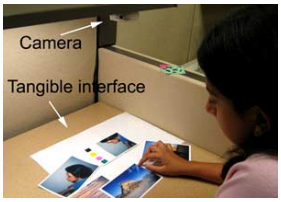

Figure 9: interacting with physical prints of digital photos

5 SITUATED MULTIMEDIA SYSTEMS

Situated multimedia systems are physically grounded systems (e.g. an intelligent home). They focus on how meaning can be acquired by a human being through novel acquisition (sensors), modeling, interactions (physical objects / touch), as well as new presentation displays (non-computer screen based feedback – e.g. changes in ambient sound).

Situated systems are motivated by the work of Rodney Brooks [17,19,20] and Pattie Maes' earlier work on autonomous agents [48]. However while there are similarities with their work, there are important differences. In a situated multimedia system, the human being (unlike a robot in Brooks' case) is the embodied agent, that is also not controlled. However, the world that the human inhabits is partially controlled by the environment (unlike Brooks' chaotic world), using ambient auditory displays and visual projective displays. It is the systems goal, to enable the user to pursue specific tasks.

A situated system requires continuous monitoring, analysis, feedback (in terms of audio and video) and interaction, and is affected by changes in the user context. The user context will change upon her interaction with the visualization / presentation. There are important questions that arise here – (a) novel interaction feedback mechanisms, (b) understanding the syntax of interaction, (c) approximations for real-time analysis, (d) the closed loop brings concerns such as convergence and stability. Convergence here implies that that the user is able to complete her task in a finite time, and stability implies that the system prevents the user from wandering off in direction completely unrelated to the goal. For example, if the user is interested in learning more about a trip to England that occurred a few years ago, the system should enable this user to complete the task in finite time. Over the next few sections, we discuss several key components of the intelligent home that form the current focus of our research. We shall discuss novel interaction paradigms, interpretation of interaction syntax, and approximation frameworks for real-time systems.

5.1 Interaction paradigms

We now discuss our current work on developing novel interaction paradigms, for our augmented home. The work is inspired by recent research into tangible user interfaces [39].

Tangible user interfaces are interaction mechanisms that involve physical artifacts *both* for representation and control. There are four characteristics that distinguish them from traditional graphical interfaces – (a) The objects are meaningful in their own right, and are coupled to computational models, (b) the objects also serve to control, (c) the objects are perceptually coupled to actively mediated digital representations, and finally (d) they also serve as

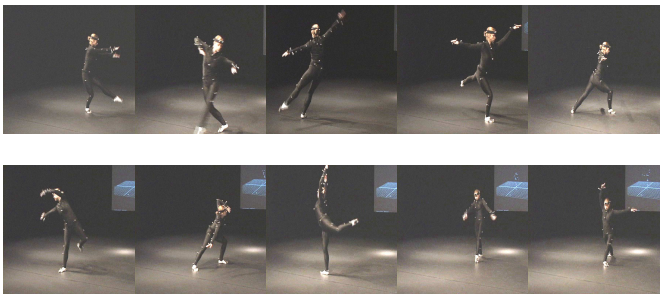

Figure 10: Movements showing phrase A (above) and movements showing phrase B (below)

representations of the current state of the system. Note that in traditional graphical interfaces, the controls (keyboard / mouse) is separate from the feedback (graphical display).

We are developing a novel interaction framework that will enable home users to explore digital photographs using physical prints. The problem is interesting since we want to give users the ability to reflect on augmented digital photos, through their interactions with the tangible interface, which incorporates dynamic audio visual feedback. We are assuming that users will have annotated the digital media perhaps using the techniques in Section 4.1.1 and will want to print out a few of the digital prints. Hence digital print will have metadata such who, when, where and what, in addition to other fields, associated with them.

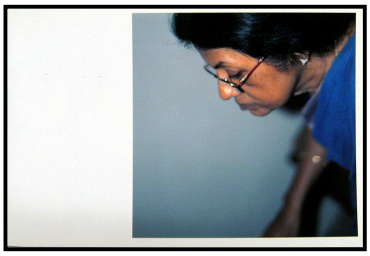

Figure 11: Each photo is printed on standard paper, and is augmented with physical markers.

