'Affective' computing and emotion recognition systems: the future of biometric surveillance?

Article · September 2005

DOI: 10.1145/1107622.1107644

Joseph Bullington
Georgia Southern University

‘Affective’ computing and emotion recognition systems: The future of biometric surveillance?

Joseph Bullington
Department of Information Systems
Georgia Southern University
Statesboro, GA 30460
912-681-0753
jbullington@georgiasouthern.edu

ABSTRACT

This paper concerns a subtopic of a larger research program called affective computing, referred to as affect recognition (the terms ‘affect recognition’ and ‘emotion recognition’ will be used interchangeably in this paper). It is proposed that computer systems based on affect recognition could play an important role in the next generation of biometric surveillance systems. In order to introduce affect recognition and its possible applications to the information security community, the present paper will explore the intersection of several groups of technologies, among them: surveillance camera networks, ubiquitous computing, biometrics, face recognition, and affective computing. Three possible scenarios for the deployment of affect recognition will then be briefly discussed. The implementation of these systems will represent the realization of an important goal for the security industry, the automation of real-time prediction of human behavior and intention. Before this goal can be achieved, however, many technical and ethical issues will have to be resolved.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues – *ethics, human safety, privacy.*

General Terms

Security, Human Factors, Legal Aspects.

Keywords

Affective computing, surveillance, biometrics.

1. INTRODUCTION

We are witnessing the emergence of the era of ‘ubiquitous’ computing (Weiser & Brown, 1997). Microprocessors have become so small and inexpensive to manufacture that eventually they will be embedded in everything from household appliances to the clothes we wear. In addition, these systems will be network ready and capable of sharing information with each other over private and public networks. Surveillance technologies have become part of this era of ubiquitous computing; in fact, surveillance cameras are rapidly becoming part of the urban infrastructure in Western society, according to many observers (Farmer & Mann, 2003; Gray, 2003; Shachtman, 2005). Though technically not computers themselves, these cameras produce searchable data capable of being stored on computer systems. Not only are there government-sponsored networks of such devices, both here and abroad, private citizens are purchasing these small cameras at expanding rates (Farmer & Mann, 2003). These private surveillance networks are largely composed of wired and wireless personal cameras and ‘web cams,’ whose images can be made accessible over the Internet, capable of monitoring one’s home or business, and the people in it.

The main justification for the presence of the public surveillance networks (and perhaps the private networks as well) is that they are a deterrent to crime. However, in England, where there are thousands of cameras installed on lampposts and at intersections in cities large and small, the data on their effectiveness have been subject to conflicting interpretations (Bowyer, 2004). In Glasgow, Scotland for example, although fewer crimes were reported where the cameras were installed, there was no evidence to suggest that the cameras were the main factor in reducing crime in those areas, given that there was a general downward trend in reported crimes overall (Ditton, 1999). The present paper will not present arguments for or against the implementation of these camera networks. Their relevance here stems from the fact that they form the infrastructure for future applications of biometric surveillance technologies.

The weakest links in surveillance systems, not surprisingly, are the people monitoring the cameras. One of the main problems, in addition to our inherent information processing limits, is simply boredom, which can cause a viewer to potentially miss details that might otherwise prove important. In addition, it is not difficult to find cases of the abuse of surveillance systems by their operators (ACLU, 2002). Even with these problems, however, the outright elimination of a human operator from the system probably cannot be done. Finding ways of taking the information-processing burden off of the operator would be a good idea,

Information Security Curriculum Development (InfoSecCD) Conference’05, September 23-24, 2005, Kennesaw, GA, USA.
Copyright 2005 ACM 1-#####-###-#/##/####…$5.00.

however, and biometric technologies are one method for achieving this goal.

In the area of authentication, biometrics have proven effective in supplementing passwords as checks of identity for accessing computer systems and other secure resources, as well as to supplement visual checks of a person's visual identity by a human observer. Biometrics, by definition, identify a person based on some feature of their biological makeup. They can be behavioral in nature, such as the way a person writes a signature, or types on a keyboard; or they can be physiological, such as a handprint or fingerprint, the arrangement of blood vessels in the retina, or the patterning of the iris. One of the most popular recent biometric technologies for authentication and identification, and the most clearly related to the subject of the present paper, is face recognition.

