Affective Pattern Classification

Elias Vyzas and Rosalind W. Picard

The Media Laboratory
Massachusetts Institute of Technology
20 Ames St., Cambridge, MA 02139

Abstract

We develop a method for recognizing the emotional state of a person who is deliberately expressing one of eight emotions. Four physiological signals were measured and six features of each of these signals were extracted. We investigated three methods for the recognition: (1) Sequential floating forward search (SFFS) feature selection with K-nearest neighbors classification, (2) Fisher projection on structured subsets of features with MAP classification, and (3) A hybrid SFFS-Fisher projection method. Each method was evaluated on the full set of eight emotions as well as on several subsets. The SFFS attained the highest rate for a trio of emotions, 2.7 times that of random guessing, while the Fisher projection with structured subsets attained the best performance on the full set of emotions, 3.9 times random. The emotion recognition problem is demonstrated to be a difficult one, with day-to-day variations within the same class often exceeding between-class variations on the same day. We present a way to take account of the day information, resulting in an improvement to the Fisher-based methods. The findings in this paper demonstrate that there is significant information in physiological signals for classifying the affective state of a person who is deliberately expressing a small set of emotions.

Introduction

This paper addresses emotion recognition, specifically the recognition by computer of affective information expressed by people. This is part of a larger effort in "affective computing," computing that relates to, arises from, or deliberately influences emotions (Picard 1997). Affective computing has numerous applications and motivations, one of which is giving computers the skills involved in so-called "emotional intelligence," such as the ability to recognize a person's emotions. Such skills have been argued to be more important in general than mathematical and verbal abilities in determining a person's success in life (Goleman 1995). Recognition of emotional information is a key step toward giving computers the ability to interact more naturally and intelligently with people.

The research described here focuses on recognition of emotional states during deliberate emotional expression by an actress. The actress, trained in guided imagery, used the Clynes method of sentic cycles to assist in eliciting the emotional states (Clynes 1977). For example, to elicit the state of "Neutral," (no emotion) she focused on a blank piece of paper or a typewriter. To elicit the state of "Anger" she focused on people who aroused anger in her. This process was adapted for the eight states: Neutral (no emotion) (N), Anger (A), Hate (H), Grief (G), Platonic Love (P), Romantic Love (L), Joy (J), and Reverence (R).

The specific states one would want a computer to recognize will depend on the particular application. The eight emotions used in this research are intended to be representative of a broad range, which can be described in terms of the "arousal-valence" space commonly used by psychologists (Lang 1995). The arousal axis ranges from calm and peaceful to active and excited, while the valence axis ranges from negative to positive. For example, anger was considered high in arousal, while reverence was considered low. Love was considered positive, while hate was considered negative.

There has been prior work on emotional expression recognition from speech and from image and video; this work, like ours, has focused on deliberately expressed emotions. The problem is a hard one when you look at the few benchmarks which exist. In general, people can recognize affect in neutral-content speech with about 60% accuracy, choosing from among about six different affective states (Scherer 1981). Computer algorithms can match this accuracy but only under more restrictive assumptions, such as when the sentence content is known. Facial expression recognition is easier, and the rates computers obtain are higher: from 80-98% accuracy when recognizing 5-7 classes of emotional expression on groups of 8-32 people (Yacoob & Davis 1996; Essa & Pentland 1997). Facial expressions are easily controlled by people, and easily exaggerated, facilitating their discrimination.

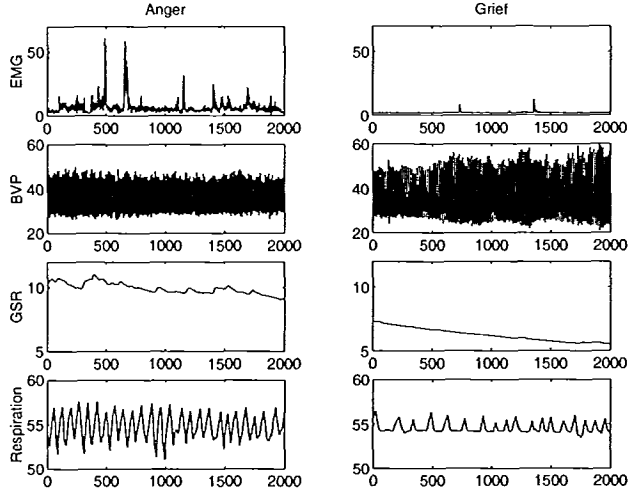Emotion recognition can also involve other modali-

