Faces of Pain: Automated Measurement of Spontaneous Facial Expressions of Genuine and Posed Pain

Gwen C. Littlewort
Machine Perception Lab, Institute for Neural Computation
University of California, San Diego
La Jolla, CA 92093-0445 USA
00-858-720-0605
gwen@mplab.ucsd.edu

Marian Stewart Bartlett
Machine Perception Lab, Institute for Neural Computation
University of California, San Diego
La Jolla, CA 92093-0445 USA
00-858-720-0605
marni@salk.edu

Kang Lee
Human Development and Applied Psychology
University of Toronto
Toronto, Ontario M5R 2X2, Canada
00-416-934-4597
kang.lee@utoronto.ca

ABSTRACT
We present initial results from the application of an automated facial expression recognition system to spontaneous facial expressions of pain. In this study, 26 participants were videotaped under three experimental conditions: baseline, posed pain, and real pain. In the real pain condition, subjects experienced cold pressor pain by submerging their arm in ice water. Our goal was to automatically determine which experimental condition was shown in a 60 second clip from a previously unseen subject. We chose a machine learning approach, previously used successfully to categorize basic emotional facial expressions in posed datasets as well as to detect individual facial actions of the Facial Action Coding System (FACS) (Littlewort et al, 2006; Bartlett et al., 2006). For this study, we trained 20 Action Unit (AU) classifiers on over 5000 images selected from a combination of posed and spontaneous facial expressions. The output of the system was a real valued number indicating the distance to the separating hyperplane for each classifier. Applying this system to the pain video data produced a 20 channel output stream, consisting of one real value for each learned AU, for each frame of the video. This data was passed to a second layer of classifiers to predict the difference between baseline and pained faces, and the difference between expressions of real pain and fake pain. Naïve human subjects tested on the same videos were at chance for differentiating faked from real pain, obtaining only 52% accuracy. The automated system was successfully able to differentiate faked from real pain. In an analysis of 26 subjects, the system obtained 72% correct for subject independent discrimination of real versus fake pain on a 2-alternative forced choice. Moreover, the most discriminative facial action in the automated system output was AU 4 (brow lower), which was consistent with findings using human expert FACS codes.

ICMI'07, November 12–15, 2007, Nagoya, Aichi, Japan.

Categories and Subject Descriptors
J.3 and J.4 [**Computer Applications**]: Life and medical sciences: Medical information systems. Social and Behavioral Sciences: Psychology.

General Terms
Measurement, Performance, Design, Experimentation, Human Factors.

Keywords
Facial expression recognition, spontaneous behavior, Facial Action Coding System, FACS, computer vision, machine learning, pain, deception

1. INTRODUCTION

The computer vision field has advanced to the point that we are now able to begin to apply automatic facial expression recognition systems to important research questions in behavioral science. This paper is among the first applications of fully automated facial expression measurement to such research questions. It explores the application of a machine learning system for automatic facial expression measurement to the task of differentiating fake from real expressions of pain.

An important issue in medicine is the ability to distinguish real pain from faked pain, (malingering). Some studies suggest that malingering rates are as high as 10% in chronic pain patients (Fishbain et a., 1999), and much higher in litigation contexts (Schmand et al., 1998). Even more important is to recognize when patients are experiencing genuine pain so that their pain is taken seriously. There is presently no reliable method for physicians to differentiate faked from real pain (Fishbain, 2006). Naïve human subjects are near chance for differentiating real from fake pain from observing facial expression (e.g. Hadjistavropoulos et al., 1996). In the absence of direct training in facial expressions, clinicians are also poor at assessing pain from the face (e.g. Prkachin et al. 2002; Grossman, 1991). However a number of studies using the Facial Action Coding System (FACS) (Ekman & Friesen, 1978) have shown that information exists in the face for differentiating real from posed pain (e.g. Hill and Craig, 2002; Craig et al., 1991; Prkachin 1992).

Recent advances in automated facial expression measurement open up the possibility of automatically differentiating posed from real pain using computer vision systems (e.g. Bartlett et al., 2006; Littlewort et al., 2006; Cohn & Schmidt, 2004; Pantic et al., 2006). This paper explores the application of a system for automatically detecting facial actions to this problem.

