Abstract

As computing technology continues to become increasingly pervasive and ubiquitous, we envision development of environments that can sense what we are doing and support our daily activities. In this article we outline our efforts toward building such environments, and discuss the importance of a sensing and signal-understanding infrastructure that leads to awareness of what is happening in an environment and how it can best be supported. Such an infrastructure supports both high- and low-end data transmission and processing, while allowing for detailed interpretation, modeling, and recognition from sensed information. We are currently prototyping several aware environments to aid in the development and study of such sensing and computation in real-world settings.

Ubiquitous Sensing for Smart and Aware Environments

IRFAN A. ESSA, GEORGIA INSTITUTE OF TECHNOLOGY

As computing technology increasingly becomes part of our daily activities, we are required to consider what is the future of computing and how will it change our lives? To address this question, we are interested in developing technologies that would allow for ubiquitous sensing and recognition of daily activities in an environment. Such environments will be aware of the activities performed within it and will be capable of supporting these activities without increasing the cognitive load on the users in the space. Toward this end, we are prototyping different types of smart and aware spaces, each supporting a different part of our daily life and each varying in function and detail. Our most significant effort in this direction is the building of the "Aware Home" at Georgia Tech.

In this article, we outline the research issues we are pursuing toward the building of such smart and aware environments, and especially the Aware Home. We are interested in developing an infrastructure for ubiquitous sensing and recognition of activities in environments. Such sensing will be transparent to everyday activities, while providing the embedded computing infrastructure with an awareness of what is happening in a space. We expect such a ubiquitous sensing infrastructure to support different environments, with varying needs and complexities. These sensors can be mobile or static, configuring their sensing to suit the task at hand while sharing relevant information with other available sensors. This configurable sensor-net will provide high-end sensory data about the status of the environment, its inhabitants, and the ongoing activities in the environment. To achieve this contextual knowledge of the space that is being sensed and to model the environment and the people within it requires methods for both low-level and high-level signal processing and interpretation. We are also building such signal-understanding methods to process the sensory data captured from these sensors and to model and recognize the space and activities in them.

Smart and Aware Environments

A significant aspect of building an aware environment is to explore easily accessible and more pervasive computing services than are available via traditional desktop computing. Computing and sensing in such environments must be reliable, persistent (always remains on), easy to interact with, and transparent (the user does not know it is there and does not need to search for it). The environment must be aware of the users it is interacting with and be capable of unencumbered and intelligent interaction. Building a computing infrastructure that supports all these needs is, therefore, one of our primary goals. Additionally, we are also developing techniques for processing and analyzing the captured sensory streams that will provide the interpretation of the data and make the environment aware. We expect to rely on rich multimodal sensory data to provide high-end awareness of what is happening in the environment.

Following are the significant aspects of our research effort in the areas of ubiquitous sensing and recognition.

Ubiquitous Sensing

The concept of ubiquitous computing seeks to develop a distributed and networked computing infrastructure to support user activities, while remaining transparent to the users. Ubiquitous sensing supports this concept by exploring potential implementations of sensing technology that provides sensing abilities to the ubiquitous computing infrastructure without increasing the burden on the users. A network of sensors that is configured with a network of processing devices can yield a rich multimodal stream of sensory data. Sensory data is, in turn, analyzed to determine the specifics of the environment and provide context for interaction between co-located and distributed users and environments. Such analysis is also useful to determine what is happening in an environment, so that it is supported effectively.

We are specifically interested in high-end multi-modal sensors that provide rich spatio-temporal information about the environment. We are interested in instrumenting the environment and the user with cameras and microphones to extract such a level of sensory data. In the next section, we discuss how content is extracted from the data streams from these high-end sensors. Here we discuss how these types of sensors are used to instrument a space. In addition to the video and audio sensors, we are also working on augmenting the environment and the user with other forms of sensors that will share information with each other.

The features that are important in developing ubiquitous sensing for an aware environment are as follows.

