Perception beyond the Here and Now

Albrecht Schmidt,
University of Stuttgart
Marc Langheinrich,
University of Lugano
Kristian Kersting,
Fraunhofer IAIS and University of Bonn

Emerging sensor-equipped computing devices are overcoming longstanding temporal and spatial boundaries to human perception.

A multitude of senses provide us with information about the here and now. What we see, hear, and feel in turn shape how we perceive our surroundings and understand the world.

Our senses are extremely limited, however, and ever since humans began creating and using technology, they have tried to enhance their natural perception in various ways. For example, hilltop castles and watch towers were early means to look farther into the distance, while optical instruments such as telescopes and microscopes enable us to see objects that are too far away or too small to be seen with the naked eye.

Emerging sensor-equipped computing devices are overcoming longstanding temporal and spatial boundaries to human perception, and this could ultimately lead to fundamental new insights about us and our environment.

EXTENDING PERCEPTION

Human perception is narrowly restricted to the five senses; unaided, we can only perceive what is in our vicinity right now—and even then we miss a lot.

Beginning with telegraphy, technology that overcomes spatial limitations has evolved at an increasingly accelerated pace. Having access to audiovisual information from other places has become part of our everyday experience. Live broadcasts on TV or over the Internet are now commonplace. In sporting events such as the World Cup, viewers around the globe get a virtual first-row seat. Soon, high-resolution 3D video feeds will provide distant observers with a highly realistic experience of such events. In addition, while today's live broadcasts are still done by professionals, many mobile phones already offer live streaming services—Qik, for example—that potentially enable average users to become broadcasters.

To overcome our senses' temporal limitations, we must be able to store and preserve what we perceive as well as access it later. This capability has been at the focus of technology development since the early days of photography and audio recordings on wax cylinders. During the past 20 years, many devices have become widely available that capture and store audiovisual data on a large scale with minimal effort and at low cost. A wearable camera that records everything we see, potentially over our entire lifetime, is already technologically and economically feasible (though selective retrieval remains a challenge).

Our senses also have limitations with respect to resolution and coverage. For example, we cannot hear ultrasound, and we do not see individual images but a movie when watching a sequence of images at more than 25 frames per second. There are technologies, however, that can increase our hearing's frequency range, or slow down motion by capturing images at a speed of hundreds of frames per second.

Furthermore, our nonaugmented perception misses many channels of potential information. We do not have a sense that tells us precisely where we are, and we have only very limited means to locate others. Technologies integrated into mobile phones such

Published by the IEEE Computer Society

as GPS and location-sharing applications such as Foursquare and Gowalla already provide these additional "senses."

TOWARD UBIQUITOUS PERCEPTION

While current technologies provide only the rudimentary ingredients for ubiquitous audiovisual perception, emerging systems will be able to capture experiences holistically. By the middle of this century, the boundaries between direct and remote perception will become blurred, yielding fundamental new understandings of our environment.

As the cost and effort of capturing experiences drops, more people will record what they see and hear. At the same time, the value of such recordings will increase with improved content-searching algorithms. No more lost keys, no more not remembering where you parked your car, and no more forgetting when you met someone. Complementing first-person recordings will be audiovisual content captured by the growing number of stationary devices, such as security cameras.

Information recorded by individuals in public areas can be valuable to the community at large. More powerful networks will enable users to ubiquitously share such data in real time. Current trends on Facebook, Twitter, and Flickr suggest that, given the limitations and difficulty of specifying access to information, more people will instead choose to share what they record publicly rather than to an individually specified group.

Researchers have already taken the first steps toward ubiquitous perception. At the 2010 TED (Technology, Entertainment, Design) conference, Photosynth cocreator and Bing Maps architect Blaise Aguera y Arcas demonstrated how to overlay live streaming video from a mobile phone onto Bing's street-view map representation (www.ted.com/talks/blaise_aguera.html).

It is easy to imagine camera-based mobile phones soon streaming viewfinder video with full orientation information and georeferences to an Internet map platform. For well-frequented and oft-captured places like the Eiffel Tower and Niagara Falls, many live streams would be available given sufficient bandwidth for deploying always-streaming apps.

