Pervasive and Unobtrusive Emotion Sensing for Human Mental Health

Rui Guo*, Shuangjiang Li*, Li He*, Wei Gao*, Hairong Qi*, Gina Owens†

*Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996
†Department of Psychology, University of Tennessee, Knoxville, TN 37996

Abstract—In this paper, we present a pervasive and unobtrusive system for sensing human emotions, which are inferred based on the recording, processing, and analysis of the Galvanic Skin Response (GSR) signal from human bodies. Being different from traditional multimodal emotion sensing systems, our proposed system recognizes human emotions with the single modularity of GSR signal, which is captured by wearable sensing devices. A comprehensive set of features is extracted from GSR signal and fed into supervised classifiers for emotion identification. Our system has been evaluated by specific experiments to investigate the characteristics of human emotions in practice. The high accuracy of emotion classification highlights the great potential of this system in improving humans' mental health in the future.

I. INTRODUCTION

Modern healthcare systems have been recently facilitated by emerging biomedical sensing technologies. A large variety of biomedical sensors have been deployed to record the biological signals of human beings. These signals are subsequently aggregated, processed, and analyzed at user's hand-held mobile devices or the back-end server residing in the remote cloud, so as to monitor the health status of human beings. There have been significant efforts to monitor the humans' physiological health from their biological signals such as physical activities [1], body temperature, blood pressure, or heart beating rate [2]. Such real-time sensing enables timely observation of disease symptom and prompts disease diagnosis in a pervasive manner.

On the other hand, mental health of human beings has been generally neglected by most of the current biomedical sensing systems. The few existing schemes aim to monitor the stress levels of human beings by analyzing their biological signals [2], [3] and daily activity patterns [4], however, are insufficient for precisely reflecting the mental health status of human beings due to the extra burdened monitoring method and coarse-gained information collection. The lack of detailed information about humans' emotions underlying such stress makes it difficult to apply appropriate remedies for stress relief, and hence reduces the practical merit of stress monitoring.

In this paper, we are aiming to explore the possibility of accurately inferring various kinds of humans' emotions from a single type of biological signal, i.e., Galvanic Skin Response, which can be monitored in a unobtrusive manner from human bodies via compact and wearable sensing devices. The information provided by our system expressively indicates the mental health status of human beings, and facilitates intelligent decisions of telemedicine or smart well-being systems.

The major challenges of developing such an emotion sensing system are two-fold. First, the emotions monitored are often disturbed by mixture type of subjective or objective factors. Second, the GSR signal recorded is easily polluted by environmental influence such as humidity and the body movement. Appropriate algorithm is hence needed to ensure the accuracy of inferring human emotions.

To tackle the above challenges, our basic idea is to let experiment participants, who wear the GSR sensing devices, watch video clips with various pre-labeled emotional themes. In this way, participants' emotional arousals during the experiments can be precisely represented by their recorded GSR signal without being interfered by other physiological or psychological activities. Our detailed contributions are listed as follows:

- We conduct experiments investigating humans' psychological behaviors and emotional changes when watching video clips. A well-defined set of emotional themes is applied to the selection of video clips in order to ensure insightful investigation on the correlation between humans' personal characteristics and their emotional patterns.
- We develop various methods for eliminating the randomness and environmental noise when processing the signal, and hence facilitate high accuracy of emotion classification.
- We design efficient machine learning algorithms to infer a variety of human emotions by classifying their GSR signals. An appropriate set of features is extracted from the GSR signal to represent the fundamental characteristics of GSR signal underpinning human emotions.

II. OVERVIEW

A. GSR Sensing

The GSR signal from human bodies indicates electrical changes measured at the surface of human skin that arise when the skin receives innervating signals from the brain, and is generally considered one of the strongest features that can be used for detecting emotional arousal [5].

Our proposed system for emotion sensing is built based on the Shimmer GSR sensors[1]. Facilitated with the Shimmer sensor, the GSR signal can be either recorded in the local storage of the sensor device, or transmitted at real-time to the back-end server via wireless radios. A Shimmer GSR sensor is also equipped with a 3-D accelerometer, whose sensed data will be used by our system to eliminate the impact of humans' physical movements on their GSR signal.