Self-Calibration – The sensors in an aware environment need to be able to calibrate automatically and adapt to the environment as needed. All the sensors in an environment

 1070-9916/00/$10.00 © 2000 IEEE

Ex.1018 / Page 1 of 3
AMAZON.COM, INC.

need to communicate their state and their coverage area to each other and develop a model of the environment. Once the static sensors are calibrated to a given environment, they can then communicate with the mobile sensors and provide them with information that they perceive and allow them to be calibrated as well. This is achieved by establishing protocols for initialization states of the sensors. We are working toward a system in which, after all the sensors are placed in a room, an automatic self-calibration process is initiated. To aid in this self-calibration, we propose to install special devices. For example, a laser light is installed in the spaces with cameras to allow for visual calibration from each camera viewpoint. Audio sensors use special speakers placed in known locations for self-calibration of all the microphones in the space. These self-calibration systems provide us with a geometric model of the space and information that allows mapping of information of each sensor relative to that model of the space.

As the spaces can dynamically change because someone has moved a piece of furniture, we need to also develop systems that allow for dynamic refinement of the model. This is done by observing people moving around in a space and measuring the changes in scene caused by such movements. In the case of measuring this using cameras, we observe the occlusions created as a person moves in a viewpoint of each camera to determine relative depths from that viewpoint [1].

Networking – Combinations of processors and sensors needed to build aware environments require an elaborate networking infrastructure. This infrastructure needs to support both high-bandwidth and low-bandwidth data transmission as determined by context and sensor/processor abilities. Sometimes video and audio needs to be transmitted, while sometimes only extracted labels need to be transmitted. It would be ideal to have these sensors in the space to set up a network dynamically as they are installed. Once a dynamic network is established, then the sensors can transparently capture relevant streams and share with other sensors or processing engines. We are working with researchers in networking to develop an infrastructure to support such computing and networking needs. Such networks, called "ad-hoc" networks, are a big topic of research for mobile and wireless networks [2].

Distributed Computing – In order to install all the above-mentioned ubiquitous computing services in our aware spaces, we need to study and develop a computing infrastructure to support these services. This infrastructure will serve as the brain for the environment where all the information regarding the space is processed. We are developing an abstraction of a virtual processing center for the space. The virtual center will connect various processors distributed throughout the space and allow for transparency in terms of processing and responsiveness. Toward this end, we are studying various parallel computing infrastructures that support real-time multimedia processing [3, 4] and are using the SKIFF and TINI boards in addition to the more traditional computing platforms.

Optical and Audio Sensors – We are interested in using video and audio sensors as high-content sensors. Traditionally these sensors are considered as recording devices. However, these sensors carry a large amount of content that is essential for interpretation of the activities in an environment. If context and the task permit, these sensors can also serve the purpose of recording interactions and allow for face-to-face interactions with spatially separated users. We will integrate these optical and audio sensors with the above mentioned networking and distributed computing architectures to provide large-array, content-rich sensing in an environment.

Mobile and Wearable Sensors – In addition to the static cameras and microphones, we also envision mobile sensors. These could be cameras and microphones either worn by the users or distributed on mobile platforms allowing the cameras to move, pan, tilt, and zoom. In addition, we are also working on bio-sensors that can be worn and which allow measurement of higher-level cues to the state of the person. These sensors will provide a mobile and first-person viewpoint in the environment and allow for focusing on the important events and activities as needed [5].

Embedded Sensors – We are working with computer engineering researchers to develop small embeddable cameras that will be mounted in the ceiling. A large number of such cameras will be distributed in a scene allowing for an elaborate coverage of the space. We are also working on instrumenting the spaces with phased-array microphones that will be embedded in the walls and ceilings. Such microphones will allow for accurate location of the speaker and will provide a higher-quality audio stream for speech recognition. Such sensors will also be aware of their own state that is communicated over the whole sensor network. We are building into these sensors attentive and foveating mechanisms so that they can process relevant information locally and transmit needed information through the network. Such sensors will help keep network traffic and computing needs limited, while assisting in power conservation.

Other Sensors – In addition to the video and audio sensors, we are also studying other types of sensors to augment the user and the environment. These include simple contact sensors to detect which furniture is in use, to a touch-sensitive carpet to track walking people. Recently, new biosensors have been developed to measure biomedical data. The users could wear these sensors and the data could be transmitted to the environment's sensor-net for higher-level content interpretation.