1.1 The Facial Action Coding System

The facial action coding system (FACS) (Ekman and Friesen, 1978) is arguably the most widely used method for coding facial expressions in the behavioral sciences. The system describes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscle movements. An example is shown in Figure 1. FACS provides an objective and comprehensive way to analyze expressions into elementary components, analagous to decomposition of speech into phonemes. Because it is comprehensive, FACS has proven useful for discovering facial movements that are indicative of cognitive and affective states. See Ekman and Rosenberg (2005) for a review of facial expression studies using FACS. The primary limitation to the widespread use of FACS is the time required to code. FACS was developed for coding by hand, using human experts. It takes over 100 hours of training to become proficient in FACS, and it takes approximately 2 hours for human experts to code each minute of video. The authors have been developing methods for fully automating the facial action coding system (e.g. Donato et al., 1999; Bartlett et al., 2006). In this paper we apply a computer vision system trained to automatically detect FACS to the problem of differentiating posed from real expressions of pain.

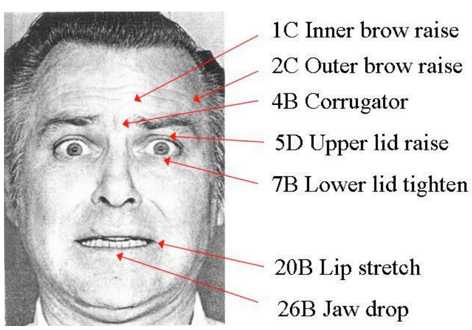

Figure 1. Example facial action decomposition from the facial action coding system. A prototypical expression of fear is decomposed into 7 component movements. Letters indicate intensity. A fear brow (1+2+4) is illustrated here.

In previous studies using manual FACS coding by human experts, at least 12 facial actions showed significant relationships with pain across multiple studies and pain modalities. Of these, the ones specifically associated with cold pressor pain were 4, 6, 7, 9, 10, 12, 25, 26 (Craig & Patrick, 1985; Prkachin, 1992). See Table 1 and Figure 2 for names and examples of these AU's. A previous study compared faked to real pain, but in a different pain modality (lower back pain). This study found that when faking subjects tended to display the following AU's: 4, 6, 7, 10, 12, 25. When faked pain expressions were compared to real pain expressions, the faked pain expressions contained significantly more brow lower (AU 4), cheek raise (AU 6), and lip corner pull (AU 12) (Craig, Hyde & Patrick, 1991). These studies also reported substantial individual differences in the expressions of both real pain and faked pain, making automated detection of faked pain a challenging problem.

1.2 Spontaneous Expressions

The machine learning system presented here was trained on spontaneous facial expressions. The importance of using spontaneous behavior for developing and testing computer vision systems becomes apparent when we examine the neurological substrate for facial expression. There are two distinct neural pathways that mediate facial expressions, each one originating in a different area of the brain. Volitional facial movements originate in the cortical motor strip, whereas spontaneous facial expressions originate in the subcortical areas of the brain (see Rinn, 1984, for a review). These two pathways have different patterns of innervation on the face, with the cortical system tending to give stronger innervation to certain muscles primarily in the lower face, while the subcortical system tends to more strongly innervate certain muscles primarily in the upper face (e.g. Morecraft et al., 2001).

The facial expressions mediated by these two pathways have differences both in which facial muscles are moved and in their dynamics (Ekman, 2001; Ekman & Rosenberg, 2005). Subcortically initiated facial expressions (the spontaneous group) are characterized by synchronized, smooth, symmetrical, consistent, and reflex-like facial muscle movements whereas cortically initiated facial expressions (posed expressions) are subject to volitional real-time control and tend to be less smooth, with more variable dynamics (Rinn, 1984; Frank, Ekman, & Friesen, 1993; Schmidt, Cohn & Tian, 2003; Cohn & Schmidt, 2004). Given the two different neural pathways for facial expressions, it is reasonable to expect to find differences between genuine and posed expressions of pain. Moreover, it is crucial that the computer vision model for detecting genuine pain is based on machine learning of spontaneous examples of real pain expressions.