Figure 1: Examples of four physiological signals measured from an actress while she intentionally expressed anger (left) and grief (right). From top to bottom: electromyogram (microvolts), blood volume pressure (percent reflectance), galvanic skin conductivity (microSiemens), and respiration (percent maximum expansion). The signals were sampled at 20 samples a second. Each box shows 100 seconds of response. The segments shown here are visibly different for the two emotions, which was not true in general.

ties such as analyzing posture, gait, gesture, and a variety of physiological features in addition to the ones described in this paper. Additionally, emotion recognition can involve prediction based on cognitive reasoning about a situation, such as "That goal is important to her, and he just prevented her from obtaining it; therefore, she might be angry at him." The best emotion recognition is likely to come from pattern recognition and reasoning applied to a combination of all of these modalities, including both low-level signal recognition, and higher-level reasoning about the situation (Picard 1997).

For the research described here, four physiological signals of an actress were recorded during deliberate emotional expression. The signals measured were electromyogram (EMG) from the jaws, representing muscular tension or jaw clenching, blood volume pressure (BVP) and skin conductivity (GSR) from the fingers, and respiration from chest expansion. Data was gathered for each of the eight emotional states for approximately 3 minutes each. This process was repeated for several weeks. The four physiological waveforms were each sampled at 20 samples a second. The experiments below use 2000 samples per signal, for each of the eight emotions, gathered over 20 days (Fig. 1). Hence there are a total of 32 signals a day, and 80 signals per emotion.

Very little work has been done on pattern recognition of emotion from physiological signals, and there is controversy among emotion theorists whether or not emotions do occur with unique patterns of physiological signals. Some psychologists have argued that emotions might be recognizable from physiological signals given suitable pattern recognition techniques (Cacioppo & Tassinary 1990), but nobody has yet to demonstrate which physiological signals, or which features of those signals, or which methods of classification, give reliable indications of an underlying emotion, if any. This paper suggests signals, features, and pattern recognition techniques for solving this problem, and presents results that emotions can be recognized from physiological signals at significantly higher than chance probabilities.

Choice of Features

A very important part in recognizing emotional states, as with any pattern recognition procedure, is to determine which features are most relevant and helpful. This helps both in reducing the amount of data stored and in improving the performance of the recognizer. recognition problem.

Let the four raw signals, the digitized EMG, BVP, GSR, and Respiration waveforms, be designated by $(S^i), i = 1, 2, 3, 4$. Each signal is gathered for 8 different emotions each session, for 20 sessions. Let S_n^i represent the value of the nth sample of the ith raw signal, where $n = 1...N$ and $N = 2000$ samples. Let $\tilde{S}_n^i$ refer to the normalized signal (zero mean, unit variance), formed as:

$$\tilde{S}_n^i = \frac{S_n^i - \mu^i}{\sigma^i} \quad i = 1, ..., 4$$

where μ^i and σ^i are the means and standard deviations explained below. We extract 6 types of features for each emotion, each session:

1. the means of the raw signals (4 values)

$$\mu^i = \frac{1}{N} \sum_{n=1}^{N} S_n^i \quad i = 1, ..., 4 \qquad (1)$$

2. the standard deviations of the raw signals (4 values)

$$\sigma^i = \left(\frac{1}{N-1} \sum_{n=1}^{N} (S_n^i - \mu^i)^2 \right)^{1/2} \quad i = 1, ..., 4 \quad (2)$$