B. Basic Idea

Our basic idea of emotion sensing is to explore the characteristics of temporal variations of humans' GSR signal that

---

[1]http://www.shimmer-research.com/p/products/development-kits/gsr-development-kit.

indicating different types of human emotions. Figure 1 shows some examples of the GSR signals from various emotions. We observe that different types of emotions exhibit distinct characteristics in the temporal variation of GSR signals, which can be extracted to identify human emotions using supervised classification techniques.
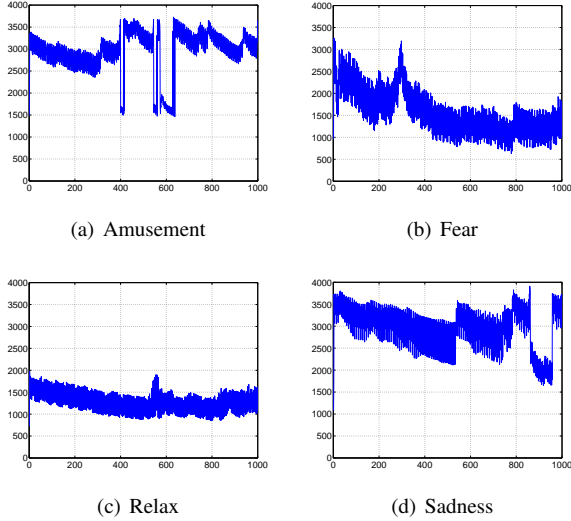

(a) Amusement

(b) Fear

(c) Relax

(d) Sadness

Fig. 1.   Raw GSR signals for 4 types of human emotions.

C. System Setup

1) Emotion Model: It is normally difficult to build a generic model for categorizing human emotions. Some basic emotions universally exist among human beings, while some other emotions are described as complex and culturally specific. Researchers usually build the emotion category models in a low dimension, e.g., Plutchik's [6] cone model and the PANA model [7]. We propose a quadrant model to represent the four basic categories of emotions: amusement, fear, relax, and sadness. These four emotions are widely accepted in modern psychology research [8]. They are also consistent with the concept that emotions can be triggered by either positive or negative factors.

2) Video Clip Selection: We adopt event-elicited methodology so as to ensure recording of subject-elicited emotions. More specifically, we choose video clips of different themes from social media websites as stimulus to arouse human emotions. As a result, the emotions of experiment participants could be affected and recorded when the participants are unaware of the purpose of the experiment. The GSR signals of experiment participants are recorded when they are watching video clips, and are then fed to supervised classifiers for emotion identification using the known labels of video clips as the ground truth.

In our system, we select 17 video clips for each category of emotions which include prank videos, thrilling movie trailers, soft music and movies with bad ending. The lengths of video clips we choose are between 5 minutes to 8 minutes to ensure that any instant or long-time fluctuations of human emotions can be appropriately captured.

III. GSR SIGNAL PREPROCESSING

After the acquisition of raw GSR signals, a preprocessing step is necessary to remove noises caused by the testing subject. This section elaborates on the GSR signal preprocessing procedure which involves mainly three steps.

A. Removing the Impact of Physical Motion

The physical movements of human beings have non-negligible impact on humans' skin conductance, and hence their GSR signals. We propose to eliminate such impact on GSR signals and the subsequent emotion classification based on the "abnormal" 3-D accelerometer data simultaneously recorded with the GSR signals. Specifically, if a sample of the accelerometer signal displays a value 5 times larger than the average, we consider this sample to be abnormal. The corresponding GSR sample, along with the 10 samples centered around it, is then considered outlier and removed from the raw signal. Since the sampling rate of the GSR sensor is set to 10 Hz, this filtering process will only remove 1-second data without noticeable change of the general patterns of the GSR signal. Note that the selection of 5 times as abnormal deviation and 10 samples as the width of the abnormal neighborhood are through extensive empirical study. Figure 2(a)(b) demonstrate the typical outliers in the raw GSR data and the accelerometer associated removal process.
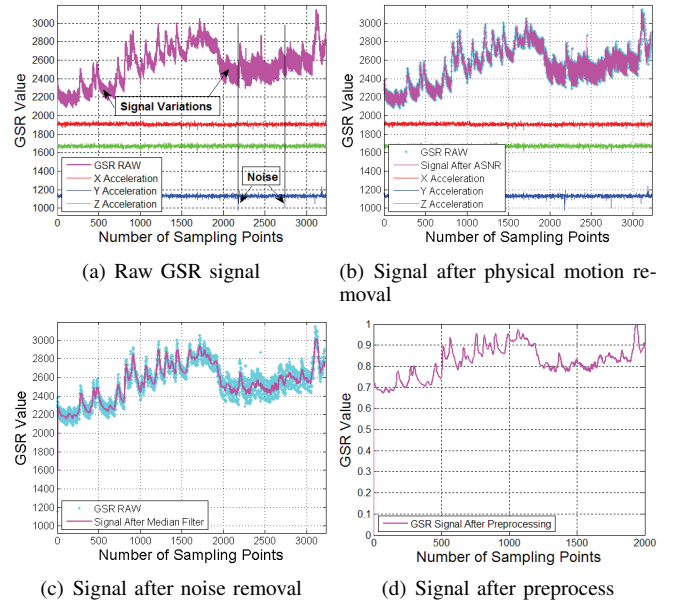