2. HUMAN SUBJECT METHODS

Video data was collected of 26 human subjects during real pain, faked pain, and baseline conditions. Human subjects were university students consisting of 6 men and 20 women. The pain condition consisted of cold pressor pain induced by immersing the arm in cold water at 3^0 Celsius. For the baseline and faked pain conditions, the water was 20^0 Celsius. Subjects were instructed to immerse their forearm into the water up to the elbow, and hold it there for 60 seconds in each of the three conditions. The order of the conditions was baseline, faked pain, and then real pain. For the faked pain condition, subjects were asked to manipulate their facial expressions so that an "expert would be convinced they were in actual pain." Participants facial expressions were recorded using a digital video camera during each condition.

A second subject group underwent the conditions in the counterbalanced order, with real pain followed by faked pain. This ordering involves immediate motor memory, which is a fundamentally different task. The present paper therefore analyzes only the first subject group. The second group will be analyzed separately in a future paper, and compared to the first group.

After the videos were collected, a set of 170 naïve observers were shown the videos and asked to guess whether each video contained faked or real pain. Subjects were undergraduates with no explicit training in facial expression measurement. They were primarily Psychology majors at U. Toronto. Mean accuracy of naïve human subjects for discriminating fake from real pain in these videos was near chance at 52%. These observers had no specific training in facial expression and were not clinicians. One might suppose that clinicians would be more accurate. However previous studies suggest that clinicians judgments of pain from the face are similarly unreliable (e.g. Grossman, 1991). Facial signals do appear to exist however (Hill & Craig, 2002, Craig et al., 1991; Prkachin 1992), and immediate corrective feedback has been shown to improve observer accuracy (Hill & Craig, 2004).

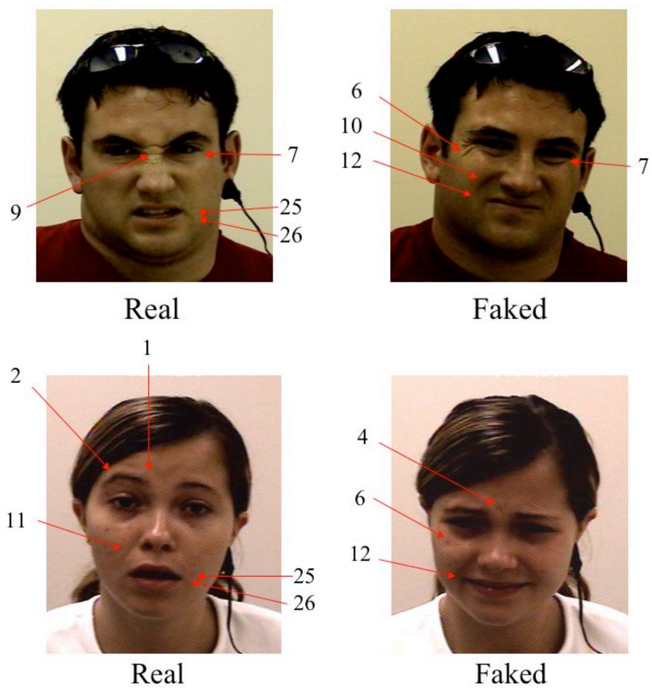

Figure 2. Sample facial behavior and facial action codes from the real and faked pain conditions.

3. COMPUTER VISION APPROACH

3.1 Overview

Here we extend a system for fully automated facial action coding developed previously by the authors (Bartlett et. Al, 2006; Littlewort et al., 2006). It is a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). The system automatically detects frontal faces in the video stream and codes each frame with respect to 20 Action units. In previous work, we conducted empirical investigations of machine learning methods applied to the related problem of classifying expressions of basic emotions. We compared image features (e.g. Donato et al., 1999), classifiers such as AdaBoost, support vector machines, and linear discriminant analysis, as well as feature selection techniques (Littlewort et al., 2006). Best results were obtained by selecting a subset of Gabor filters using AdaBoost and then training Support Vector Machines on the outputs of the filters selected by AdaBoost. An overview of the system is shown in Figure 3.