In our proposed scenario, each physical print of the digital photo, is augmented with a set of easily tracked markers. The markers allow us to quickly identify the photo and hence associating it with metadata stored in the database. The user can interact with the prints in real time, using natural hand gestures. The system *overlays* visualizations using a projector, that is optically co-located with the camera. The visualization is displayed on top of the photos, thus encouraging the user to interact with the digital and projected images, thus augmenting the mechanism of interacting with the photo. We are also planning to extend this work in two ways – (a) allowing a group to interact with the system in a natural manner, (b) allow users in *remote* locations to discuss digital photographs using physical artifacts.

While our current research interest is in tangible interactions, we foresee exciting opportunities for other interfaces such as speech and touch, depending upon the user and application context. Interaction with tangible media brings other interesting challenges of *interpreting a sequence of gestures* for control. We now present concurrent activities in structure detection in our group.

5.2 Structure in interaction

We now discuss our attempts to detect structure in movement [33]. This problem is relevant to our framework, since we are interested in parsing the syntax of the interaction of the user with the system, and detect structures meaningful to the interaction. For example, a specific sequence of hand gestures may trigger of a specific feedback (change of TV programs / initiation of a telepresence session.)

Our focus thus far has been on detecting phrasal structures in dance. The dance domain contains many examples of highly structured movement. The problem is important since structure plays an important role in representing and synthesizing meaning in dance [14]. The solution to structure detection will impact multimedia systems that attempt to summarize the contents of captured dance, as well as other surveillance systems that aim to detect structures in movement.

There has been prior research on gesture boundary segmentation and detection but limited work on phrase detection. In [54], the authors have developed a real–time system that can be used for posture recognition in dance. Work in [43], dealt with gesture segmentation using a dynamic hierarchical layered structure to model the human body and activity measures in human body segments to find gestures in a motion sequence. Note that *a phrase is a sequence of movements, that exists at higher level of semantic abstraction than gestures*.

We solve the problem of phrasal structure detection in the following way. First, we identify fundamental structures in cotemporary western dance – ABA and the Rondo. Then we determine robust features (kinetic energy, momentum, and force) that incorporates both the hierarchical body structure and segment movement. The distance between two phrases is computed using dynamic programming, since the phrases are of unequal length due to human variation.

We present the idea of a topological graph [33] for phrasal structure detection. Given the associated topological matrix, we show how to derive an objective function whose minimization shall determine the phrase change boundaries, thus confirming the presence of structure. The dances for the experiments, was created by an expert dancer, and was acquired using an eight camera VICON motion-capture setup. The results show that algorithm is robust – the structures are detected with very low median error 7% (ABA) and 15%. (Rondo).

Our goal to incorporate knowledge gained in this experiment, to detecting other temporal structures due to interaction, that are sensed using video / auditory / pressure sensors.

5.3 Real time issues

A situated system needs a real-time response. We have been collaborating with other researchers on the problem of creating a real-time situated space architecture with QoS guarantees – ARIA [4]. ARIA will enable design, simulation, and execution of interactive performances. ARIA provides a language interface for specifying intended mappings of the sensory inputs to audio-visual responses. The specific outcomes of this project include the following: (1) an adaptive and programmable kernel that can extract, process, fuse, and map media-flows while ensuring quality of service guarantees, (2) a specification language and user interface capable of specifying the components of the media-flow network, (3) QoS scalable fusion and filter

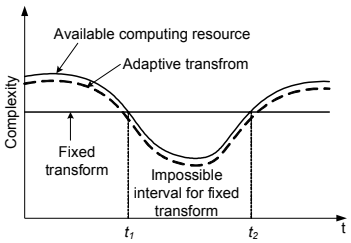

Figure 12: complexity for fixed and adaptive transforms

operators, (4) a framework that integrates the input sensors, media pathways, external data sources, output sensors, run-time kernel, and the specification language into a real-time quality-adaptive media-flow architecture.

