A Preliminary System for Recognizing Boredom

Allison M. Jacobs
George Mason University
4400 University Drive
Fairfax, VA 22203
+1 (267) 424-5910
ajacobs8@gmu.edu

Benjamin Fransen
Naval Research Lab
4555 Overlook Avenue
Washington, DC 20375
+1 (202) 863-5872
fransen@aic.nrl.navy.mil

J. Malcolm McCurry
Naval Research Lab
4555 Overlook Avenue
Washington, DC 20375
+1 (202) 404-3923
mccurry@itd.nrl.navy.mil

Frederick W. P. Heckel
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223
+1 (704) 687-7445
fheckel@uncc.edu

Alan R. Wagner
Georgia Institute of Technology
85 Fifth Street NW
Atlanta, GA 30308
+1 (404) 894-9311
alan.wagner@cc.gatech.edu

J. Gregory Trafton
Naval Research Lab
4555 Overlook Avenue
Washington, DC 20375
+1 (202) 767-1748
trafton@itd.nrl.navy.mil

ABSTRACT

A 3D optical flow tracking system was used to track participants as they watched a series of boring videos. The video stream of the participants was rated for boredom events. Ratings and head position data were combined to predict boredom events.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding

General Terms

Experimentation

Keywords

Human-Robot Interaction

1. INTRODUCTION

Head pose and gesture clearly provide important communication channels in social interaction. For example, a listener may lean forward when interested in what you are saying and slouch back if bored. In fact, some researchers have found empirical support for this intuition [1], though the overall effect seems to be modest. There are, of course, many possible predictors for boredom. Some people may close their eyes and start to fall asleep, while others may show subtle facial expressions. Our goal is to explore automated methods of predicting a person's boredom level and then use that information on a robotic platform. Specifically, if a robot can detect a conversational partner's boredom level, it should be able to adjust its own style for better communication.

2. BACKGROUND

Other researchers have attempted to determine affective state from body pose. Mota and Picard studied the pressure on the back and seat of a chair from seated computer users with children aged 8-11. Teachers identified the states of high interest, low interest, and taking a break. These states were matched with the pressure map and the system was trained in identification. They cross-validated the system and achieved a 75% success rate [5].

D'Mello et al. used the same pressure map on a chair and found a match between interest level and body position. An increase in interest was associated with increased seat pressure, or the body position of leaning forward. An increase in boredom was associated with increased back pressure and a rapid change in seat pressure, or the body position of leaning back and shifting weight. The system was trained and able to identify interest and boredom at a rate higher than chance. [2].

El Kaliouby and Robinson developed a system to recognize emotional state from video using facial and head action units. The changes in facial and head movements over time were translated into action units. Using videos from a database of emotion examples, their system was trained and able to recognize 87% of the example emotions [3].

3. APPROACH

A visual tracking system was used to analyze videos of participants experiencing varying levels of boredom to extract head pose and position data. Pose and position were extracted from a video sequence using a 3D optical flow tracking system [4]. The tracking system is capable of tracking faces at distances of up to six feet with orientation changes of +/- 45 degrees. For training, the tracking system requires 3D images of faces (acquired with a depth camera in the current experiment). During run time, 320x 240 or larger video is necessary for accurate tracking.

An adult habituation paradigm was used where participants were shown a series of repetitive videos. While these videos

Conference'04, Month 1–2, 2004, City, State, Country.

were intended to be initially interesting, they quickly became boring. Two judges coded videos of the participants for changes in boredom level. By combining these ratings and the information obtained from our vision system, we attempted to predict boredom and interest.

4. EXPERIMENT

Twenty-three undergraduates from George Mason University participated in our experiment. For this initial analysis of the data, only one participant's data was used. For each participant, 3D face images were recorded using a Swiss Ranger camera integrated with a color camera for use by the tracking system to obtain depth information. Each participant was shown nine video clips in a random order for a total of 22.5 minutes. Participants were seated in a small room in front of a computer monitor and viewed all videos in one session. A Tobii© eyetracker recorded participants' gaze and fixation data and two LogicTech© QuickCam® Pro 9000 webcams on either side of the monitor recorded the participants' faces. After viewing the video clips, participants rated how bored they were in each video on a 7-point Likert scale. During debriefing, participants were informed about the goal to gather video of varying levels of engagement.