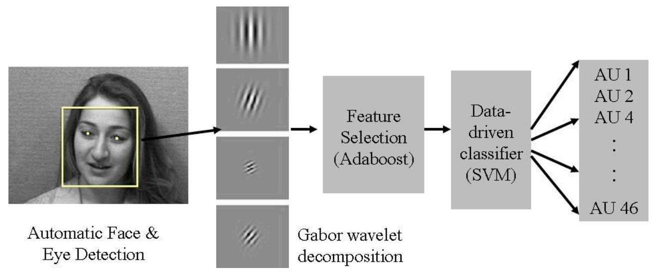

Figure 3. Overview of the automated facial action recognition system.

3.2 Real Time Face and Feature Detection

We employed a real-time face detection system that uses boosting techniques in a generative framework (Fasel et al.) and extends 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. Source code for the face detector is freely available at http://kolmogorov.sourceforge.net. Accuracy on the CMU-MIT dataset, a standard public data set for benchmarking frontal face detection systems (Schneiderman & Kanade, 1998), is 90% detections and 1/million false alarms, which is state-of-the-art accuracy. The CMU test set has unconstrained lighting and background. With controlled lighting and background, such as the facial expression data employed here, detection accuracy is much higher. All faces in the training datasets, for example, were successfully detected. The system presently operates at 24 frames/second on a 3 GHz Pentium IV for 320x240 images. The automatically located faces were rescaled to 96x96 pixels. The typical distance between the centers of the eyes was roughly 48 pixels. Automatic eye detection (Fasel et al., 2005) was employed to align the eyes in each image. The images were 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 passed to the action unit classifiers.

3.3 Automated Facial Action Classification

The training data for the facial action classifiers came from three posed datasets and one dataset of spontaneous expressions. The facial expressions in each dataset were FACS coded by certified FACS coders. The first posed datasets was the Cohn-Kanade DFAT-504 dataset (Kanade, Cohn & Tian, 2000). This dataset consists of 100 university students who were instructed by an experimenter to perform a series of 23 facial displays, including expressions of seven basic emotions. The second posed dataset consisted of directed facial actions from 24 subjects collected by Ekman and Hager. Subjects were instructed by a FACS expert on the display of individual facial actions and action combinations, and they practiced with a mirror. The resulting video was verified for AU content by two certified FACS coders. The third posed dataset consisted of a subset of 50 videos from 20 subjects from the MMI database (Pantic et al., 2005). The spontaneous expression dataset consisted of the FACS-101 dataset collected by Mark Frank (Bartlett et. Al. 2006). 33 subjects underwent an interview about political opinions on which they felt strongly. Two minutes of each subject were FACS coded. The total training set consisted of 5500 examples, 2500 from posed databases and 3000 from the spontaneous set.

Twenty linear Support Vector Machines were trained for each of 20 facial actions. Separate binary classifiers, one for each action, were trained to detect the presence of the action in a one versus all manner. Positive examples consisted of the apex frame for the target AU. Negative examples consisted of all apex frames that did not contain the target AU plus neutral images obtained from the first frame of each sequence. Eighteen of the detectors were for individual action units, and two of the detectors were for specific brow region combinations: fear brow (1+2+4) and distress brow (1 alone or 1+4). All other detectors were trained to detect the presence of the target action regardless of co-occurring actions. A list is shown in Table 1.

Table 1. AU detection performance on posed and spontaneous facial actions. Values are Area under the roc (A') for generalization to novel subjects.