As part of ARIA, we have developed a highly novel framework for linear transforms that adapt to changing computational resources [26]. The problem is important since in multimedia systems, the computational resources available to content analysis algorithms are not fixed, and a *generic* computationally scalable framework for content analysis algorithms is needed.

For example in Figure 12, we show an example of a system shows computational resources are changing over time. However, for a fixed analysis transform (e.g. FFT / DCT) there will be a time between t_1 and t_2 when the transform cannot operate at all. We seek an adaptive transform that is able to gracefully adapt to the resources available, but with greater error. The problem is made difficult since the relationship between computational resources and distortion depends on the specific content. Our experimental results indicate a *convex, complexity- distortion relationship* and our preliminary results indicate that our adaptation technique performs well.

In the previous three major sections, we have discussed our research in the three focal areas – (a) context models, (b) novel communication paradigms and (c) new interactive, multimodal systems. We now briefly address some other concerns.

6 OTHER CONCERNS

This paper deals with a narrow set of topics of interest to us at the Arts Media an Engineering Program. We now discuss other important concerns, not covered in this paper.

- **Privacy:** The sensors and the context models discussed here do raise privacy concerns. While this is not a current focus of our research, we expect that other research groups will address issues of access control, encryption among other issues. Interestingly, recent studies show that young consumers today are not as concerned about privacy, as are members of earlier generations [63]. This manifests itself in various ways – people are now blogging details about their lives online, and are increasingly used to communicating using webcams.
- **Intellectual property:** In our framework, we have assumed that media will freely travel amongst devices in a home, and with the wide personal network of each user. However, for paid content, this raises an important intellectual property issue that will need to be addressed appropriately.
- **Affect / behavioral analysis:** This paper does not deal with research that is focused on the analysis of behavior inside a home, or event the emotional state of its residents. These are certainly important indicators of media consumed (or communicated), but we are just attempting to limit the scope of our research.

7 CONCLUSIONS

We envision an augmented user and environment-context adaptive, networked home that enables the user to, reflect upon, interact with, and communicate everyday experiences through multimedia. We believe that the practice of interacting with different interaction communication paradigms, are affected by the user context – the user does not watch television or use the personal computer independent of the other.

We have focused our research around three areas: (a) context, (b) communication of meaning and (c) situated interaction. We explained the challenges of modeling context and examined different context frameworks. We presented our multimodal, semantic net as a potential solution for modeling user context. We discussed problems and proposed solutions relating to media acquisition (annotation, message based analysis), presentation (multimodal collages, ambient auditory displays for reflection), and sharing (networked interaction for media browsing, telepresence). Finally we presented out situated interaction paradigm. These refer to physically grounded multimedia systems, and we discussed three problems relating to interaction, parsing interaction syntax and we also discussed real-time issues.

A networked home, with a high speed network is already having a profound influence in countries such as Korea and Japan. It is our belief that the emerging networked home requires new mechanisms of consumption and interaction. This paper is an attempt to formulate some of the applications that are needed as we look forward to redefining the home as an environment for pervasive multimedia construction, consumption and communication. While the specific technological scenarios in our vision are still fairly conservative (they use current technological trends, and look at the home in just a few years from now), we believe that networked, multi-sensory interaction will have a profound effect on our lives. Importantly, there are many issues relating to privacy, intellectual property and behavioral understanding that are beyond the scope of this investigation. These issues are important to us as well, and we shall be examining research possibilities in these areas in the future.

8 REFERENCES

[1] *Merriam Webster Dictionary* http://www.m-w.com.

[2] *TIGER maps* http://imagemaps.mle.ie/.

[3] *Friendster* http://www.friendster.com.

[4] *ARIA* http://aria.asu.edu.

[5] *napster* http://www.napster.com.

[6] E. ADAR, D. R. KARGER and L. STEIN (1999). *Haystack: Per-User Information Environments*, Conference on Information and Knowledge Management, 1999.