The popularity of face recognition as a biometric is based on the fact that data can be gathered passively, without the knowledge or permission of the person who is being observed. The ease with which the data can be gathered, however, has created privacy concerns. While some have suggested that the use of face recognition technology constitutes an invasion of privacy, the courts have consistently ruled that a person does not have reasonable expectation of privacy with regard to physical characteristics (like facial features) that are constantly exposed to the public (Bowyer, 2000).

In brief, face recognition systems used in surveillance are capable of extracting information about human faces from still images or live video feeds and generating a composite image called a *probe.* The *probe* is then compared to previously acquired images that have been stored in a database (referred to as a *gallery* or *watch list*). The system can be set up to trigger 'matches' to *gallery* images with varying levels of sensitivity, thus generating positive as well as false positive matches. A positive match results in an alert being sent to a human operator for further analysis. The possibility of false positives, that is, suggesting that someone has been matched with a terrorist's image, has also fueled privacy concerns. While important in the area of face recognition, the present paper will not include a discussion of the trade-offs between security and privacy that different system sensitivity settings achieve; the interested reader should consult Bowyer, 2004 for more detail. For a literature review of face recognition research, including a discussion of the algorithms used, see Pentland, 2000; and Zhao, Chellappa, Rosenfeld, & Phillips, 2003.

2. PROBLEM STATEMENT

As discussed above, networked, ubiquitous, private and public surveillance networks provide the necessary infrastructure for the application of different techniques for analyzing and acting upon their accompanying image databases. Face recognition is currently perhaps the most popular and widely applied of the new surveillance technologies that can build on this infrastructure. It remains a controversial, though much improved technology. Its importance to the present paper, however, is that it furthers the goal of taking the human observer out of the observation-identification-action loop, and secondly, it provides the foundation for the next generation of biometric devices based on the process of the recognition of human emotion or affect.

So despite privacy concerns, with the growth of surveillance networks and ubiquitous computing, one of the paths for future security technologies lies in advancing and building on automated biometric 'recognition' technologies. The next generation of such devices may include systems that are capable of going beyond the matching of individuals' faces to images in a database, to the interpretation or prediction of a person's underlying motives and/or emotional state and subsequent behavior. The ultimate goal would be a system that can take real-time multi-modal (visual, auditory, physiological, kinesthetic) input from crowded urban environments, and combine the input to come up with accurate predictions of the underlying motives/states of individual agents. Human-computer interface researchers refer to this type of ubiquitous system as a *computer environment*, where the richness of the information available to a human actor – gesture, speech, social context, and affect – is available to the computer system through the interface (Lisetti & Schiano, 2000). In ubiquitous computing environments such as this, the human agents will be unaware that the system is present.

3. LITERATURE REVIEW

In an oft-used (perhaps overused) analogy, the kind of 'affect recognition' surveillance system that is envisioned would be reminiscent of the 'HAL 9000' computer from "2001: A Space Odyssey" ("I'm afraid, Dave…"). Though, as described above, these systems sound like science fiction, there is interest in making them a reality. For example, a recent solicitation from DARPA's Small Business Innovation Research Center calls for the development of a "non invasive emotion recognition system…suitable for deployment in military/operational environments or in environments in which discrete observation of potential enemy threats is desired" (DARPA SB032-038). In addition, SRI lists 'Affective Computing' as one of their 'Next Generation Technologies:' "Affective computing is an important development in computing, because as pervasive or ubiquitous computing becomes mainstream, computers will be far more invisible and natural in their interactions with humans" (SRI Business Intelligence web site, 2005).