Two judges coded the videos. During coding, judges viewed the participants' head and shoulders, including facial expressions, from the left video stream, using the right video stream as a reference if needed. The two judges identified events together but rated each event separately (see [2] for a similar methodology). Judges rated the events as either a change to boredom or a neutral event that did not indicate a change in state.

Participants' faces were manually segmented from the 3D facial image to produce training data used for tracking. Pose and position were recovered automatically using face recognition to initialize the tracking system. Recordings of pose and position were then registered with event ratings for analysis purposes. For this preliminary analysis, only the participants' head velocity was used as a predictor.

4.1 Results

The judges achieved an average of 76.9% agreement after rating the first participant's video. The judges then went back and re-watched the events where there was disagreement and re-rated these events. This improved agreement to an average of 96.7%, resulting in a Cohen's Kappa score of .91.

As shown in Figure 1, the (smoothed) height of this participant's head gradually dropped as the experiment progressed. This can be interpreted as the participant slouching, which other researchers have shown as a sign of boredom [2].

The window around a boredom event was classified as 30 frames prior to the boredom rating and 90 frames after. A support vector machine was trained on the first three quarters of the data and tested on the unused data. The d' of the current system was .894.

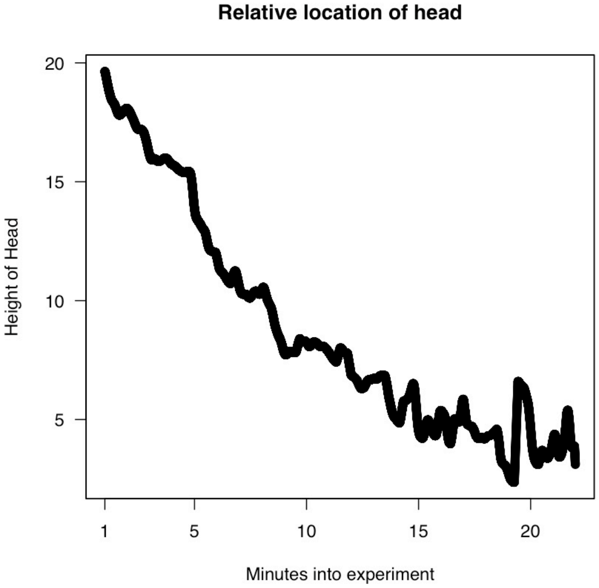

Figure 1. Head height over time

5. CONCLUSION

From our collected dataset, we developed a method for coding events and achieved high agreement between raters. Event ratings and data from our vision system were used to predict boredom and interest. These results remain preliminary because of the single participant examined. Velocity was the only predictor analyzed. In future analyses, we will examine other features of boredom including eye gaze and spatial positioning of the head.

6. ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research under funding document N0001409WX20173 to J. Gregory Trafton. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies of the U. S. Navy.

7. REFERENCES

[1] P. E. Bull. *Posture and Gesture*. Pergamon Press, 1987.

[2] S. D'Mello, P. Chapman, and A. Graesser. Posture as a predictor of learner's affective engagement. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, pages 905–910, Austin, TX, 2007. Cognitive Science Society.

[3] R. el Kaliouby and P. Robinson. Mind reading machines: Automated inference of cognitive mental states from video. In *Proceedings of The 2004 IEEE International Conference on Systems, Man and Cybernetics*, 2004.

[4] Fransen, B., Evan, Harrison, A., & Trafton, J. G. (under review). 3D position and pose tracking. ICRA 2009.

[5] S. Mota and R. W. Picard. Automated posture analysis for detecting learner's interest level. *cvprw*, 05:49, 2003.