Some may argue that recording and sharing the collective experiences of millions, perhaps billions, of users would generate too much data to handle efficiently. However, massive datasets also provide an opportunity, as they are likely to capture even very rare aspects of the problem at hand (A. Halevy, P. Norvig, and F. Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems*, Mar./Apr. 2009, pp. 8-12).

By the middle of this century, the boundaries between direct and remote perception will become blurred.

In fact, researchers could apply emerging real-time query-matching techniques that identify similarities in massive datasets to the data generated by ubiquitous sensor-equipped computing devices to make predictions about the near future. For example, Yair Weiss, Antonio Torralba, and Robert Fergus have developed a spectral hashing algorithm to process millions of image queries per second on standard computers ("Spectral Hashing," *Advances in Neural Information Processing Systems 21*, NIPS Foundation, 2008, pp. 1753-1760). This technique uses compact binary codes to represent images such that similar images have similar codes and are within a small Hamming distance of a query image's code.

So, just as researchers use machine-learning techniques to analyze historical data, they could also apply them to real-time data to further extend human perception and to better understand human behavior (T.M. Mitchell, "Mining Our Reality," *Science*, 18 Dec. 2009, pp. 1644-1645).

ETHICS AS A DESIGN FACTOR

Technological progress invariably has both positive and negative aspects, often prompting new ethical guidelines and legal mandates. At the end of the 19th century, portable cameras coupled with the large-scale distribution of newspapers led to demands for explicit privacy rights. In the 1960s, when computers evolved from calculators to data processors, many institutions implemented data protection schemes. While extending perception beyond traditional spatial and temporal boundaries provides new opportunities to individuals and businesses, it also threatens many fundamental societal concepts.

Given the many technical choices required to realize ubiquitous perception, designers must carefully consider the ethical implications of their decisions. Current discussions about Web 2.0 developments such as mashups and social networking demonstrate the conflicting values at stake in such applications and how individuals often pragmatically trade off socioeconomic benefits and costs. Achieving systems that offer the functionality users seek without forcing them to surrender their privacy will be at least as challenging as creating the underlying technology.

We will need powerful tools that dynamically control the amount of information we disclose to both friends and strangers, not simple on-off switches. To be meaningful, these tools must be created with substantial input from sociologists, political scientists, and legal and ethical scholars.

EXAMPLE APPLICATION DOMAIN

The long-term vision of ubiquitous perception is not far from reality in many application domains. Mobile Internet connectivity and emerging intervehicle communication standards make such systems especially feasible in a car. For example, by accessing live-stream video from a camera installed in a vehicle ahead, a driver could see what lies around an upcoming corner—though it is hard to imagine visualizing such data in a way that would not be distracting.

Shared data need not be restricted to audiovisual information. A modern car features a plethora of sensors that measure everything from tire pressure to driver interaction to environmental conditions. This information, again georeferenced, could lead to new applications such as a real-time map that indicates road surface quality based on a vehicle's traction and tire vibration.

For these and other scenarios, only a fraction of vehicles would have to share data; not every car would have to be connected to the network. Traffic information systems use a similar approach, and it appears that finding enough volunteers to share data is not difficult.

Magicians, seers, and prophets are often described as having perception beyond the here and now, and if technology can give this ability to everyone, there will likely be a large market for it. Those who have the ability to see what is happening at a place without going there, or to hear what was said some time back at a certain location, will have a competitive advantage. In the long term, it might be as difficult to "opt out" of such technology as it is reject the Internet today. ■

Albrecht Schmidt, *Invisible Computing column editor, is a professor of human-computer interaction at the University of Stuttgart, Germany. Contact him at albrecht@computer.org.*

Marc Langheinrich *is an assistant professor in the Faculty of Informatics at the University of Lugano, Switzerland. Contact him at marc.langheinrich@usi.ch.*

Kritian Kersting *is a senior researcher in the Knowledge Discovery Department of the Fraunhofer Institute for Intelligent Analysis and Information Systems, Sangt Augustin, Germany, and in the Computer Science Department of the University of Bonn, Germany. Contact him at kristian.kersting@iais.fraunhofer.de.*