3. the means of the absolute values of the first differences of the raw signals (4 values)

$$\delta_1^i = \frac{1}{N-1} \sum_{n=1}^{N-1} |S_{n+1}^i - S_n^i| \quad i = 1, ..., 4 \qquad (3)$$

4. the means of the absolute values of the first differences of the normalized signals (4 values)

$$\tilde{\delta}_1^i = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{S}_{n+1}^i - \tilde{S}_n^i| = \frac{\delta_1^i}{\sigma^i} \quad i = 1, ..., 4.$$

(4)

5. the means of the absolute values of the second differences of the raw signals (4 values)

$$\delta_2^i = \frac{1}{N-2} \sum_{n=1}^{N-2} |S_{n+2}^i - S_n^i| \quad i = 1, ..., 4$$

(5)

6. the means of the absolute values of the second differences of the normalized signals (4 values)

$$\tilde{\delta}_2^i = \frac{1}{N-2} \sum_{n=1}^{N-2} |\tilde{S}_{n+2}^i - \tilde{S}_n^i| = \frac{\delta_2^i}{\sigma^i} \quad i = 1, ..., 4$$

(6)

Therefore, each emotion is characterized by 24 features, corresponding to a point in a 24-dimensional space. The classification can take place in this space, in an arbitrary subspace of it, or in a space otherwise constructed from these features. The total number of data in all cases is 20 points per class for each of the 8 classes, 160 data points in total.

Note that the features are not independent; in particular, two of the features are nonlinear combinations of the other features. We expect that dimensionality reduction techniques will be useful in selecting which of the proposed features contain the most significant discriminatory information.

Dimensionality reduction

There is no guarantee that the features chosen above are all appropriate for emotion recognition. Nor is it guaranteed that emotion recognition from physiological signals is possible. Furthermore, a very limited number of data points—20 per class—is available. Hence, we expect that the classification error may be high, and may further increase when too many features are used. Therefore, reductions in the dimensionality of the feature space need to be explored, among with other options. In this paper we focus on three methods for reducing the dimensionality, and evaluate the performance of these methods.

Sequential Floating Forward Search

The Sequential Floating Forward Search (SFFS) method (Pudil, Novovicova, & Kittler 1994) is chosen due to its consistent success in previous evaluations of feature selection algorithms, where it has recently been shown to outperform methods such as Sequential Forward and Sequential Backward Search (SFS, SBS), Generalized SFS and SBS, and Max-Min, (Jain & Zongker 1997) in several benchmarks. Of course the performance of SFFS is data dependent and the data here is new and difficult; hence, the SFFS may not be the best method to use. Nonetheless, because of its well documented success in other pattern recognition problems, it will help establish a benchmark for the new field of emotion recognition and assess the quality of other methods.

The SFFS method takes as input the values of n features. It then does a non-exhaustive search on the feature space by iteratively adding and subtracting features. It outputs one subset of m features for each m, $2 \leq m \leq n$, together with its classification rate. The algorithm is described in detail in (Pudil, Novovicova, & Kittler 1994).

Fisher Projection

Fisher projection is a well-known method of reducing the dimensionality of the problem in hand, which involves less computation than SFFS. The goal is to find a projection of the data to a space of fewer dimensions than the original where the classes are well separated.

Due to the nature of the Fisher projection method, the data can only be projected down to $c-1$ (or fewer if one wants) dimensions, assuming that originally there are more than $c-1$ dimensions and c is the number of classes.

It is important to keep in mind that if the amount of training data is inadequate, or the quality of some of the features is questionable, then some of the dimensions of the Fisher projection may be a result of noise rather than a result of differences among the classes. In this case, Fisher might find a meaningless projection which reduces the error in the training data but performs poorly in the testing data. For this reason, projections down to fewer than $c-1$ dimensions are also evaluated in the paper.

Furthermore, since 24 features is high for the amount of training data here, and since the nature of the data is so little understood that these features may contain superfluous measures, we decided to try an additional approach: applying the Fisher projection not only to the original 24 features, but also to several "structured subsets" of the 24 features, which are described further below. Although in theory the Fisher method finds its own most relevant projections, the evaluation conducted below indicates that better results are obtained with the structured subsets approach.