Computational Perception for Recognition of Activities

Using the ubiquitous sensors, the space senses the activities of its occupants and "learns" their routines. Using computer vision and audio processing techniques, the environment locates and identifies its occupants and recognizes their activities [6]. These audio and video sensors will also help establish a natural interface with the space as the space will be able to interpret speech and gestures. These same sensors will allow for face-to-face interactions with other users at other locations.

Following are a few aspects of perceptual processing of video and audio streams that can lead to awareness.

- **Identity**: Using speaker identification and face recognition techniques, the users can be identified. User identity is a strong cue for context awareness. Both speaker ID and face recognition technologies are commercially available at present, and we are using these systems within our frameworks.
- **Location**: Using visual and tactile tracking, users can be located. Where the users are will serve as a strong cue about what they are doing. Phased-array microphones are used to locate a speaking person [7].
- **Activity and behavior**: By processing a video stream, we can track the users as they move. These movements could be over short terms (minutes and seconds) or over long terms (days). Each type of movement provides important cues about what is happening in a scene. Recognition of short-term activities could be used for determining what a user wants at a specific moment, while long-term activities could be learned as routines and then deviations could be measured [8, 9].
- **Pose Tracking**: Recognizing where a user is looking is a significant cue about what a user is doing. This is achieved in

Ex.1018 / Page 2 of 3
AMAZON.COM, INC.

an aware space by tracking the body and face as it moves. We have developed various video-processing techniques to address this problem [10–12]. One of these methods uses a special camera that accentuates pupils by shinning an IR light and tracking its reflections.

- **Expression/gesture recognition**: Visual tracking and recognition of facial expression and hand gesticulation is essential in determining the state of the users. Facial expressions help with interpreting the users' affective state, while gesture recognition will aid in analyzing the communicative intent of the users. We have developed various methods for this form of spatio-temporal analysis of the video streams [13].
- **Audio processing**: Speech recognition is at present a commercially available technology. Though not perfect, these commercially available systems can be used for recognition of spoken words. Using the phased array microphones, we can extract reliable speech signals for such speech recognition systems. In addition to speech recognition, we are also studying non-verbal analysis of the speech signal to extract stylistic and emotive information that is encoded in speech [14].

All of the above technologies are based on our research in the area of signal processing and interpretation. We are prototyping these technologies in our aware and smart spaces and are studying the implications of technologies in an every day (24/7) setting.

Applications

We are developing several smart and aware environments, include a classroom, a living room, a meeting room, and an office [15]. These spaces will serve as "living laboratories" for development and experimentation. We will incorporate into these environments various applications, including:

- Anytime, anywhere videophones (so the occupant does not have to search for a communications device)
- Intelligent surveillance and monitoring (to monitor children, the elderly, and those in need of health care, and to prevent unauthorized access)
- Speech and gestural interfaces (to support natural interfaces)
- Education (to provide connection to school and to facilitate distance learning)
- Entertainment (to support collaborative game-playing)

Our most significant step in this direction is the building of a "Residential Laboratory," under the direction of the Georgia Tech Broadband Institute on the Georgia Tech campus. This laboratory is our attempt at building a home that is a dedicated, large-scale intelligent environment that is aware of the activities of its inhabitants. Initially, this home will serve as a living laboratory while inhabited by our students. However, in the long run we expect this space to become an environment to support the elderly and the sick. An aware home, with embedded abilities to perceive its inhabitant's activities, can provide a richer quality of life to the elderly, allowing them to stay in their home longer at less cost then an assisted living facility. Awareness of daily routines and deviations from some typical behaviors can trigger warnings. Crisis intervention would also be possible as the aware home could detect dangerous circumstances. The aware home can also serve to provide a familial connection between distant family members and provide a ubiquitous support structure [16,17].

Summary

We are building various types of intelligent environments that are aware of their inhabitants. Each aware environment has varying needs and complexity and therefore requires a unified sensing and recognition infrastructure. We are building a large-scale "aware home" as a prototypical intelligent space. This space will provide us with a testbed for implementing our technologies of ubiquitous sensing to make the environment aware of different types of activities. We are pursuing various applications in this aware home. Our most significant effort in this direction is to make this aware home an assisted living facility for the elderly.