Rosalind Picard's group at MIT is the most representative of the efforts to develop 'affective computing' as an area of research in computer science. They define affective computing as "…computing that relates to, arises from, or deliberately influences emotions" (Picard, 1997). Using sensors and wearable devices to record information like facial expression, facial and other muscle activity, and auditory input, the goal of this research group is to create systems that are capable of interpreting emotional responses on the fly, to do things like tailor a learning environment to an individual student or heighten one's experience while playing video games. A second goal is to produce systems that are not only able to recognize emotion, but are capable of responding emotionally as well. To accomplish these goals, their research draws on work in the fields of neuroscience, the psychology of emotion (including emotion recognition), cognitive science, artificial intelligence, and human-computer interface design. The focus of the present paper, 'affect recognition,' represents a sub-field within this broader research area (Picard, 1997; Picard & Klein, 2002).

Building an affect recognition system for the purpose of surveillance and possible intervention has never been attempted. Therefore, what follows represents speculation on the present author's part about where such systems are likely to be deployed, as well as what their capabilities and limitations might be. To put together a system capable of interpreting emotional states in a crowded setting such as an airport would require networked surveillance cameras, along with massive amounts of storage capacity to hold the data retrieved from the cameras (although, as noted above, this infrastructure is becoming more commonplace in urban settings, and less expensive to acquire). Second, software would be required to process the data in real time, in all kinds of environments, to reach judgments about emotional states and behavioral intentions. Finally, human observers would be needed to support the technology and possibly implement the recommendations. Because these emotion recognition systems are likely to be very costly to develop and deploy initially, we are unlikely to see such systems for monitoring the psychological states of people at airports and other public places any time soon. The recognition problem in these types of environments is difficult for current face recognition technology, let alone a technology seeking to predict a person's 'state of mind' in a crowd based in part on facial expression (Pentland, 2000). Until such a time when these general, group-oriented systems are available, researchers may concentrate instead on developing systems that use machine learning algorithms to analyze and learn about the distinctive patterns of affective responding of *individual users* (Pantic & Rothkrantz, 2001; Picard & Klein, 2002).

These individually oriented systems could potentially be deployed in high risk environments where operator error could lead to injuries or fatalities among the general public, as in the transportation industry, or where potential widespread damage to public health could occur, as in the nuclear power industry. Other settings, such as military or other high security environments, could also see this kind of technology deployed. Thus, applications of affect recognition systems will most likely be built for specific problem domains (e.g., nuclear power plant personnel, the military, commercial airline pilots) using learned sets of rules derived from historical observation and analysis of specific users or small groups of users.

In what follows, three possible scenarios for the implementation of emotion recognition technology will be discussed. Because of the psychologically invasive nature of this technology, each scenario is fraught with privacy concerns, which would have to be addressed to the satisfaction of any parties participating in the system. Picard's affective computing research group has anticipated these ethical concerns, and contributed to the ongoing discussion (Picard, 1997; Reynolds & Picard, 2004; see also Bowyer, 2004, and Gray 2003, for a discussion of privacy issues in the use of face recognition technology). The ethical and privacy issues in the use of emotion recognition technologies will be discussed more fully later in the paper.

1) The first (and most likely) scenario for affect recognition systems might include cameras in the engine room of a train (or cockpit of a plane) that could relay footage of a possibly sleepy, intoxicated or distressed driver/pilot to a system designed either to alert security personnel or deliver a warning to the driver/pilot and crew about the nature of the person's state of mind. The system might also include wearable sensors in the driver's clothing to relay physiological data. Were such an automated system in place, the train's (or plane's or ferry's) crew could take over the control of the vehicle in such a situation, or perhaps an autopilot system could be engaged. Such a system will, based on analysis of facial and bodily response data from a surveillance camera (supplemented by perhaps an audio recording and/or physiological data), be capable of: a) deciding that a person is under severe stress, emotional distress, or impaired in some fashion, b) what the person's intentions might be based on this determination, and c) recommending a possible course of intervention. A similar system for use in personal automobiles has been suggested by Lisetti and Schiano (2000).

2) Affect recognition technology could also potentially become part of Group Decision Support Systems. Group decision making, though not specifically related to security concerns, is often susceptible to psychological factors such as 'groupthink.' An affect recognition system might provide a useful source of feedback to group members about the overall emotional state of the group. To preserve anonymity, such information could be synthesized and generated so that individual group members would not be identified. This type of information could also allow members of a decision making group, or a facilitator, to gauge the strength of support or dissent for particular positions as they are being discussed, thus providing a more reliable indicator of the strength of people's feelings about an issue than a more typical survey-based measure of support or dissent.