(a) Raw GSR signal

(b) Signal after physical motion removal

(c) Signal after noise removal

(d) Signal after preprocess

Fig. 2.   GSR signal preprocessing.

B. Noise Removal Using the Median Filter

The raw GSR signal usually demonstrates a large amount of small signal fluctuations, caused by the transient fluctuations of the physiological status of human bodies that, are inevitably recorded due to the high sampling rate of the GSR sensors. Removal of such noise is important for ensuring high accuracy of emotion classification.

We remove such undesirable noise from the raw GSR signal by applying the median filter over a sliding window of size W, as suggested by [9]. We set the window size as $W = 20$ data

points in order to remove the fluctuations while not blurring the signal too much. Figure 2(c) and (d) demonstrate the process of noise removal using median filter and the processed GSR signal.

C. Data Rescaling and Resampling

The last step of GSR signal preprocessing includes the rescaling and resampling. The GSR signal is rescaled to $[0, 1]$ to reduce the subjective difference. The data series are then resampled into the same length for comparison purpose. In the experiment, the sample length is defined as 2000 data points.

IV. EMOTION CLASSIFICATION

To better characterize the patterns of GSR signals for various emotions and also to reduce the computational overhead, the feature extraction step is developed to extract a lower-dimensional, well-defined set of features from the preprocessed GSR signal. Such feature vector is used as the input to the supervised classifiers.

A. Statistical Features

The primary feature set we use includes 6 statistical features described in [10]. These features have been used in the literature on the raw GSR signal, but our feature extraction is based on the GSR signal after preprocessing. The six statistical features, means and standard deviations of the raw signals, means of the absolute values of the first and second differences of the raw signals and the normalized signals, cover the typical range, variation and gradient-like characteristics of the signal, and the measurements of these features are cost-effective.

B. Running Means in the Time Domain

We add the mean values of the sliced GSR signal to the set of features, so as to capture such temporal variations. More specifically, we divide the GSR signal into several adjacent but non-overlapping slices, each of which has a size $L = 200$ data points. The mean value of each slice is used as a feature. Therefore, the running average will add 10 more features to the final feature set.

C. Running Means in the Frequency Domain

Periodic excitation is another significant characteristic when a human being experiences the amusement or fear emotion, which can be represented by frequency-domain features. We transform the preprocessed GSR signal into the frequency domain using the Fast Fourier Transform (FFT), followed by a median filter with the same specification as in Sec. III to smooth out the frequency content. We then use the similar method as in Sec. IV-B to extract the 10 mean values of the transformed GSR signal. The difference is that we only keep the first 4 mean values as they represent low-frequency content in the original GSR signal that are most related to the expression of human emotions.

D. Feature Optimization and Selection

The final feature vector generated from the preprocessed GSR signal includes the 6-dimensional statistical feature, the 10-dimensional time-domain running means, and the 4-dimensional frequency-domain running means. We demonstrate the normalized features from each type of emotions in Figure 3.