[7] P. APPAN and H. SUNDARAM (2004). *Networked event exploration and interaction summarization*, to appear in ACM Multimedia 2004, also AME-TR-2004-10, New York, New York, Oct. 2004.

[8] P. APPAN, H. SUNDARAM and D. BIRCHFIELD (2004). *Communicating everyday experiences*. Arts Media and Engineering Program, Arizona State University, AME-TR-2004-07, Jun. 2004

http://ame2.asu.edu/groups/xdg/pubs/ame-tr-2004-07.pdf.

[9] J. BALLAS and J. HOWARD (1987). *Interpreting the Language of Environmental Sounds.* Environment and Behavior.

[10] K. BELSON and M. RICHEL (2003). America's Broadband Dream Is Alive in Korea. The New York Times. New York.

[11] D. BIRCHFIELD (2002). for Susan.

[12] D. BIRCHFIELD (2003). *Genetic Algorithm for the Evolution of Feature Trajectories in Time-Dependent Arts*. 6th International conferene on Generative Art. Milan, Italy,

[13] D. BIRCHFIELD (2003). *Generative Model for the Creation of Musical Emotion, Meaning, and Form*, ACM SIGMM 2003 Workshop on Experiential Telepresence, Berkeley, CA,

[14] L. A. BLOM and L. T. CHAPLIN (1982). The intimate act of choreography. Pittsburgh, Pa., University of Pittsburgh Press.

[15] D. BOYD (2004). *Friendster and Publicly Articulated Social Networks*, Conference on Human Factors and Computing Systems (CHI 2004), Vienna, Austria, April 24-29, 2004.

[16] A. BREGMAN (1990). Auditory Scene Analysis, MIT Press.

[17] R. BROOKS (1986). *A robust layered control system for a mobile robot.* IEEE Journal of Robotics and Automation **2**(1): 14-23.

[18] R. BROOKS (1997). *The Intelligent Room Project*, 2nd International Cognitive Technology Conference, Aizu, Japan, 1997.

[19] R. A. BROOKS (1991). *Intelligence Without Reason*, International Joint Conference on Articial Intelligence, Sydney, Austrailia, pp. 569-595, Aug. 1991.

[20] R. A. BROOKS, C. BREAZEAL, R. IRIE, et al. (1998). *Alternate Essences of Intelligence*, Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), Madison, Wisconsin, pp. 961-976, July 1998.

[21] B. BRUMITT, B. MEYERS, J. KRUMM, et al. (2000). *EasyLiving: Technologies for Intelligent Environments*, Handheld and Ubiquitous Computing, September, 2000.

[22] J. BRUNGART, H. SRIDHARAN, A. MANI, et al. (2004). *Adapting Multimedia Design To Context: A design framework for interactive, user context-adaptive multimodal learning environments,*. Arts Media and Engineering Program, Arizona State University, AME-TR-2004-08, Jun. 2004

http://ame2.asu.edu/groups/xdg/pubs/ame-tr-2004-08.pdf.

[23] V. BUSH (1945). *As We May Think*. The Atlantic Monthly. **176:** 101-108, http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm.

[24] S. W. K. CHAN and J. FRANKLIN (2003). *Dynamic Context Generation for Natural Language Understanding: A Multifaceted Knowledge Approach.* IEEE Transactions on Systems, Man, and Cybernetics **33**(1): 23-41.

[25] E. CHANG, K. GOH, G. SYCHAY, et al. (2003). *CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines.* IEEE Transactions on Circuits and Systems for Video Technology **13**(1): 26-38.

[26] Y. CHEN and H. SUNDARAM (2004). *Approximate Multimedia Transform for Real Time Applications*. Arts Media and Engineering Program, Arizona State University, AME-TR-2004-06, Apr. 2004

http://ame2.asu.edu/groups/xdg/pubs/ame-tr-2004-06.pdf.