3) A final, though only remotely likely deployment scenario, could involve the installation of emotion recognition systems in the offices of senior executives at major corporations in the financial industry, or in the offices of high-level government officials. Because fraud and security violations perpetrated by 'insiders' is such a big problem, such a system might serve as a deterrent or provide early warning signs of potential problems. This scenario, more so than the other two mentioned, would probably generate more resistance because of its perceived invasive nature. A major point of contention would be the control of the system. Would the person being observed be able to switch it off during sensitive meetings or private phone conversations for instance? Who would have access to any data produced by the system, and how would that data be stored? Though this application represents a viable use of this technology, in our present social and ethical environment, privacy concerns would outweigh any security benefits we could achieve. In the future, should the security/privacy scale tip the other way, however, we could very likely see this type of system implemented.

4. DISCUSSION AND CONCLUSIONS

In summary, we have a surveillance and storage infrastructure being rapidly put into place to provide support for biometric surveillance systems like face recognition, and building on these technologies, emotion recognition systems. Initially these systems will be capable of monitoring individuals in specific environmental contexts, but eventually they will be suitable for the general population in urban settings. The part of the system that is not yet in place is the software to carry out the data analysis and recommend possible interventions. This kind of task, correctly recognizing and responding to internal psychological states given external cues, goes far beyond facial recognition, which may be extraordinarily difficult. One of the problems with multimodal affect recognition in the general case is that it is context dependent. Our behavior and outward expressions of psychological states are a function of social and cultural context, as well as of individual differences in the way we respond to these contexts. Thus, we may express emotions or signal intentions through our emotional states in different ways. Further, there is continuing controversy in the research literature on emotion over the role of facial and bodily expression in emotional experience (Russell, Bachorowski, Fernandez-Dols, 2003).

Tracking and learning the habits and expressions of a single individual in a limited context may help mitigate these problems. However, just as there is the possibility of false positives in face recognition, the problems will be magnified in emotion recognition systems. In this case, a false positive would mean that the system will be indicating the presence of a particular psychological state, along with possible intentions, that the person is *not* experiencing, or reports not experiencing. The presence of historical data and the fact that systems will have 'learned' about how a person typically responds in a given context will help qualify such judgments, but until the system has enough data about a person to make more accurate judgments, false positives will remain a problem. This is further complicated by the fact that people may try to deliberately deceive the system, by either remaining 'poker-faced' much of the time, or by simply reporting to an investigator that the machine was reporting inaccuracies about the way they were feeling at a particular time. Will human observers, along with the problems that they introduce, be necessary to oversee the monitoring process for a certain amount of time until the system is trained? How will such system 'training' be undertaken, and for how long? Are the self-reports of the person whose state is being observed to be trusted? These are among the questions that will need to be answered.

Assuming that the technical problems can be, for the most part, overcome, there remains the ethical question, are the gains in public safety worth the loss of privacy. The answer will depend on the context in which these systems are implemented. In the post- 9/11 world, people seem more tolerant (or perhaps unaware) of ever-greater levels of surveillance in their everyday lives, particularly if it is felt that the technology can improve security. For example, an informal survey on CNN's web site was conducted concerning a move by the TSA to test the use of low-level x-ray scanners to reveal hidden weapons as part of routine airport surveillance. Such scans would also have the by-product of removing the scanned person's clothes on the image. The results of the survey revealed that 63% of responders (106,666) "…approved of airport security scanning equipment that can see through your clothing" (CNN.com, 2002).

