Sentient Future Competition: Ambient Intelligence by Collaborative Eye Tracking

Eiko Yoneki
University of Cambridge
Computer Laboratory
Cambridge CB3 0FD, UK
eiko.yoneki@cl.cam.ac.uk

ABSTRACT

A key aspect for the design of a future sentient computing application is providing ambient intelligence for non-expert users. Automatic, self-organizing and self-managing systems will be essential for such ubiquitous environments, where billions of computers are embedded in everyday life. Eye tracking provides information on both explicit and implicit subconscious social interactions and indicates directions when other communication is inappropriate. Integration of eye tracking and sentient technology will create a powerful paradigm to control and navigate applications. In a public setting, the aggregation of people's observations and knowledge provides a valuable asset. Ten years progress on sensor device hardware and software should realize this paradigm, and numerous applications can be integrated into this technology.

1. INTRODUCTION

We witness a rapid evolution in wireless devices and ubiquitous computing (aka ambient computing), with small computers becoming embedded throughout our environment. Wireless Sensor Networks (WSNs) are composed of multiple, interconnected nodes that are equipped with sensors, wireless communication transceivers, power supply units, and microcontrollers on chips of only a few millimeters square. The sensors are used to gather different types of data such as pictures, motion, sound, temperature, radioactivity, and pressure. In ten years, we imagine that more advanced sensors will appear and those sensors are able to capture 3D images from far distances with high accuracy. Sensors interconnect to establish multi-hop wireless networks streaming captured multimedia data. This heterogeneous collection of devices will interact with sensors embedded in our homes, offices and transportation systems. They will form an intelligent ubiquitous environment. People have to interact with invisible, ambient technology, which must be usable by non-experts. Thanks to Ambient Intelligence, the system will need less input from users and fewer mistakes will occur, because it will take note of the user's history and context and can make 'educated' guesses of the user's needs. Thus, the system will come up with suggestions and questions like 'I think you will need this', or 'Would you like me to adapt for this context?'

The vision of an activated world is action oriented, and, rather than precisely planned, it follows responding human behavior. The social implications are substantial. For example, is the person looking directly at you, to the ground or simply past you, showing interest or bored, aggressiveness or submissiveness [14]? These habits form a powerful method of subtle communication. This new dimension of ubiquitous computing requires more complex communication mechanisms and, most importantly, intelligent processing of information collected from sensors. Network environments for ubiquitous computing will be highly decentralized, distributed over different devices that can be dynamically networked together and will interact in an event-driven mode.

This paper describes future ambient computing, where sentient applications are controlled and coordinated by human eye tracking in many different ways such as forming group communication, sequence of interactions, consensus of the next action, and so forth. Coordination of eye tracking can be between two people, between a person and an object, or among several people. Research requires extensive work with interactive robotics, computer vision, image recognition-understanding-generation, machine learning, data mining, as well as human behavioral studies and cognitive modeling. Ten years will give ample progress in these areas.

Eye (Gaze) tracking is an important human social skill. It is believed that the form of the human eye has evolved in such a way as to allow other humans to infer the direction of other people's view with ease [6]. Especially the high contrast between the sclera (the white part of the eye) and the iris is unusual and cannot be seen in this form in other species [9]. Eye tracking is used in explicit and implicit subconscious social interactions as well as to point and indicate directions when vocal communication is inappropriate. People can immediately recognize if their communication partner is looking at them or past them and infer characteristics of the partner such as interest, fear, or unease from it. Sensitive eye movements can act as a language of emotional states and therefore their detectability in visible light was an important gain in evolution. There are also numerous application fields of eye tracking and they can be grouped into two main tasks, point of interest detection and information transmission via eye movement, although spanning across these fields is common.

Determining attention focus is probably the most common use, although attention is not directly coupled with the line of sight. Point of interest information can be used in a multitude of applications of which marketing, psychophysical experiments, and verification of attention to critical situations such as traffic while driving [18] are the most common uses.