Note that if the number of features n is smaller than the number of classes c, the Fisher projection is meaningful only up to at most $n-1$ dimensions. Therefore in general the number of Fisher projection dimensions d is $1 \leq d \leq \min(n, c) - 1$. For example, when 24 features are used on all 8 classes, all $d = [1, 7]$ are tried. When 4 features are used on 8 classes, all $d = [1, 3]$ are tried.

178

Hybrid SFFS with Fisher Projection (SFFS-FP)

As mentioned above, the SFFS algorithm proposes one subset of m features for each m, $2 \leq m \leq n$. Therefore, instead of feeding the Fisher algorithm with all 24 features or with structured subsets, we can use the subsets that the SFFS algorithm proposes as our input to the Fisher Algorithm. Note that the SFFS method is used here as a simple preprocessor for reducing the number of features fed into the Fisher algorithm, and not as a classification method. We call this hybrid method SFFS-FP.

Evaluation

We now describe how we obtained the results shown in Table 1. A discussion of these results follows below.

Methodology

The Maximum a Posteriori (MAP) classification is used for all Fisher Projection methods. The leave-one-out method is chosen for cross validation because of the small amount of data available. More specifically, here is the algorithm that is applied to every data point:

1. The data point to be classified (the testing set only includes one point) is excluded from the data set. The remaining data set will be used as the training set.

2. In the case where a Fisher projection is to be used, the projection matrix is calculated from only the training set. Then both the training and testing set are projected down to the d dimensions found by Fisher.

3. Given the feature space, original or reduced, the data in that space is assumed to be Gaussian. The respective means and covariance matrices of the classes are estimated from the training data.

4. The posterior probability of the testing set is calculated: the probability the test point belongs to a specific class, depending on the specific probability distribution of the class and the priors.

5. The data point is then classified as coming from the class with the highest posterior probability.

The above algorithm is first applied on the original 24 features (**Fisher-24**). Because this feature set was expected to contain a lot of redundancy and noise, we also chose to apply the above algorithm on various "structured subsets" of 4, 6 and 18 features defined as follows:

Fisher-4 All combinations of 4 features are tried, with the constraint that each feature is from a different signal (EMG, BVP, GSR, Respiration). This gives a total of $6^4 = 1296$ combinations, which substantially reduces the (24 choose 4)=10626 that would result if all combinations were to be tried. The results of this

evaluation may give us an indication of which type of feature is most useful for each physiological signal.

Fisher-6 All combinations of 6 features are tried, with the constraint that each feature has to be of a different type: (1)-(6). This gives a total of $4^6 = 4096$ combinations instead of (24 choose 6)=134596 if all combinations were to be tried. The results of this evaluation may give us an indication which physiological signal is most useful for each type of feature.

Fisher-18 All possible combinations of 18 features are tried, with the constraint that exactly 3 features are chosen from each of the types (1)-(6). That again gives a total of $4^6 = 4096$ combinations, instead of (24 choose 18)=134596 if all combinations were to be tried. The results of this evaluation may give us an indication which physiological signal is least useful for each feature.

The SFFS software we used included its own evaluation method, K-nearest neighbors, in choosing which features were best. For the SFFS-FP method, the procedure below was followed: The SFFS algorithm outputs one set of m features for each $2 \leq m \leq n$, and for each $1 \leq k \leq 20$. All possible Fisher projections are then calculated for each such set.

Another case, not shown in Table 1, was investigated. Instead of using a Fisher projection, we tried all possible **2-feature subsets**, and evaluated their class according to the maximum a posteriori probability, using cross-validation. The best classification in this case was consistently obtained when using the mean of the EMG signal (feature μ^1 above) and the mean of the absolute value of the first difference of the normalized Respiration signal (feature $\bar{\delta}_1^4$ above) as the 2 features. The only result almost comparable to other methods was obtained when discriminating among Anger, Joy and Reverence where a linear classifier scores 71.66% (43/60). When trying to discriminate among more than 3 emotions, the results were not significantly better than random guessing, while the algorithm consumed too much time in an exhaustive search.