AU	Name	Posed	Spont
1	Inner brow raise	.90	.88
2	Outer brow raise	.94	.81
4	Brow Lower	.98	.73
5	Upper Lid Raise	.98	.80
6	Cheek Raise	.85	.89
7	Lids tight	.96	.77
9	Nose wrinkle	.99	.88
10	Upper lip raise	.98	.78
12	Lip corner pull	.97	.92
14	Dimpler	.90	.77
15	Lip corner Depress	.80	.83
17	Chin Raise	.92	.80
18	Lip Pucker	.87	.70
20	Lip stretch	.98	.60
23	Lip tighten	.89	.63
24	Lip press	.84	.80
25	Lips part	.98	.71
26	Jaw drop	.98	.71
1,1+4	Distress brow	.94	.70
1+2+4	Fear brow	.95	.63
Mean:		.93	.77

The output of the system was a real valued number indicating the distance to the separating hyperplane for each classifier. Previous work showed that the distance to the separating hyperplane (the margin) contained information about action unit intensity (e.g. Bartlett et al., 2006).

In this paper, area under the ROC (A') is used to assess performance rather than overall percent correct, since percent correct can be an unreliable measure of performance, as it depends on the proportion of targets to non-targets, and also on the decision threshold. Similarly, other statistics such as true positive and false positive rates depend on decision threshold, which can complicate comparisons across systems. A' is a measure is derived from signal detection theory and characterizes the discriminative capacity of the signal, independent of decision threshold. The ROC curve is obtained by plotting true positives against false positives as the decision threshold shifts from 0 to 100% detections. The area under the ROC (A') ranges from 0.5 (chance) to 1 (perfect discrimination). A' can also be interpreted in terms of percent correct. A' is equivalent to the theoretical maximum percent correct achievable with the information provided by the system when using a 2-Alternative Forced Choice testing paradigm.

Table 1 shows performance for detecting facial actions in posed and spontaneous facial actions. Generalization to novel subjects was tested using 3-fold cross-validation on the images in the training set. Performance was separated into the posed set, which was 2,500 images, and a spontaneous set, which was 1100 images from the FACS-101 database which includes speech.

4. CLASSIFICATION OF REAL VERSUS FAKE PAIN EXPRESSIONS

Applying this system to the pain video data produced a 20 channel output stream, consisting of one real value for each learned AU, for each frame of the video. This data was further analyzed to predict the difference between baseline and pained faces, and the difference between expressions of real pain and fake pain.

4.1 Characterizing the Difference Between Real and Fake Pain

We first examined which facial action detectors were elevated in real pain compared to the baseline condition. Z-scores for each subject and each AU detector were computed as $Z=(x-\mu)/\sigma$, where (μ,σ) are the mean and variance for the output of frames 100-1100 in the baseline condition (warm water, no faked expressions). The mean difference in Z-score between the baseline and pain conditions was computed across the 26 subjects. Table 2 shows the action detectors with the largest difference in Z-scores. We observed that the actions with the largest Z-scores for genuine pain were Mouth opening and jaw drop (25 and 26), lip corner puller by zygomatic (12), nose wrinkle (9), and to a lesser extent, lip raise (10) and cheek raise (6). These facial actions have been previously associated with cold pressor pain (e.g. Prkachin, 1992; Craig & Patrick 1985).

The Z-score analysis was next repeated for faked versus baseline. We observed that in faked pain there was relatively more facial activity than in real pain. The facial action outputs with the highest z-scores for faked pain relative to baseline were brow lower (4), distress brow (1 or 1+4), inner brow raise (1), mouth open and jaw drop (25 and 26), cheek raise (6), lip raise (10), fear brow (1+2+4), nose wrinkle (9), mouth stretch (20), and lower lid raise (7).

Differences between real and faked pain were examined by computing the difference of the two z-scores. Differences were observed primarily in the outputs of action unit 4 (brow lower), as well as distress brow (1 or 1+4) and inner brow raise (1 in any combination).

Table 2. Z-score differences of the three pain conditions, averaged across subjects. FB: Fear brow 1+2+4. DB: Distress brow (1,1+4).

A. Real Pain vs baseline:

Action Unit	25	12	9	26	10	6
Z-score	1.4	1.4	1.3	1.2	0.9	0.9

B. Faked Pain vs Baseline:

Action Unit	4	DB	1	25	12	6	26	10	FB	9	20	7
Z-score	2.7	2.1	1.7	1.5	1.4	1.4	1.3	1.3	1.2	1.1	1.0	0.9

C. Real Pain vs Faked Pain:

Action Unit	4	DB	1
Z-score difference	1.8	1.7	1.0

Table 3. Individual subject differences between faked and genuine pain. Differences greater than 2 standard deviations are shown. F>P: Number of subjects in which the output for the given AU was greater in faked than genuine pain. P>F: Number of subjects for which the output was greater in genuine than faked pain. FB: Fear brow 1+2+4. DB: Distress brow (1,1+4).