The second main area is to use movement as a direct channel of information, encoding bits as eye movement to the left or to the right. As this method has a relatively low bit rate, it is most often used if other methods of communication are no longer available. This case usually arises from medical conditions when patients have no voluntary control over large parts of their muscles, such as after penalization or with Amyotrophic Lateral Sclerosis (ALS) [2]. ALS for example is a degenerative neural disease causing total loss of muscle control, but sometimes before the terminal stage of locked-in syndrome, eye movement is still possible. For these people, a technical solution for communication via eye tracking can mean at least a little normality in an otherwise difficult situation [5, 11].

Applications combining these two are becoming increasingly popular, as they use point of interest detection as a way of controlling systems. This finer scale resolution allows for a higher bit rate and makes such systems susceptible to more advanced Human Computer Interaction (HCI) devices. For example, the military has used helmet-based eye tracking to act as an additional input and free the pilots' hands to perform other duties. Civilian applications of gaze-based HCIs exist as well, for example for video conferencing or civilian avionics. Furthermore, technical applications of gaze tracking would be necessary in artificial intelligence and social robotics. To mimic human behavior, a robot would have to be capable of reading the emotional language encoded in the movement of the eyes and the direction of gaze.

A significant amount of research literature exists on eye tracking, but most of the earlier approaches have required special hardware and have been to some extent invasive [1, 16, 10]. Those limitations have prevented widespread use of gaze tracking and the technique is currently only used in specialist areas. I envision that the evolution of wireless sensor hardware will overcome many limitations. Sensors will be able to capture 3D images from a distance.

This paper continues as follows: Section 2 describes key aspects of technologies supporting sentient applications using eye tracking. Section 3 describes examples of application scenarios demonstrating the idea and Section 4 contains conclusions.

2. TECHNOLOGY

Research in ubiquitous computing covers diverse research areas, including distributed system design, distributed robotics, wireless communication, signal processing, information theory, P2P networking, embedded systems, data mining, language technology, intelligent agents, and optical technologies. Capturing the eye movement by cameras is possible with current technology, and advanced sensors for this purpose will appear within ten years. The challenge is to capture them from a distant sensor location and establish efficient real-time operation of wireless sensor networks. Outdoors, aerial robots can be used to collect such data. We have the essential technology already and what is needed is to make it scalable, reliable and deployable.

2.1 Eye Tracking

The first eye tracking method was proposed in 1969 for visual targeting of weapon systems by aircraft pilots, allowing them to keep their hands free to control the plane [12]. The first device was built by the US military [13]. These devices exploit that a person's direction of gaze is directly related to the relative positions of the centre of their pupil. The accuracy of these systems depends largely on how precisely the relative positions of the pupil centre and the corneal reflection can be located. To locate the pupil accurately, early systems used a light source at the side of the user and camera with a semi-silvered mirror mounted at 45 degrees to reflect light from the source along the camera axis and into the eye of the operator. In 1989, an eye tracking device [7] was proposed which used a tiny infrared LED mounted in the center of an infra-red sensitive camera, eliminating the need for semi-silvered mirrors. In the more recent vision-based system, the tracking is performed by algorithmically analyzing images coming from video cameras. The use of IR LEDs and a cameras conferred the additional benefit of a lighting independent image. The most recent development is to use a purely vision-based approach. These systems use the natural illumination of the scene and record images with normal video cameras to infer gaze direction. The most popular algorithms have been neural networks to learn the mapping of small images of the eye to the 2D gaze direction. One of the first groups was Baluja et al. [3] at CMU in 1994. The work was continued by Stiefelhage and Waibe [20] and achieved a high accuracy of about 1.4–2.0 degrees using a standard back propagation algorithm on images of size 20 x 10. More recently, other labs have developed similar methods, Xu et al. [23], and application fields are expanding, such as attention tracking during meetings [19].

2.2 Progress of Sensor Hardware

Current research envisages a multitude of inexpensive cameras and projectors embedded in the environment. The cameras infer the geometry and reflective properties of the visible surfaces and the projectors create 3D imagery for a user whose eye positions are tracked in 3D. Current limitations include sensitivity of the camera and narrow fields of camera view, and only few people can be tracked. The cost of cameras and eye contact sensors will fall in ten years, and more sophisticated sensors with eye tracking capability will appear for highly detailed observation of eyes.