Acknowledgments

The author is funded by the National Science Foundation grants #EIA-9806822 and #CAREER-9984847. The Broadband Institute's Residential Laboratory is funded by a grant from the Georgia Research Alliance. Aware Home Research is funded by an Aware Home Research Initiative.

References

[1] G. Brostow and I. Essa, "Motion-based Video Decompositing," *Proc. IEEE Int'l. Conf. Computer Vision 1999*, Corfu, Greece, Mar. 1999.

[2] C.-K. Toh, *et al.*, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Pers. Commun.*, Apr 1999.

[3] U. Ramachandran *et al.*, "Space-Time Memory: A Parallel Programming Abstraction for Interactive Multimedia Applications," *10th ACM SIGPLAN Symp. Principles and Practice of Parallel Programming*, May 1999.

[4] J. M. Rehg *et al.*, "Integrated Task and Data Parallel Support for Dynamic Applications," accepted for publication in the *Journal of Scientific Programming*, to appear.

[5] T. Starner, "Contextual Awareness and Wearable Computing," *PhD Thesis*, Massachusetts Institute of Technology, Media Laboratory, 1999.

[6] I. Essa, "Computers Seeing People," *AI Magazine*, vol. 20, no. 1, Summer 1999, pp. 69–82.

[7] S. Stillman, R. Tanawongsuwan, and I. Essa, "A System for Tracking and Recognizing Multiple People with Multiple Cameras," *Proc. 2nd Int'l. Conf. Audio- Vision-based Person Authentication*, Washington, DC, Apr. 1999.

[8] D. Moore, I. Essa, and M. Hayes, "Exploiting Human Actions and Object Context for Recognition Tasks," *Proc. IEEE Int'l. Conf. Computer Vision 1999 (ICCV'99)*, Corfu, Greece, Mar. 1999

[9] T. Darrell, I. Essa, and A. Pentland, "Task-specific Gesture Modeling using Interpolated Views," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, IEEE Computer Society Press, Dec. 1996, pp. 1236–42.

[10] A. Schödl, A. Haro, and I. Essa, "Head Tracking using a Textured Polygonal Model," *Proc. Perceptual User Interfaces Wksp.*, (held in conjunction with ACM UIST 1998), San Francisco, CA., Nov. 1998

[11] I. Essa and S. Basu. "Modeling, Tracking and Interactive Animation of Facial Expressions and Head Movements using Input from Video," appears, *Proc. Computer Animation 1996 Conf.*, Geneva, Switzerland, June 1996.

[12] A. Haro, M. Flickner, and I. Essa, "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance," *Proc. IEEE Conf. Computer Vision and Pattern Recognition 2000*, Hilton Head, SC, USA, June 2000.

[13] I. Essa and A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, IEEE Computer Society Press, July, 1997.

[14] A. Gardner and I. Essa, "Prosody Analysis for Speaker Affect Determination," *Proc. Perceptual User Interfaces Wksp. (in conjunction with UIST 1997 Conf.) 1997*, Banff, Canada, Oct. 1997.

[15] G. Abowd *et al.*, "Living Laboratories: The Future Computing Environments Group at Georgia Institute of Technology," *Proc. ACM CHI 2000*, (Organizational Overview), The Hague, Netherlands, Apr. 2000.

[16] E. Mynatt, I. Essa, and W. Rogers, "Increasing the Opportunities for Aging in Place," *Proc. ACM Universal Usability 2000 Conf.*, Arlington, VA, Nov. 2000.

[17] Aware Home Research Initiative: http://www.cc.gatech.edu/fce/ahri

Biography

IRFAN ESSA [M] (irfan@cc.gatech.edu) is an assistant professor at the College of Computing, Georgia Institute of Technology. He is affiliated with the Graphics, Visualization, and Usability Center and the Broadband Institute. He is also an active member of the Future Environments Group and has founded the Computational Perception Laboratory. Prior to joining the Georgia Institute of Technology he was a student and a member of the research staff at the MIT Media Laboratory. His research interests are in computer vision, computer graphics, and intelligent, interactive, and aware environments. He is a member of ACM (http://www.cc.gatech.edu/~irfan)

Ex.1018 / Page 3 of 3
AMAZON.COM, INC.