Even with the public's tolerance, however, the capabilities of affect recognition systems would seem to border on what some would consider 'mind-reading.' That is, given a person's outward expressions (along with physiological data in some cases) and context, to accurately predict their underlying psychological state and motivation. Would this constitute an invasion of privacy of a different order of magnitude, with an accompanying public outcry? To explore the ethical implications of computers capable of affect recognition, Reynolds and Picard (2004) conducted a study in which participants read several scenarios presenting details of an 'affect-sensing' computer system that would give them recommendations for music purchases as a result of their emotional reactions to particular songs. Half of the subjects were also provided with a 'contract' describing exactly how the system would behave, and what the data that was collected by the system would be used for. Measures of the participants' comfort levels with the system and feelings that their privacy had been violated revealed similar levels of comfort with the system among members of the 'contract' and 'no-contract' groups, however, the 'contract' group reported significantly lower feelings of privacy violation than the 'no-contract' group. Though these findings suggest that informed consent is important to the acceptance of affect recognition systems, more research is needed into the general public's feelings about the more 'ubiquitous' surveillance systems described in the present paper.

Finally, even if people accept the presence of emotion recognition systems as part of the general surveillance environment, there remain questions about the larger implications of these technologies for our society. Imagine systems that know about your habits and patterns of emotional responding, at home and possibly at work, tied together through the public networks, capable of sharing information about you with each other and with other parties. What are the privacy implications, let alone the security implications of such a network of systems? Where and how would the information be stored? Could you choose to 'opt-out' of such a system? Though emotion recognition systems of this type have not yet been built, it would behoove us as a society to have the discussion about their implications now, rather than after they have been put in place. The present paper is an attempt to begin this discussion by introducing these technologies to the information security community.

5. ACKNOWLEDGMENTS

The author wishes to thank three anonymous reviewers for their helpful comments and suggestions for the improvement of an earlier version of this paper.

6. REFERENCES

[1] ACLU (2002). What's wrong with public video surveillance? http://www.aclu.org/Privacy/Privacy.cfm?ID=13482&c=130

[2] Bowyer, K.W. (2004). Face recognition technology: Security versus privacy. IEEE Technology and Society Magazine, Spring, 9-20.

[3] **CNN.com. (2002). See-through scanner sets off alarms: 'Virtual strip search' testing provokes outcries. http://archives.cnn.com/2002/TRAVEL/NEWS/03/18/rec.airport.xray/index.html**

[4] Ditton, J. (1999). The effect of closed circuit television cameras on recorded crime rates and public concern about crime in Glasgow, The Scottish Office Central Research Unit Main Findings, No. 30, 4 pp.

[5] DARPA SB032-038. Integrated system for emotional state recognition for the enhancement of human performance and detection of criminal intent. http://www.dodsbir.net/solicitation/darpa032.htm

[6] Farmer, D., & Mann, C. (2003). Surveillance nation. Technology Review, 106, (3), 34-43.

[7] Gray, M. (2003). Urban surveillance and panopticism: Will we recognize the facial recognition society. Surveillance and Society, 1, 314-330.

[8] Lisetti, C., & Schiano, D. (2000). Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect. Pragmatics and Cognition, 8, 185-235.

[9] Pantic, M., & Rothkrantz, L. (2001). Affect-sensitive multi-modal monitoring in ubiquitous computing: Advances and challenges. ICEIS, 1, 466-474.

[10] Pentland, A. (2000). Looking at people: Sensing for ubiquitous and wearable computing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 107-119.

[11] Picard, R. (1997). Affective computing. Cambridge, MA: MIT Press.

[12] Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: Theoretical and practical implications. Interacting With Computers, 14, 141-169.

[13] Reynolds, C., & Picard, R. (2004). Affective sensors, privacy, and ethical contracts. Conference on Human Factors in Computing Systems, 1103-1106.

[14] Russell, J.A., Bachorowski, J., & Fernandez-Dols, J. (2003). Facial and vocal expressions of emotion. Annual Review of Psychology, 54, 329-349.

[15] Shachtman, N. (2005). Spycam force. Wired, 13, no. 5, 154-155, 170-171.

[16] Weiser, M., & Brown, J.S. (1997). The coming age of calm technology. In Denning, P.J., and Metcalfe, R.M. (eds.), Beyond Calculation: The Next Fifty Years of Computing. New York, NY: Springer-Verlag.

[17] Zhao, W., Chellapa, A., Rosenfeld, A., & Phillips, P.J. (2003). Face recognition: A literature survey. ACM Computing Surveys, 35, 399-458.