2.3 Ubiquitous Computing

The Internet and computer hardware/software made large scale distributed computing possible. The evolution of ubiquitous computing will make a change in a different dimension. Individual systems have to scale down to support ubiquity. Data from sensor networks need to provide pervasive access through a variety of wireless networks. There are inherent resource limitations in the technologies for processing, storage and communications (and power) in this context, and these lead to novel system performance requirements. A new platform needs to cover the range from tiny MEMS to Internet scale P2P systems and must include not only quantitative performance but also quality of service as a critical issue. A total system view will be based on information from a variety of heterogeneous sources and will require knowledge fusion; a reactive system between sensing, decision making and acting will be a common application feature. The architecture of global computing is a fine-grained, open, component-based structure that is highly configurable and self-adaptive. A difficult issue here is that current applications are tied to sensor deployments (see [22] for more details). A new type of open platform is required, where sensed data can be shared among different applications over large-scale environments. Data management over heterogeneous networks, computing, and social environments will be crucial.

Ubiquitous computing infrastructures require software technologies that enable ad hoc assemblies of devices to spontaneously form a coherent group of cooperating components. This is a challenge if the individual components are heterogeneous and have to engage in complex activity sequences to achieve a goal. Today, the interaction between the components is carefully designed by hand. Most sensor network applications are implemented as complex, low level programs that dictate the behavior of individual sensor nodes. WSNs need to organize themselves from components built by different applications. Programming for WSNs raises two main issues; programming abstractions and programming support. The former focuses on providing programmers with abstractions of sensors and sensor data. The latter is providing additional runtime mechanisms that simplify program execution.

Context Awareness: Next-generation conference rooms are designed to include the new rich media presentation hardware. A multiple thin and bright monitor screens, along with interactive whiteboards will be used for smart remote conferencing systems. Smart spaces and interactive furniture design projects have shown systems where tables, floors, walls, chairs and ceilings. Exploiting the capabilities of all these technologies is complex. For example, faced with several monitors, all but a few participants are likely to choose simply replicating the same image on all of them. Even more difficulty is the design challenge: how to choose which capabilities are vital to particular tasks, or for a particular room, or are well suited to a particular culture. The coordination of media-rich evolvement strategies in meetings requires a need to support a way to navigate a meeting among meeting participants with appropriate tools for managing these monitors. Research in areas such as context awareness in pervasive computing, interactive furniture, and mobile devices is increasingly popular. People expect to find the adaptable ease of use that they get from their personal devices in all the technology they encounter.

2.4 Security

Wireless networks are becoming more pervasive and devices more programmable, thereby facilitating malicious and selfish behavior. A ubiquitous application may involve collaboration between ad hoc groups of members. New encounters occur and there are complex issues in knowing what members to trust. Based on predefined trust, recommendations, risk evaluation and experience from past interactions, an entity may derive new trust metrics to use as the basis for authorization policies for access control (see SECURE project [4]). This raises serious concerns about privacy, surveillance and freedom of action. While providing location information can be a one-way system where the location providing tools do not track who are the receivers, once your devices, or other device receive information, the information of your location is potentially available in public. The design of the system will need to consider multi-disciplinary efforts among technologists, social scientists, and societal observers.

3. APPLICATION SCENARIOS

There will be many different uses of the approach described in the previous sections. Several potential scenarios are described below. This list is not exhausted.

3.1 Intruder Detection

By using the correlation between gaze-direction and point of interest, it is possible to find an unobtrusive way of determining attention. This allows gathering the data while people continue their normal task unaware of monitoring, providing data of conscious and subconscious awareness in a natural environment. At an airport, for example, the sensors on the wall sense people's gaze movement. People notice any strange incident, other people's behavior, or objects unconsciously for ~ 0.000001 seconds. This information can be collected to use the detection of any security violation such as suspicious behavior of people or objects left alone.