Although the feature vector is formulated from different perspectives, they are not completely independent. We incorporate the Sequential Floating Forward Selection (SFFS) technique [11] to find the optimal feature subspace for emotion classification. The SFFS always constructs the "best" combination of features to maximize the inter-class distance in emotion classification. Usually the k-nearest-neighbor (kNN) classifier and its classification rate are used as the criterion for SFFS. Without any pre-knowledge about the feature dimension, the SFFS helps construct the appropriate feature vector and calculate the classification rate using a non-exhaustive way.
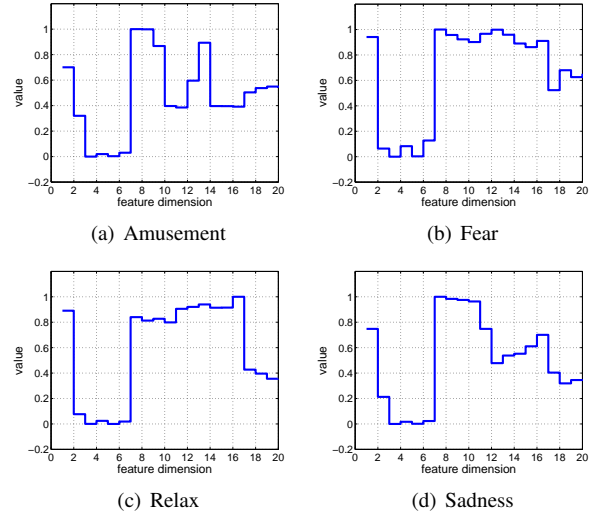

(a) Amusement   (b) Fear

(c) Relax   (d) Sadness

Fig. 3. Illustration of feature vectors for different emotions. The first six dimensions are statistic features. The $7th$ to $16th$ features are the time-domain means of the sliced GSR signal. The rest part consists of the first 4 dimensions of the frequency-domain means of the sliced GSR signal.

V. EXPERIMENTS AND EVALUATION

In this section, we present our experimental evaluation results regarding the accuracy of our proposed emotion classification approach. The experiments are conducted over 4 subjects with their GSR signals collected when they are watching the video clips.

A. Experiment Setup

Different human beings exhibit heterogeneous characteristics in their emotion patterns. In this section, we try the supervised classifier and perform emotion classification for each experiment subject individually. A two-fold cross validation strategy is adopted in all experiments. For each emotion, we randomly divide the recorded GSR signal into two parts with equal numbers as training and testing sets. We use the kNN classifier to evaluate the benchmark test performance. The average and maximal classification rates are calculated.

To implement the SFFS-based classification, we employ the 10-fold cross validation strategy on the individual GSR datasets. The kNN classifier is kept in the SFFS process. Once the best subset of the features from the training set is selected, it will be automatically applied to the test data to calculate the classification accuracy. To reduce the variation caused by the random training and test data selection, the SFFS-based classification runs for 100 times.

B. Experimental Results

The comparison of the overall classification accuracy of different schemes is listed in Table I. The overall classification accuracy is the average value which is computed from all experimental subjects. From Table I, we can see that the classification accuracy of our proposed scheme is generally 10% higher than that in the existing work [10] which relied on sensed data from multiple sources. For comparison, we also design the experiment on emotion dependent scheme which mix the emotion samples from different experiment subjects as the baseline. It is clear to see the baseline experiment receives the lowest classification accuracy. The benchmark experiment which adopts the subject dependent strategy effectively increases the classification accuracy by limiting the inter-subject difference, as shown in Table II.

Result comparison	
Emotions	4
kNN baseline	30.70%
Picard 1 [10]	68.70%
Picard 2 [10]	67.50%
Proposed	**79.45%**

TABLE I
RESULT COMPARISON OF DIFFERENT METHODS. KNN BASELINE IS THE EMOTION DEPENDENT EXPERIMENT WHICH MIX THE SAME EMOTION SAMPLES FROM DIFFERENT SUBJECTS. PICARD 1 AND 2 GIVE THE 4-EMOTION CLASSIFICATION RATES IN [10].

The utilization of SFFS does not only reduce the length of the feature vectors, it also improves the overall classification accuracy. When counting usage frequency of each dimension of the feature vector, the time-domain means are the most selected feature in SFFS and play a dominant role in emotion classification.