[27] M. G. CHRISTEL, M. A. SMITH, C. R. TAYLOR, et al. (1998). *Evolving video skims into useful multimedia abstractions*, Proceedings of the SIGCHI conference on Human factors in computing systems, Los Angeles, California, United States, 171-178, 1998.

[28] M. G. CHRISTEL, A. G. HAUPTMANN, H. D. WACTLAR, et al. (2002). *Collages as dynamic summaries for news video*, Proceedings of the 10th ACM international conference on Multimedia, Juan Les-Pins, France, 561-569, Dec. 2002.

[29] A. K. DEY and G. D. ABOWD (1999). *Towards a Better Understanding of Context and Context-Awareness*, Proceedings of the 3rd International Symposium on Wearable Computers, San Francisco, CA, pp. 21-28, October 20-21, 1999.

[30] A. K. DEY (2000). *Providing Architectural Support for Building Context-Aware Applications*. College of Computing. Atlanta, Georgia Institute of Technology, PhD.

[31] A. K. DEY (2001). *Understanding and Using Context.* Personal and Ubiquitous Computing Journal **5**(1): 4-7.

[32] K. DOBSON, D. BOYD, W. JU, et al. (2001). *Creating visceral personal and social interactions in mediated spaces*. CHI '01 extended abstracts on Human factors in computer systems, ACM Press**:** 151--152.

[33] V. M. DYABERI, H. SUNDARAM, J. JAMES, et al. (2004). *Phrase Structure Detection in Dance*, to appear in ACM Multimedia 2004, also AME-TR-2004-05, New York, New York, Oct. 2004.

[34] S. FELD (1982). Sound and Sentiment: Birds, Weeping, Poetics, and Sound in Kaluli Expression. Philadelphia, PA, University of Pennsylvania Press.

[35] D. GARLAN, D. SIEWIOREK, A. SMAILAGIC, et al. (2002). *Project Aura: Toward Distraction-Free Pervasive Computing.* IEEE Pervasive Computing.

[36] J. GEMMELL, G. BELL, R. LUEDER, et al. (2002). *MyLifeBits: fulfilling the Memex vision*, Proceedings of the 10th ACM international conference on Multimedia, Juan Les-Pins, France, pp. 235-238, Dec. 2002.

[37] B. GOODEY (1974). Images of Place: Essays on Environmental Perception, Communications and Education. Birmingham, England, University of Birminghan, Centre for Urban and Regional Studies.

[38] S. S. INTILLE (2002). *Designing a Home of the Future.* IEEE Pervasive Computing: pp. 80-86.

[39] H. ISHII and B. ULLMER (1997). *Tangible bits: towards seamless interfaces between people, bits and atoms*. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press**:** 234--241.

[40] H. ISHII, C. WISNESKI, S. BRAVE, et al. (1998). *ambientROOM: integrating ambient media with architectural space*. CHI 98 conference summary on Human factors in computing systems, ACM Press**:** 173--174.

[41] R. JAIN (2000). *Real reality.* IEEE Computer Graphics and Applications **20**(1): 40-41.

[42] R. JAIN (2003). *Experiential Computing.* Communications of the ACM **46**(7): 48-55.

[43] K. KAHOL, P. TRIPATHI and S. PANCHANATHAN (2003). *Gesture Segmentation in Complex motion sequences*, Proc. IEEE International Conference on Image Processing 2003, Barcelona, Spain, Sep. 2003.

[44] C. D. KIDD, R. ORR, G. D. ABOWD, et al. (1999). *The Aware Home: A Living Laboratory for Ubiquitous Computing Research*, International Workshop on Cooperative Buildings, October 1999.

[45] R. KJELDSEN, C. PINHANEZ, G. PINGALI, et al. (2002). *Interacting with steerable projected displays*, Proceedings. Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002., pp. 387-392, 2002.

[46] H. LIEBERMAN and T. SELKER (2000). *Out of Context : Computer Systems that Adapt To, and Learn From, Context.* IBM Systems Journal **39**(3,4): 617-631.