3.2 Screen Navigation

The multiplicity of recent displays makes it difficult to control what should be shown on specific featured displays. This can be controlled through an eye tracking orchestrated interface. An eye tracking mechanism calculates the user's gaze direction and this can be used to control the volume of the sound of the video streams either by increasing the volume of the stream, or scaling the volumes by temporal distance from the stream. Thus, the less recently selected streams are less chances to be increased volume of the sound. Gaze direction can also be used to identify the stream at which the user is looking and zoom in on it, following a time-out period. The centre of the user's attention can be drawn in higher detail than the rest of the picture by eye tracking in non-uniform rendering of images.

- You are at an airport lounge waiting for the gate assignment of the plane. When the plane delays, the rescheduled time could be in 10 minutes or in 1 hour. The flight schedule screen senses gazes, and when it gets high hits, it releases additional information on the screen or delivers information to the customer's mobile phone. Eyes are used to communicate with the tag on the screen.
- At the main plaza during a nice summer evening, people are watching a football game on a big screen. The screen zooming or selecting the angle are chosen based on the gaze movement of the audience. What is on the screen will thus reflect the interest of the majority of people.
- Interactive TV: sensors are embedded in the TV, which senses movement of gaze for navigation such as zooming or changing channels.

Rennison et al. [17] have worked on gestural navigation of multidimensional information space in 1995 at MIT Media lab. Figure 1 shows 3D Internet browsing navigated by hand gestures. The evolution of sensor networks since 1995 indicates that the next ten years will bring further dramatic progress in technology.

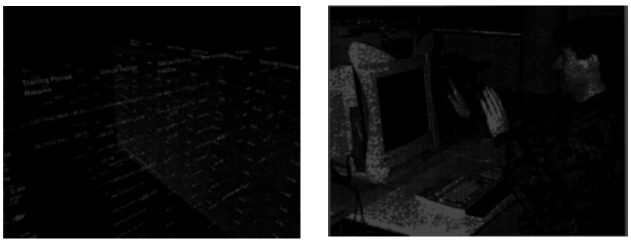

Figure 1: 3D Internet navigation by hand gesture (from [17])

3.3 Effective Video Conference

Video conferencing is a useful tool to conduct meetings without travel. One potential problem is that video conferencing does not necessarily support the coordination of conversational turn taking. In multi-party conversations, when the current speaker finishes his/her talks, it is not clear who will be the next speaker. According to statistics, the previous speaker's eye direction decides who will be the next speaker in many cases, where the next speaker will be the conversation partner in many cases. Thus, their eye contact with each other plays an important role in determining who is to be the next speaker in group conversations [8, 21]. Using low cost eye tracking and measuring eye contacts among the participants could accurately provide more natural video conferencing. There has been an approach to solve this problem by setting multiple cameras, but using 3D image sensors with networking capability will make this even simpler and more effective.

3.4 Medical Applications

There are some medical conditions such as paraplegia, that make other, more traditional methods of communication increasingly difficult, or even impossible. For these patients, eye tracking devices could be a method of communication and improve their quality of life.

Another example is for patients with psychologically complex problems where logging eye tracking data for certain periods could help to find hidden mental problems.

3.5 Unsafe Driving Detection

Eye tracking by sensors detects unawareness of traffic conditions by the driver and offers the potential of improving safety by alerting the driver of tedious but potentially dangerous situations if not sufficient attention is given to the traffic.

3.6 Shop Assistance

Customers' gaze movements are captured to determine which colors of clothes their eyes are on, which types of DVD recorders their eyes are on, and how long their eyes are kept on those objects. The shop could assist the customers by showing more products of interest based on these observations. Furthermore, this information could be used for future improvement of the products or layout of the shop display.

