A Neural Network Model for Human Facial Expression Recognition

Matthew N. Dailey
Computer Science and Engineering Department
University of California, San Diego
9500 Gilman Dr., La Jolla CA 92093-0114 USA
Phone (619) 453-4364
FAX (619) 534-7029
MDAILEY@CS.UCSD.EDU

Garrison W. Cottrell
Computer Science and Engineering Department
University of California, San Diego
9500 Gilman Dr., La Jolla CA 92093-0114 USA
Phone (619) 534-6640
FAX (619) 534-7029
GARY@CS.UCSD.EDU

Preferred presentation format: Oral

Topics: Artificial systems, Data Analysis/Pattern Recognition, Applications, Vision
A neural network model for human facial expression recognition

Matthew N. Dailey
Garrison W. Cottrell
Computer Science and Engineering, UC San Diego

Purpose
We design and evaluate computational models that categorize emotional expressions in facial images for two reasons: 1) the practical advantages of eventually endowing software systems with the ability to recognize their users’ emotions, and 2) to further our understanding of the mechanisms the human brain utilizes for facial expression recognition and, more generally, vision. In this work, we compare the performance of an existing neural network-based expression recognition system with the performance of humans doing the same tasks with the same images. Interestingly, humans perceive expressions “categorically:” when they view images from a sequence produced by morphing, say, an “afraid” face into a “happy” face, subjects see a sharp transition between fear and happiness, rather than a smooth mix of the two or any significant intrusion of other emotions. We compare the performance of our networks with that of the human subjects in a large study, “Megamix,” by Young et al., involving categorization of faces in morph sequences between pairs of “basic” expressions (happy, sad, afraid, surprised, angry, disgust, neutral). We take similarities between the way humans categorize facial expressions and the way our networks categorize the same stimuli as evidence that 1) phenomena like categorical perception of emotional expression are simply artifacts of the process of learning to recognize facial expressions and 2) both the human visual system and our networks use similar mechanisms to recognize emotional expressions.

Methods
The Megamix study used image-quality morphs between pairs of the seven basic emotions portrayed by the model “JJ” in Ekman and Friesen’s Pictures of Facial Affect (PFA) database. In order to evaluate how our networks classify JJ’s facial expressions, we first train networks on the rest of the PFA database then test them on the JJ images and morphs between pairs of JJ images. The morph sequence below is from fear to happiness. In previous work, Padgett and Cottrell have shown that local features extracted from the facial regions around the eyes and mouth perform quite well for expression recognition. Using the same approach, we first find the principal components of a large set of 32x32 pixel patches randomly sampled from the face images. For each face, we then project seven overlapping pixel patches from the eye and mouth regions onto the top 15 principal components of the random block set. This process represents each face with a 105-element input vector. We employ fairly simply neural networks for classification. They contain 10 hidden units and either 6 or 7 outputs depending on whether “Neutral” is interpreted as an expression or not. Since the PFA is small (there are as few as 12 examples of some expressions in the database), we employ early stopping and cross validation to prevent overfitting to the training set.

Results
We have preliminary results utilizing dissolves (linear pixel value interpolation) between facial expression images rather than image-quality morphs. For experiments in which the humans in the Megamix experiments performed a 6-way forced-choice task (happy / sad / disgust / anger / surprise / fear), the networks (with 6 outputs) closely match the humans’ sharp transitions between emotion categories with few intrusions by other emotions. However, when “neutral” is treated as an emotion itself, the networks do not match the human data as well. We are currently running the same experiments with the true morphs to determine whether neutral intrusions are an artifact of the dissolve process.

New or Breakthrough Aspect of Work
This work is novel because it examines similarities and differences between neural network and human facial expression recognition performance. Having a working computational model of the expression recognition system helps us formulate and test hypotheses about the biological system.

Conclusions
The preliminary results of comparisons between our computational models of the expression recognition system and human data suggest that some human phenomena like categorical perception are simply artifacts of the generic learning task. The results also suggest that the expression recognition system may handle “neutral” faces differently than other “true” emotional expressions. Our future work will investigate these points in more detail.