Self-Cam: Feedback from what would be your social partner

Alea Teeters
MIT Media Lab
teeters@media.mit.edu

Rana El Kaliouby
MIT Media Lab
kaliouby@media.mit.edu

Rosalind Picard
MIT Media Lab
picard@media.mit.edu

Summary

We present *Self-Cam*, a novel wearable camera system that analyzes, in real-time, the facial expressions and head gestures of its wearer and infers six underlying affective-cognitive states of mind: agreeing, disagreeing, interested, confused, concentrating, and thinking. A graphical interface summarizes the wearer's states over time. The system allows you to explore what others may see in your face during natural interactions, e.g., "I looked 'agreeing' a lot when I spoke with that researcher I admire." *Self-Cams* will be available for SIGGRAPH attendees to explore this technology's novel point-of-view, and gain insight into what their facial expressions reveal during natural human interaction.

1 Introduction

How does your countenance look during interaction? What do your head movements and facial expressions reveal about your thinking? While we may mask our feelings with some "put on" facial expressions, we also sincerely express many internal states that people readily interpret to mean that we are interested, confused, agreeing, and more. These affective-cognitive signals are fundamental to communication and serve both to regulate conversation and to help others get to know us better. But how do we use these signals naturally as we engage with others? Inspired by Steve Mann's EyeTap – a wearable camera that uses machine vision to provide the wearer with a mediated view of the world [Eyetap 2004] – we developed *Self-Cam* to provide the wearer with a mediated view of the self. *Self-Cam* is novel in providing a new personal point of view and allowing you to gather and analyze your facial signals. *Self-Cam* works in a way that is discrete, recording and revealing information only to you, and maintaining privacy for those around you.

2 Description

Self-Cam consists of a lightweight chest-mounted camera and a belt-mounted handtop together with real-time mental state inference and visual summary output. At 30 fps, the inference system locates and tracks 24 feature points on the face and uses motion, shape and color deformations of these features to identify 20 facial and head movements (e.g., head pitch, lip corner pull) and 11 communicative gestures (e.g., head nod, smile, eyebrow flash). Dynamic Bayesian Networks model these head and facial movements over time, and infer the wearer's hidden affective-cognitive state. The system currently infers a probability for each of the six states of agreeing, disagreeing, interested, confused, concentrating, and thinking. In a six-way classification, it performs as accurately as the top six percent of humans [el Kaliouby (2005)]. The probability distribution of each state is summarized visually, allowing for overlapping states and showing the overall trends while *Self-Cam* was worn (Fig. 1). The wearer can see which states occur most frequently, when these occur, and pair this with any annotations he or she made about activities or social partners during those times.

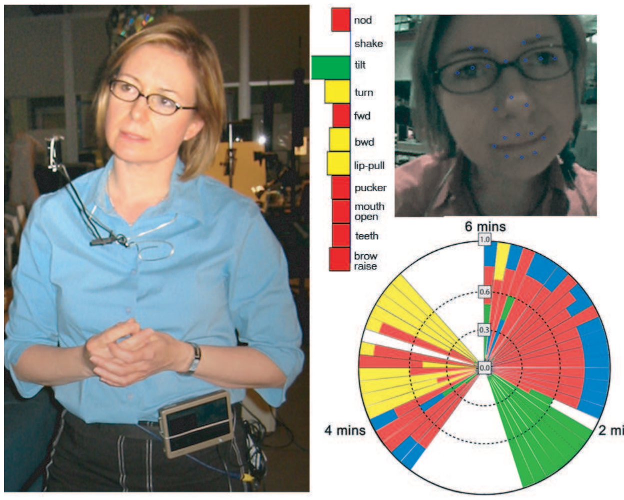

Figure 1: *(clockwise from top-left)* (a) *Self-Cam* pointing at wearer and connected to a belt-mounted handtop; (b) analysis of wearer's head gestures and facial expressions – a head tilt is being recognized ■ (c) radial probability plot of inferred affective-cognitive states of wearer. Between 0-2 mins, the wearer is predominantly concentrating ■ and also mostly agreeing ■; between 2-3 mins, the speaker is thinking-about-something ■. At 4 mins, the wearer is interested ■. (http://affect.media.mit.edu/projectpages/esp/self-cam.mov)

In pilot studies, participants wore *Self-Cam* in one-on-one social interaction and at larger social gatherings. During the latter, people were drawn to converse with the wearers. Sample comments included "It is so tiny – is it a microphone?" People expressed satisfaction that it was only recording the wearer, voiced comfort being around it, and maintained normal social distance. *Self-Cam* was useful both as a real-time **self-awareness** tool, where people thought about how they presented themselves, and as an off-line **self-reflection** tool, where video paired with mental state analysis provoked self-evaluation.

We are currently adapting *Self-Cam* to help persons with autism learn to better understand affective-cognitive expressions through relating their own expressions to their own internal state. The next step in developing the *Self-Cam* system is to implement real-time feedback – a private audio cue alerts you to user-chosen events (e.g. "you look 'disagreeing' again..."). *Self-Cam* builds on people's fascination with their self. Give a person a recording of her voice and she will react strongly ("I sound like that!?"). Take a picture, and she will want to see it. *Self-Cam* lets you explore your appearance during natural interaction, indexed by your affective-cognitive expressions. It gives you a point of view that might help you understand new things about yourself, and how you interact.

References

EYETAP, 2004. http://www.eyetap.org.

EL KALIOUBY, R. (2005). *Mind-Reading Machines: Automated Inference of Complex Mental States*. Phd thesis, University of Cambridge, Computer Laboratory.