AU	1	2	4	5	6	7	9	10	12	14	15	17	18	20	23	24	25	26	FB	DB
F>P	6	4	9	1	7	4	3	6	5	3	5	5	1	4	3	4	4	4	6	5
P>F	3	3	0	0	4	0	4	4	4	2	3	1	3	1	1	1	2	4	2	0

Individual subject differences between faked and real pain are shown in Table 3. Difference-of-Z-scores between the genuine and faked pain conditions were computed for each subject and each AU. There was considerable variation among subjects in the difference between their faked and real pain expressions. However the most consistent finding is that 9 of the 26 subjects showed significantly more brow lowering activity (AU4) during the faked pain condition, whereas none of the subjects showed significantly more AU4 activity during the real pain condition. Also 7 subjects showed more cheek raise (AU6), and 6 subjects showed more inner brow raise (AU1), and the fear brow combination (1+2+4). The next most common differences were to show more 12, 15, 17, and distress brow (1 alone or 1+4) during faked pain.

Paired t-tests were conducted for each AU to assess whether it was a reliable indicator of genuine versus faked pain in a within-subjects design. Of the 20 actions tested, the difference was statistically significant for three actions. It was highly significant for AU 4 ($p < .001$), and marginally significant for AU 7 and distress brow ($p < .05$).

In order to characterize action unit combinations that relate to the difference between fake and real pain expressions, principal component analysis was conducted on the difference-of-Z-scores. The first eigenvector had the largest loading on distress brow and inner brow raise (AU 1). The second eigenvector had the largest loading on lip corner puller (12) and cheek raise (6) and was *lower* for fake pain expressions. The third eigenvector had the largest loading on brow lower (AU 4). Thus when analyzed singly, the action unit channel with the most information for discriminating fake from real pain was brow lower (AU 4). However when correlations were assessed through PCA, the largest variance was attributed to two combinations, and AU 4 accounted for the third most variance.

Overall, the outputs of the automated system showed similar patterns to previous studies of real and faked pain using manual FACS coding by human experts. Exaggerated activity of the brow lower (AU 4) during faked pain is consistent with previous studies in which the real pain condition was exacerbated lower back pain (Craig et al. 1991, Hill & Craig, 2002). Another study performed a FACS analysis of fake and real pain expressions with cold pressor pain, but with children ages 8-12 (LaRochette et al., 2006). This study observed significant elevation in the following AU's for fake pain relative to baseline: 1 4 6 7 10 12 20 23 25 26. This closely matches the AU's with the highest z-scores in the automated system output of the present study (Table 2B). LaRochette et al. did not measure AU 9 or the brow combinations. When faked pain expressions were compared with real cold pressor pain in children, LaRochette et al found significant differences in AU's 1 4 7 10. Again the findings of the present study using the automated system are similar, as the AU channels with the highest z-scores were 1, 4, and 1+4 (Table 2C), and the t-tests were significant for 4, 1+4 and 7.

4.2 Subject Independent Classification

The above analysis examined which AU outputs contained information about genuine versus faked pain. We next turned to the problem of discriminating genuine from faked pain expressions in a subject-independent manner. If the task were to simply detect the presence of a red-flag set of facial actions, then differentiating fake from real pain expressions would be relatively simple. However, it should be noted that subjects display actions such as AU 4, for example, in both real and fake pain, and the distinction is in the quantity of AU 4. Also, there is inter-subject variation in expressions of both real and fake pain, there may be combinatorial differences in the sets of actions displayed during real and fake pain, and the subjects may cluster. We therefore applied machine learning to the task of discriminating real from faked pain expressions in a subject-independent manner.