4. CONCLUSTIONS

Automatic, self-organizing and self-managing systems will be essential for supporting ubiquitous environments, where billions of computers are embedded in everyday life. A key aspect to design a future sentient computing application is to provide ambient intelligence for non-expert users. Eye tracking provides explicit and implicit subconscious social interactions and indicates directions when other communication is inappropriate. Integration of eye tracking with sentient technology will create a new paradigm to control and navigate applications. In our society, there is information overload while people are not getting the information they need. They might not even know what exactly they want or need. In a public setting, the aggregation of people's observations and knowledge is a useful and important asset, which can be harvested without the conscious contribution of anybody. Using ambient intelligence, a consensus of knowledge can be obtained and used for good purpose without interfering with people. The applications proposed in this paper aim to construct ambient intelligence by eye tracking, providing ways of effectively coordinating humans, objects, and environments in invisible ways by sentient objects.

Acknowledgement: This research is funded by EPSRC (Engineering and Physical Sciences Research Council) under grant GR/557303. I thank Earl Rennison for showing his work (in 1995) and inspiring the proposed ideas and Jean Bacon (University of Cambridge) for providing helpful suggestions on wireless sensor networks.

5. REFERENCES

[1] A. S. Laboratories. http://www.a-s-l.com.

[2] Amyotrophic lateral sclerosis fact sheet. http://www.ninds.nih.gov/health_and_medical/pubs/als.htm

[3] Baluja, S et al. *Non-intrusive gaze tracking using artificial neural networks,* Technical Report CMU-CS-94-102, Computer Science Department, USA, 1994.

[4] Cahill, V. et al. *Using trust for secure collaboration in uncertain environments,* IEEE Pervasive Computing, 2(3): 52-61, 2003.

[5] Chapman, J.E. *Use of an eye-operated computer system in locked-in syndrome,* CSUN's International Conference: Technology and Persons with Disabilities, 1991.

[6] Emery, N. J. *The eyes have it: the neuroethology, function and evolution of social gaze,* Neuroscience and Biobehavioral Reviews, 2000.

[7] Hutchnson, T.E. *Eye Movement Detector,* U.S. Patent 4,836,670, 1989.

[8] Isaacs, E. et al. *What Video Can and Can't Do for Collaboration,* Proc. ACM Multimedia, 1993.

[9] Kobayashi, H. et al. *Unique morphology of the human eye,* Nature, 1997.

[10] L. Technologies. http://www.eyegaze.com.

[11] Majaranta, P. et al. *Twenty years of eye typing: systems and design issues,* Proc. ETRA, 2002.

[12] Mason, K. A. *Control Apparatus Sensitive to Eye Movement,* U.S. Patent 3,462,604, 1969.

[13] Merchant, J. et al. *A Remote Oculometer Permitting Head Movement,* Aerospace Medical Research Laboratory, WrightPatterson Air Force Base, AMRL-TR-73-69, 1973.

[14] Nummenmaa, T. *The language of the Face. University of Jyvaskyla Studies in Education,* Psychology and Social Research, 1964.

[15] Ramesh Raskar, G. et al. *The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays,* Proc. SIGGRAPH, 1998.

[16] S. Instruments. http://www.smi.de.

[17] Rennison, E. et al. *Issues of Gestural Navigation in Abstract Information Spaces,* ACM CHI Conference, 1995.

[18] Smith, M. et al. *Determining driver visual attention with one camera,* IEEE Transaction on Intelligent Transportation Systems, 2003.

[19] Stiefelhagen, R. *Tracking and Modeling Focus of Attention in Meetings,* PhD thesis, University of Karlsruhe, 2002.

[20] Stiefelhagen, R. et al. *Tracking eyes and monitoring eye gaze,* Proc. Workshop on Perceptual User Interfaces, 1997.

[21] Vertegaal, R. et al. *Effects of Gaze on Multiparty Mediated Communication,* Proc. Of Graphics Interface, 2000.

[22] Yoneki, E. and Bacon, J. *A Survey of Wireless Sensor Network Technologies: Research Trends and Middleware's Role,* Technical Report UCAM-CL-TR646, University of Cambridge, 2005.

[23] Xu, L.-Q. et al. *A novel approach to real-time non-intrusive gaze finding,* British Machine Vision Conf., 1998.