Attempting to discriminate among 8 different emotional states is unnecessary for many applications, where 3 or 4 emotions may be all that is needed. We therefore evaluated the three methods here not only for the full set of eight emotion classes, but also for sets of three, four, and five classes that seemed the most promising in preliminary tests.

Results

The results of all the emotion subsets and classification algorithms are shown in Table 1. All methods performed significantly better than random guessing, indicating that there is emotional discriminatory information in the physiological signals.

When Fisher was applied to structured subsets of features, the results were always better than when Fisher was applied to the original 24 features.

179

3 emotions In runs using the Fisher-24 algorithm, the two best 3-emotion subsets turned out to be the *Anger-Grief-Reverence (AGR)* and the *Anger-Joy-Reverence (AJR)*. All the other methods are applied on just these two triplets for comparison.

4 emotions In order to avoid trying all the possible quadruplets with all the possible methods, we use the following arguments for our choices:

Anger-Grief-Joy-Reverence (AGJR): These are the emotions included in the best-classified triplets. Furthermore, the features used in obtaining the best results above were not the same for the two cases. Therefore a combination of these features may be discriminative for all 5 emotions. Finally, these emotions can be seen as placed in the four corners of a valence-arousal plot, a common taxonomy used by psychologists in categorizing the space of emotions:
Anger: High Arousal, Negative Valence
Grief: Low Arousal, Negative Valence
Joy: High Arousal, Positive Valence
Reverence: Low Arousal, Positive Valence

Neutral-Anger-Grief-Reverence (NAGR) In results from the 8-emotion classification using the Fisher-24 algorithm, the resulting confusion matrix shows that Neutral, Anger, Grief, and Reverence are the four emotions best classified and least confused with each other.

5-emotions The 5-emotion subset examined is the one including the emotions in the 2 quadruplets chosen above, namely the *Neutral-Anger-Grief-Joy-Reverence (NAGJR)* set.

The best classification rates obtained by SFFS and SFFS-FP are reported in Table 1, while the number of features used in producing these rates can be seen in Table 2. We can see that in SFFS a small number m_{SFFS} of the 24 original features gave the best results. For SFFS-FP a slightly larger number $m_{SFFS-FP}$ of features tended to give the best results, but still smaller than 24. These extra features found useful in SFFS-FP, could be interpreted as containing some useful information, but together with a lot of noise. That is because feature selection methods like SFFS can only accept/reject features, while the Fisher algorithm can also scale them appropriately, performing a kind of "soft" feature selection and thus making use of such noisy features.

In Table 3 one can see that for greater numbers of emotions and greater numbers of features, the best-performing number of Fisher dimensions tends to be less than the maximum number of dimensions Fisher can calculate, confirming our earlier expectations (Section).

Day Dependence

As mentioned previously, the data were gathered in 20 different sessions, one session each day. During

Number of Emotions	m_{SFFS}	$m_{SFFS-FP}$
8	13	17
5 (NAGJR)	12-17	15
4 (NAGR)	9-15,18	19
4 (AGJR)	7-8	12
3 (AGR)	2-16	12
3 (AJR)	6-14	7

Table 2: Number of features m used in the SFFS algorithms which gave the best results. When a range is shown, this indicates that the performance was the same for the whole range.

their classification procedure, we noticed high correlation between the values of the features of different emotions in the same session. In this section we quantify this phenomenon in an effort to use it to improve the classification results, by first building a day (session) classifier.

Day Classifier
We use the same set of 24 features, the Fisher algorithm, and the leave-one-out method as before, only now there are $c = 20$ classes instead of 8. Therefore the Fisher projection is meaningful from 1 to 19 dimensions. The resulting "day classifier" using the Fisher projection and the leave-one-out method with MAP classification, yields a classification accuracy of 133/160 (83%), when projecting down to 6,9,10 and 11 Fisher dimensions. This is better than all but one of the results reported above, and far better than random guessing (5%). We note the following on this result:

- It should be expected that a more sophisticated algorithm would give even better results. For example we only tried using all 24 features, rather than a subset of them.

