Computer Expression Recognition Toolbox

Marian Bartlett, Gwen Littlewort, Tingfan Wu, Javier Movellan
University of California, San Diego
San Diego, CA
{mbartlett, tingfan}@ucsd.edu; {gwen, movellan}@mplab.ucsd.edu

Abstract

We present a live demo of the Computer Expression Recognition Toolbox (CERT) developed at University of California, San Diego. CERT measures facial expressions in real-time, and codes them with respect to expressions of basic emotion, as well as over 20 facial actions from the Facial Action Coding System (Ekman & Friesen, 1978). Head pose (yaw, pitch, and roll) is also detected using an algorithm presented at this conference (Whitehill & Movellan, 2008). A sample output is shown in Figure 1.

This system was employed in some of the earliest experiments in which spontaneous behavior was analyzed with automated expression recognition to extract new information about facial expression that was previously unknown (Bartlett et al., 2008). These experiments addressed automated discrimination of posed from genuine expressions of pain, and automated detection of driver drowsiness. The analysis revealed information about facial behavior during these conditions that were previously unknown, including the coupling of movements. Automated classifiers were able to differentiate real from fake pain significantly better than naïve human subjects, and to detect driver drowsiness above 98% accuracy. Another experiment showed that facial expression was able to predict perceived difficulty of a video lecture and preferred presentation speed (Whitehill et al., 2008). CERT is also being employed in a project to give feedback on facial expression production to children with autism. A prototype game, called SmileMaze, requires the player to produce a smile to enable a character to pass though doors and obtain rewards (Tanaka et al., 2008).

Statistical pattern recognition on large quantities of video data can reveal emergent behavioral patterns that previously would have required hundreds of coding hours by human experts, and would be unattainable by the non-expert. Moreover, automated facial expression analysis will enable investigations into facial expression dynamics that were previously intractable by human coding because of the time required to code intensity changes.

The technical approach to CERT is a texture-based discriminative approach. Such approaches have proven highly robust and fast for face detection and tracking (e.g. Viola & Jones, 2001). Face and detection and detection of internal facial features is first performed on each frame using boosting techniques in a generative framework (Fasel et al.) extending work by Viola and Jones (2001). Enhancements to Viola and Jones include employing Gentleboost instead of AdaBoost, smart feature search, and a novel cascade training procedure, combined in a generative framework. The automatically located faces are then rescaled to 96x96 pixels. The typical distance between the centers of the eyes is roughly 48 pixels. Faces are then aligned using a fast least squares fit on the detected features, and then passed through a bank of Gabor filters 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps). Output magnitudes were then normalized and passed to the facial action classifiers.