[47] P. LYMAN and H. R. VARIAN (2003) *How Much Information* http://www.sims.berkeley.edu/how-much-info-2003.

[48] P. MAES (1994). *Modeling adaptive autonomous agents.* Artif. Life **1**(1-2): 135--162.

[49] L. MANOVICH (2001). The language of new media. Cambridge, Mass., MIT Press.

[50] G. A. MILLER, R. BECKWITH, C. FELLBAUM, et al. (1993). *Introduction to WordNet: An On-line Lexical Database.* International Journal of Lexicography **3**(4): 235-244.

[51] E. MYNATT, BACK, M., WANT, R., BAER, M., ELLIS, J.B. (1998). *Designing Audio Aura*, SIGCHI conference on Human factors in computing systems,

[52] S. PETERS and H. SHROBE (2003). *Using Semantic Networks for Knowledge Representation in an Intelligent Environment*, 1st Annual IEEE International Conference on Pervasive Computing and Communications, Ft. Worth, TX, USA, March, 2003.

[53] G. PINGALI, C. PINHANEZ, A. LEVAS, et al. (2003). *Steerable Interfaces for Pervasive Computing Spaces*, IEEE International Conference on Pervasive Computing and Communications, Dallas-Fort Worth, Texas, March 2003.

[54] G. QIAN, F. GUO, T. INGALLS, et al. (2004). *A Gesture Driven Multimodal Dance System*, IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, June 2004.

[55] R. RASKAR, G. WELCH, M. CUTTS, et al. (1998). *The Office of the Future : A Unified Approach to Image-Based Modeling and Spatially Immersive Displays*, ACM SIGGRAPH, Orlando, Fl, USA,

[56] R. M. SCHAFER (1973-78). The Music of the Environment Series. R. M. SCHAFER. Vancouver, A.R.C. Publications.

[57] B. SHEVADE and H. SUNDARAM (2003). *Vidya: An Experiential Annotation System*, 1st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia 2003, Berkeley, CA, Nov. 2003.

[58] B. SHEVADE and H. SUNDARAM (2004). *Incentive Based Image Annotation*. Arts Media and Engineering Program, Arizona State University, AME-TR-2004-02, Jan. 2004 http://ame2.asu.edu/groups/xdg/pubs/ame-tr-2004-02.pdf.

[59] F. SPARACINO, K. LARSON, R. MACNEIL, et al. (1999). *Technologies and methods for interactive exhibit design: from wireless object and body tracking to wearable computers*, International Cultural Heritage Informatics Meeting, Washington D.C., Sep. 1999.

[60] H. SRIDHARAN, H. SUNDARAM and T. RIKAKIS (2003). *Computational models for experiences in the arts and multimedia*, 1st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia 2003, Berkeley CA,

[61] H. SUNDARAM, L. XIE and S.-F. CHANG (2002). *A utility framework for the automatic generation of audio-visual skims*, ACM Multimedia 2002, Juan-les-Pins, France, ACM Press, 189-198, Dec. 2002.

[62] B. TRUAX (1996). *Soundscape, Acoustic Communication and Environmental Sound Composition.* Contemporary Music Review **7(1)**: 49-65.

[63] S. TURKLE (2004). *How Computers Have Changed the Way We Think*. The Chronicle of Higher Education: Information Technology

[64] S. UCHIHASHI, J. FOOTE, A. GIRGENSOHN, et al. (1999). *Video Manga: generating semantically meaningful video summaries*, Proceedings of the 7th ACM international conference on Multimedia, Orlando, Florida, USA, 383-392, 1999.

[65] J. UNDERKOFFLER, B. ULLMER and H. ISHII (1999). *Emancipated pixels: real-world graphics in the luminous room*. Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co.**:** 385--392.

[66] M. WEISER (1991). *The Computer for the 21st Century.* Scientific American.

[67] K. WRIGHTSON (2000). *An Introduction to Acoustic Ecology.* Soundscape, The Journal of Acoustic Ecology(1(1)): 10-13.