- Either the signals or the features extracted from them are highly dependent on the day the experiment is held.

- This can be because, even if the actress is intentionally expressing a specific emotion, there is still an underlying emotional and physiological state which affects the overall results of the day.

- This may also be related to technical issues, like the amount of gel used in the sensing equipment (for the BVP and GSR signals), or external issues like the temperature in a given day, affecting the perspiration and possibly the blood pressure of the actress.

Whichever the case, a possible model for the emotions could then be thought of as follows: At any point in time the physiological signals are a combination of a long-term slow-changing mood (for example a day-

Number of Emotions	Random Guessing (%)	SFFS (%)	Fisher-24 (%)	Structured subsets (%)			SFFS-FP (%)
				4-feature	6-feature	18-feature	
8	12.50	40.62	40.00	34.38	41.25	48.75	46.25
5 (NAGJR)	20.00	64.00	60.00	53.00	63.00	71.00	65.00
4 (NAGR)	25.00	70.00	61.25	61.25	70.00	72.50	68.75
4 (AGJR)	25.00	72.50	60.00	58.75	70.00	68.75	67.50
3 (AGR)	33.33	83.33	71.67	75.00	83.33	81.67	80.00
3 (AJR)	33.33	88.33	66.67	73.33	83.33	81.67	83.33

Table 1: Classification rates for several algorithms and emotion subsets.

Number of Emotions	Structured subsets			Fisher-24	SFFS-FP	Ratio
	4-feature	6-feature	18-feature			
8	3/3	3/5	5/7	6/7	4,5/7	4:1
5 (NAGJR)	3/3	4/4	3/4	3/4	3/4	3:2
4 (NAGR)	3/3	3/3	3/3	3/3	3/3	0:5
4 (AGJR)	3/3	2/3	2,3/3	3/3	2/3	3:2
3 (AGR)	2/2	2/2	2/2	2/2	2/2	0:5
3 (AJR)	2/2	2/2	2/2	1/2	2/2	1:4
Ratio	0:6	2:4	3:3	3:3	3:3	11:19

Table 3: Number of dimensions used in the Fisher Projections which gave the best results, over the maximum number of dimensions that could be used. The last row and column give the ratio of cases where these two values were not equal, over the cases that they were.

long frustration) or physiological situation (for example lack of sleep) and of a short-term emotion caused by changes in the environment (for example the arrival of some bad news). In the current context, it seems that knowledge of the day (as part of the features) may help in establishing a baseline which could in turn help in recognizing the different emotions within a day. This baseline may be as simple as subtracting a different value depending on the day, or something more complicated.

It is also relevant to consider conditioning the recognition tests on only the day's data, as there are many applications where the computer wants to know the person's emotional response *right now* so that it can change its behavior accordingly. In such applications, not only are interactive-time recognition algorithms needed, but they need to be able to work based on only present and past information, i.e., causally. In particular, they will probably need to know what range of responses is typical for this person, and base recognition upon deviations from this typical behavior. The ability to estimate a good "baseline" response, and to compare the present state to this baseline is important.

Establishing a day-dependent baseline
According to the results of the previous section, the features extracted from the signals are highly dependent on the day the experiment was held. Therefore, we would like to augment the set of features to include

both the **Original** set of 24 features and a second set incorporating information on the day the signals were extracted. A **Day Matrix** was constructed, which includes a 20-number long vector for each emotion, each day. It is the same for all emotions recorded the same day, and differs among days. There are several possibilities for this matrix. In this work, we chose the 20-number vector as follows: For all emotions of day i all entries are equal to 0 except the i'th entry which is equal to a given constant C. This gives a 20x20 diagonal matrix for each emotion.