Facial action detectors were then developed by training separate support vector machines to detect the presence or absence of each facial action. The training set consisted of over 8000 images from both posed and spontaneous expressions, which were coded for facial actions from the Facial Action Coding System. The datasets were the Cohn-Kanade DFAT-504 dataset (Kanade, Cohn & Tian, 2000); The Ekman, Hager dataset of directed facial actions (Bartlett et al., 1999); A subset of 50 videos from 20 subjects from the MMI database (Pantic et al., 2005); and three spontaneous expression datasets collected by Mark Frank D005, D006, D007 (Bartlett et. al. 2006). Performances on a benchmark datasets (Cohn-Kanade) show state of the art performance for both recognition of basic emotions (98% correct detection for 1 vs all, and 93% correct for 7 alternative forced choice), and for recognizing facial actions from the Facial Action Coding System (mean .93 area under the ROC over 8 facial actions, equivalent to percent correct on a 2-alernative forced choice). More information about the facial expression detection system can be found in Bartlett et al., 2006.
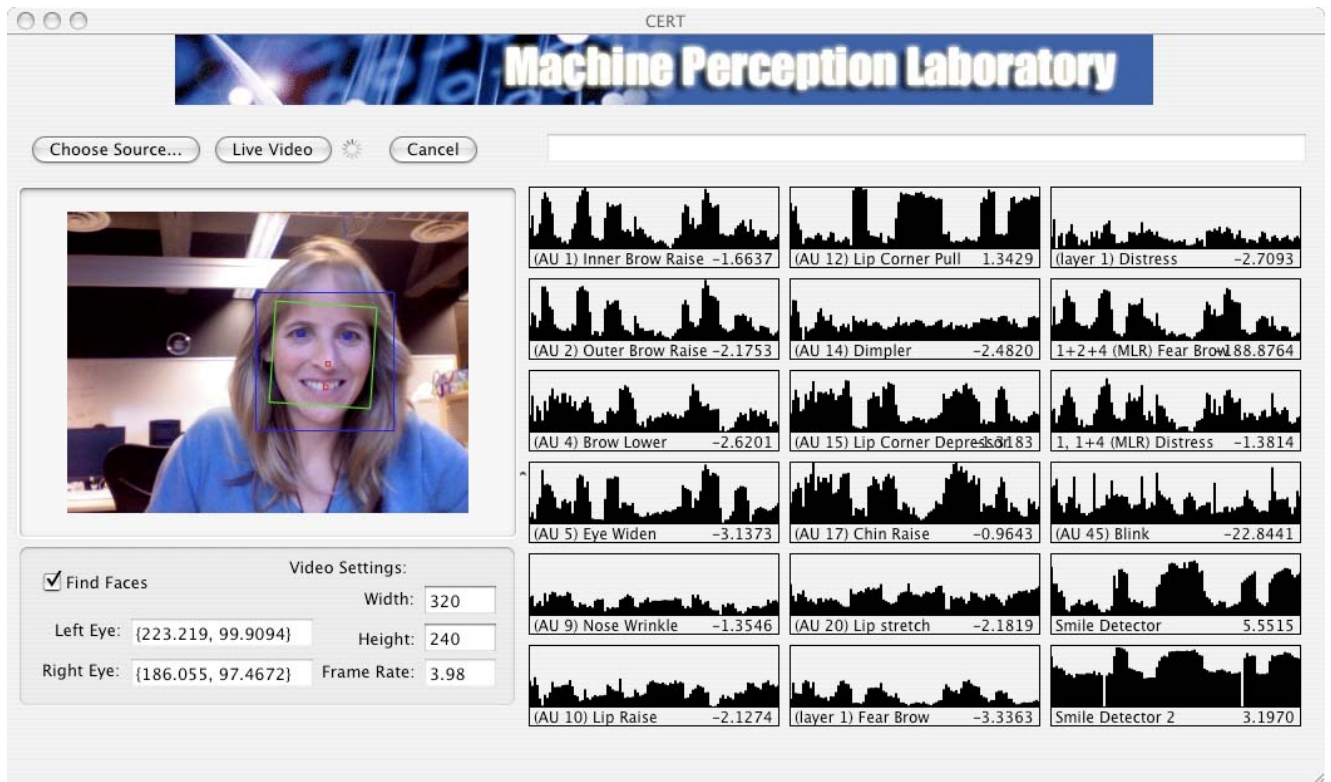
Figure 1. Example of CERT running on live video. Each subplot has time in the horizontal axis and bar height indicates intensity of a particular facial movement.

References

Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology* 36, p. 253-263.

Bartlett M.S., Littlewort G.C., Frank M.G., Lainscsek C., Fasel I., and Movellan J.R., (2006). Automatic recognition of facial actions in spontaneous expressions., Journal of Multimedia., 1(6) p. 22-35.

Bartlett, M. Littlewort, G. Vural, E., Lee, K., Cetin, M., Ercil, A., and Movellan, M. (2008). Data mining spontaneous facial behavior with automatic expression coding. Lecture Notes in Computer Science, Vol 5042: p. 1-21.

Ekman P. and Friesen, W. Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto, CA, 1978.

Fasel I., Fortenberry B., Movellan J.R. "A generative framework for real-time object detection and classification.," Computer Vision and Image Understanding 98, 2005.

Kanade, T., Cohn, J.F. and Tian, Y., "Comprehensive database for facial expression analysis," in Proceedings of the fourth IEEE International conference on automatic face and gesture recognition (FG'00), Grenoble, France, 2000, pp. 46–53.

Pantic, M.F. Valstar, R. Rademaker and L. Maat, "Web- based Database for Facial Expression Analysis", Proc. IEEE Int'l Conf. Multmedia and Expo (ICME'05), Amsterdam, The Netherlands, July 2005.

Tanaka, J., Movellan, J., Bartlett, M., Cockburn, J., & Pierce, M. (2008). SmileMaze: A real-time program training in expression recognition. Demo, 8th Annual Meeting of Vision Science Society, Naples, Florida.

Viola, P. & Jones, M. (2004). Robust real-time face detection. J. Computer Vision, Vol. 57, No. 2, pp. 137-154.

Whitehill, J., Bartlett, M., and Movellan, J. (in press). Automated teacher feedback using facial expression recognition. CVPR workshop, 2008.

Whitehill, J. & Movellan. J. (2008). A discriminative approach to frame-by-frame head pose estimation. 8th IEEE International Conference on Automatic Face and Gesture Recognition.