A second-layer classifier was trained to discriminate genuine pain from faked pain based on the 20-channel output stream. For this second-layer classification step, we explored SVM's, Adaboost, and linear discriminant analysis. Nonlinear SVM's with radial basis function kernels gave the best performance. System performance for generalization to novel subjects was assessed using leave-one-out cross-validation, in which all the data from one subject was reserved for testing on each trial.

Prior to learning, the system performed an automatic reliability estimate based on the smoothness of the eye positions, and those frames with low reliability were automatically excluded from training and testing the real pain / fake pain classifier. Those frames with abrupt shifts of 2 or more pixels in the returned eye positions were automatically detected and labeled unreliable. This tends to occur during eyeblinks with the current eye detector. However future versions of the eye detector will correct that issue. The reliability filter had a relatively small effect on performance. The analysis of Table 2 was repeated under this criterion, and the Z-scores improved by about 0.1. Note also that the reliability filter on the frames is not to be confused with dropping the difficult trials since a real pain / fake pain decision was always made for each subject.

The 60 second video from each condition was broken up into 6 overlapping segments of 500 frames. For each segment, the following 5 statistics were measured for each of the 20 AU's: median, maximum, range, first to third quartile difference, 90 to 100 percentile difference. Thus the input to the SVM for each segment contained 100 dimensions. Each cross-validation trial contained 300 training samples (25 subjects x 2 conditions x 6 segments).

A nonlinear SVM trained to discriminate posed from real facial expressions of pain obtained an area under the ROC of .72 for generalization to novel subjects. This was significantly higher than performance of naïve human subjects, who obtained a mean accuracy of 52% correct for discriminating faked from real pain on the same set of videos.

4.3 Comparison with Human Expert

In order to assess the validity of the system findings, we obtained FACS codes for a portion of the video from a human FACS expert certified in the Facial Action Coding System. For each subject, the last 500 frames of the fake pain and real pain conditions were FACS coded (about 15 seconds each). It took 60 man hours to collect the human codes, over the course of more than 3 months, since human coders can only code up to 2 hours per day before having negative repercussions in accuracy and coder burn-out.

The sum of the frames containing each action unit were collected for each subject condition, as well as a weighted sum, multiplied by the intensity of the action on a 1-5 scale. To investigate whether any action units successfully differentiated real from faked pain, paired t-tests were computed on each individual action unit. (Tests on specific brow region combinations 1+2+4 and 1,1+4 have not yet been conducted.) The one action unit that significantly differentiated the two conditions was AU 4, brow lower, ($p<.01$) for both the sum and weighted sum measures. This finding is consistent with the analysis of the automated system, which also found action unit 4 most discriminative.

5. DISCUSSION

The field of automatic facial expression analysis has advanced to the point that we can begin to apply it to address research questions in behavioral science. Here we describe a pioneering effort to apply fully automated facial action coding to the problem of differentiating fake from real expressions of pain. While naïve human subjects were only at 52% accuracy for distinguishing fake from real pain, the automated system obtained .72 area under the ROC, which is equivalent to 72% correct on a 2-alternative forced choice. Moreover, the pattern of results in terms of which facial actions may be involved in real pain, fake pain, and differentiating real from fake pain is similar to previous findings in the psychology literature using manual FACS coding.

Here we applied machine learning on a 20-channel output stream of facial action detectors. The machine learning was applied to samples of spontaneous expressions during the subject state in question. Here the state in question was fake versus real pain. The same approach can be applied to learn about other subject states, given a set of spontaneous expression samples. For example, we recently developed a related system to detect driver drowsiness from facial expression (Vural et al., 2007).

While the accuracy of individual facial action detectors is still below that of human experts, automated systems can be applied to large quantities of video data. Statistical pattern recognition on this large quantity of 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. Future work in automatic discrimination of fake and real pain will include investigations into facial expression dynamics.

6. ACKNOWLEDGMENTS

Portions of the research in this paper use the MMI Facial Expression Database collected by M. Pantic & M.F. Valstar. Support for this work was provided by NSF CNS-0454233 and NSF ADVANCE award 0340851. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES

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

[2] Cohn, J.F. & Schmidt, K.L. (2004). The timing of facial motion in posed and spontaneous smiles. *J. Wavelets, Multi-resolution & Information Processing*, Vol. 2, No. 2, pp. 121-132.