Subj.	kNN benchmark		SFFS		SFFS+Label		F.S.
	Avg.	Max.	Avg.	Max.	Avg.	Max.	
1	38.89%	48.15%	58.00%	68.57%	70.86%	77.14%	10/20
2	35.26%	53.57%	57.22%	61.11%	68.23%	77.78%	8/20
3	42.22%	55.56%	65.14%	70.14%	71.43%	82.86%	5/20
4	47.78%	66.67%	56.67%	65.71%	70.29%	80.00%	9/20

TABLE II
RESULTS FOR EMOTION CLASSIFICATION. KNN BENCHMARK IS THE INITIAL EXPERIMENT. THE COLUMNS SFFS GIVE THE CLASSIFICATION RESULTS WHEN SFFS IS USED. THE COLUMNS SFFS+LABEL GIVE THE CLASSIFICATION RESULTS WITH AUXILIARY QUESTIONNAIRE DATA. AVG. IS THE AVERAGE CLASSIFICATION RATE OF 100 TRIALS WITH RANDOM TRAINING AND TEST SETS. MAX. IS THE HIGHEST CLASSIFICATION RATE. F.S. STANDS FOR FEATURE SELECTION

We further improve the accuracy of emotion classification by using auxiliary data of subjective feedback from human beings about their feelings. The mental state of human beings after watching the video clips is an acquired emotion rather than a proprioceptive sense. Therefore, we design the questionnaires which record the real feelings after the subject watches the video. They can evaluate the feelings by assigning a point from 0 to 5 to the videos. 0 indicates that the pre-defined video label is not related to their feelings after watching at all. The real emotion should be explicitly given in this case. 5 means that the label of the video exactly matches humans' feelings. Accessing the external data like the questionnaire points, we change the labels according to the real emotion acquired rather than the pre-defined label. The interactive evaluation makes the labeling closer to the ground truth.

Having access to the questionnaire data, we rebuild experiments with the interactive labeling. The SFFS-based classification accuracy for four emotion categories generally reaches around 80% as shown in Table II, which is the *best* performance reported in the single modality GSR data emotion classification.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a single-modality human emotion sensing system based on GSR sensing. Our experimental results based on GSR signals were encouraging, given its simplicity, pervasiveness, and unobtrusiveness. The extracted features from the affective GSR signal sequence were verified to be effective in emotion classification within the subject-dependent context. Without the external input from other modalities, the proposed algorithm was boosted by SFFS and the interactive labeling scheme. The final classification accuracy for four classes of emotion reached the highest rate compared to the state-of-the-art. Inspired from the experimental framework and results, our future work focuses on developing the subject-dependent mobile applications to monitor the personal emotion changes at real-time with an online learning strategy. The goal of such a mobile application is to tweet the stream to personal device once detected his/her emotion with some intentional content. For instance, the user with sadness emotion may have a negative impact on the work. The soft music and delight color content should be tweeted to alleviate the mood of the user. The emotion sensing and classification trigger the application automatically after the user configures and trains the model. The feasibility of the proposed system makes it applicable in mobile sensing devices and humans mental health monitoring.

REFERENCES

[1] Rodrigo de Oliveira and Nuria Oliver, "Triplebeat: enhancing exercise performance with persuasion," in *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. ACM, 2008, pp. 255–264.

[2] Emre Ertin et al., "Autosense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2011, pp. 274–287.

[3] Andrew Raij et al., "mstress: Supporting continuous collection of objective and subjective measures of psychosocial stress on mobile devices," *ACM Wireless Health 2010 San Diego, California USA*, 2010.

[4] Yadid Ayzenberg et al., "Feel: frequent eda and event logging–a mobile social interaction stress monitoring system," in *Proceedings of the ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, 2012, pp. 2357–2362.

[5] Wolfram Boucsein, *Electrodermal activity*, Springer, 2011.

[6] Robert Plutchik, "The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89, no. 4, 2001.

[7] David Watson and Auke Tellegen, "Toward a consensual structure of mood.," *Psychological bulletin*, vol. 98, no. 2, pp. 219, 1985.

[8] Posner J et al., "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, pp. 714–734, 2005.

[9] Jorn Bakker et al., "What's your current stress level? detection of stress patterns from gsr sensor data," in *Data Mining Workshops, 2011 IEEE 11th International Conference on*, pp. 573–580.

[10] R.W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 10, pp. 1175–1191, 2001.

[11] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recogn. Lett.*, vol. 15, no. 11, 1994.