It must be noted that when the feature space includes the Day Matrix, the Fisher projection algorithm encounters manipulations of a matrix which is close to singular. We can still proceed with the calculations but they will be less accurate. Nevertheless, the results are consistently better than when the Day Matrix is not included. A way to get around the problem is the addition of small-scale noise to C. Unfortunately this makes the results dependent on the noise values, in such an extent that consecutive runs with just different random values of noise coming from the same distribution give results with up to about 3% fluctuations in performance.

Another approach that we investigated involves constructing a **Baseline Matrix** where the Neutral (no emotion) features of each day are used as a baseline for (subtracted from) the respective features of the remaining 7 emotions of the same day. This gives

Feature Space	SFFS (%)	Fisher (%)	SFFS-FP (%)
Original (24)	40.62	40.00	46.25
Original+Day (44)	N/A	49.38	50.62

Table 4: Classification Rates for the 8-emotion case using several algorithms and methods for incorporating the day information. The "N/A" is to denote that SFFS feature selection is meaningless if applied to the Day Matrix.

Feature Space	SFFS (%)	Fisher (%)	SFFS-FP (%)
Original (24)	42.86	39.29	45.00
Orig.+Day (44)	N/A	39.29	45.71
Orig.+Base. (48)	49.29	40.71	54.29
Orig.+Base.+Day (68)	N/A	35.00	49.29

Table 5: Classification Rates for the 7-emotion case using several algorithms and methods for incorporating the day information. The "N/A" is to denote that SFFS feature selection is meaningless if applied to the Day Matrix.

an additional 24x20 matrix for each emotion.

The complete 8-emotion classification results can be seen in Table 4, while the 7-emotion classification results can be seen in Table 5. Random guessing would be 12.50% and 14.29% respectively. The results are several times that for random guessing, indicating that significant emotion classification information has been found in this data.

Conclusions & Further work

As one can see from the results, these methods of affect classification demonstrate that there is significant information in physiological signals for classifying the affective state of a person who is deliberately expressing a small set of emotions. Nevertheless more work has to be done until a robust and easy-to-use emotion recognizer is built. This work should be directed towards:

Experimenting with other signals: Facial, vocal, gestural, and other physiological signals should be investigated in combination with the signals used here.
Better choice of features: Besides the features already used, there are more that could be of interest. An example is the overall slope of the signals during the expression of an emotional state (upward or downward trend), for both the raw and the normalized signals.
Real-time emotion recognition Emotion recognition can be very useful if it occurs in real time. That is, if the computer can sense the emotional state of the user the moment he actually is in this state, rather than whenever the data is analyzed. Therefore we are

interested in examining the possibility of online recognition. This should be considered in combination with the model of an underlying mood, which may change over longer periods of time. In that respect, the classification rate of a time window using given a previous time window can yield useful information. The question is how frequently should the estimates of the baseline be updated to accommodate for the changes in the underlying mood. In addition, it appears that although the underlying mood changes the features' values for all emotions, it affects much less the relative positions with respect to each other. We are currently investigating ways of exploring this, expecting much higher recognition results.

Acknowledgments

References

Cacioppo, J. T., and Tassinary, L. G. 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45(1):16–28.

Clynes, D. M. 1977. *Sentics: The Touch of the Emotions.* Anchor Press/Doubleday.

Essa, I., and Pentland, A. 1997. Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):757–763.

Goleman, D. 1995. *Emotional Intelligence.* New York: Bantam Books.

Jain, A. K., and Zongker, D. 1997. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(2):153–158.

Lang, P. J. 1995. The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5):372–385.

Picard, R. W. 1997. *Affective Computing.* Cambridge, MA: The MIT Press.

Pudil, P.; Novovicova, J.; and Kittler, J. 1994. Floating search methods in feature selection. *Pattern Recognition Letters* 15:1119–1125.

Scherer, K. R. 1981. Ch. 10: Speech and emotional states. In Darby, J. K., ed., *Speech Evaluation in Psychiatry*, 189–220. Grune and Stratton, Inc.

Yacoob, Y., and Davis, L. S. 1996. Recognizing human facial expressions from log image sequences using optical flow. *IEEE T. Patt. Analy. and Mach. Intell.* 18(6):636–642.