[3] Craig KD, Hyde S, Patrick CJ. (1991). Genuine, supressed, and faked facial behaviour during exacerbation of chronic low back pain. Pain 46:161– 172.

[4] Craig KD, Patrick CJ. (1985). Facial expression during induced pain. J Pers Soc Psychol. 48(4):1080-91.

[5] Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P. & Sejnowski, T.J. (1999). Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, pp. 974-989.

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

[7] Ekman, P. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton, New York, USA.

[8] Ekman, P. & Rosenberg, E.L., (Eds.), (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the FACS*, Oxford University Press, Oxford, UK.

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

[10] Fishbain DA, Cutler R, Rosomoff HL, Rosomoff RS. (1999). Chronic pain disability exaggeration/malingering and submaximal effort research. Clin J Pain. 15(4):244-74.

[11] Fishbain DA, Cutler R, Rosomoff HL, Rosomoff RS. (2006). Accuracy of deception judgments. Pers Soc Psychol Rev. 10(3):214-34.

[12] Frank MG, Ekman P, Friesen WV. (1993). Behavioral markers and recognizability of the smile of enjoyment. J Pers Soc Psychol. 64(1):83-93.

[13] Grossman, S., Shielder, V., Swedeen, K., Mucenski, J. (1991). Correlation of patient and caregiver ratings of cancer pain. Journal of Pain and Symptom Management 6(2), p. 53-57.

[14] Hadjistavropoulos HD, Craig KD, Hadjistavropoulos T, Poole GD. (1996). Subjective judgments of deception in pain expression: accuracy and errors. Pain. 65(2-3):251-8.

[15] Hill ML, Craig KD (2002) Detecting deception in pain expressions: the structure of genuine and deceptive facial displays. Pain. 98(1-2):135-44.

[16] 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.

[17] Larochette AC, Chambers CT, Craig KD (2006). Genuine, suppressed and faked facial expressions of pain in children. Pain. 2006 Dec 15;126(1-3):64-71.

[18] Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J. & Movellan, J. (2006). Dynamics of facial expression extracted automatically from video. *J. Image & Vision Computing,* Vol. 24, No. 6, pp. 615-625.

[19] Morecraft RJ, Louie JL, Herrick JL, Stilwell-Morecraft KS. (2001). Cortical innervation of the facial nucleus in the non-human primate: a new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. Brain 124(Pt 1):176-208.

[20] Pantic, M., Pentland, A., Nijholt, A. & Huang, T. (2006). Human Computing and machine understanding of human behaviour: A Survey, *Proc. ACM Int'l Conf. Multimodal Interfaces*, pp. 239-248.

[21] 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.

[21] Prkachin KM. (1992). The consistency of facial expressions of pain: a comparison across modalities. Pain. 51(3):297-306.

[23] Prkachin KM, Schultz I, Berkowitz J, Hughes E, Hunt D. Assessing pain behaviour of low-back pain patients in real time: concurrent validity and examiner sensitivity. Behav Res Ther. 40(5):595-607.

[24] Rinn WE. The neuropsyhology of facial expression: a review of the neurological and psychological mechanisms for producing facial expression. Psychol Bull 95:52–77.

[25] Schmand B, Lindeboom J, Schagen S, Heijt R, Koene T, Hamburger HL. Cognitive complaints in patients after whiplash injury: the impact of malingering. J Neurol Neurosurg Psychiatry. 64(3):339-43.

[26] Schmidt KL, Cohn JF, Tian Y. (2003). Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. Biol Psychol. 65(1):49-66.

[27] Schneiderman, H. and Kanade, T. (1998). Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 45-51.

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

[29] Vural, E., Ercil, A., Littlewort, G.C., Bartlett, M.S., and Movellan, J.R. (2007). Machine learning systems for detecting driver drowsiness. Proceedings of the Biennial Conference on Digital Signal Processing for in-Vehicle and Mobile Systems.