



Techniques in Pattern Recognition for School Bullying Prevention: Review and Outlook

Liang Ye

*Communication Research Center, Harbin Institute of Technology,
P.O.Box 3043, No.2 Yikuang Street, Harbin, 150080, China;
Department of Electrical Engineering, University of Oulu,
Erkki Koiso-Kanttilan katu 3 P.O.Box 4500, Oulu, 90014, Finland*

coldwound@163.com

Hany Ferdinando

*University of Oulu, Erkki Koiso-Kanttilan katu 3 P.O.Box 4500, Oulu, 90014, Finland;
Petra Christian University, Siwalankerto 121-131, Surabaya, 60236, Indonesia*

hanyf@peter.petra.ac.id

Esko Alasaarela

University of Oulu, Erkki Koiso-Kanttilan katu 3 P.O.Box 4500, Oulu, 90014, Finland

esko.alasaarela@ee.oulu.fi

Abstract

School bullying is a serious problem among teenagers. With the development of sensor technology and pattern recognition algorithms, several approaches for detecting school bullying have been developed, namely speech emotion recognition, mental stress recognition, and activity recognition. This paper reviews some related work and makes some comparisons among these three aspects. The paper analyzes commonly used features and classifiers, and describes some examples. The Gaussian Mixture Model and the Double Threshold classifiers provided high accuracies in many experiments. By using a combined architecture of classifiers, the results could be further improved. According to the results of the experiments, the six basic emotions, high mental stress and irregular movements can be recognized with high accuracies. So the three types of pattern recognition can be used for school bullying detection effectively. And these techniques can be used on consumer devices such as smartphones to protect teenagers.

Keywords: pattern recognition, speech emotion recognition, mental stress recognition, activity recognition, school bullying

1. Introduction

School bullying means aggressive behaviors or words, which hurt another person intentionally. It is often applied by the stronger upon the weaker, or by the elder upon the younger. School bullying is a serious social problem among teenagers, especially in high grades of primary schools and junior middle schools, for example, in China. It is very common not only in Eastern countries but also in Western countries. Victims in school bullying suffer from both mental and physical violence. School bullying is considered as one of the main reasons for depression, dropping out of school and adolescent suicide.

Preventing school bullying is an important and timeless topic. School bullying has been studied since the 1960s in Sweden, Finland, and Norway. As sensor technology develops, there are now new approaches for detecting and preventing school bullying. School bullying can occur through several situations, for example, cursing, physical violence, and so on. Correspondingly, there are improved techniques to detect these. These techniques are involved in building a school bullying prevention system.

The effect of bullying upon victims includes increased levels of depression and anxiety [1]. The human body gives apparent signals indicating these. The most common signals are

reactions by the heart, voice, and lungs and therefore such signals can be used to indicate bullying. When a physical bullying event happens, the most intuitive signals are the motion signals.

Pattern recognition plays an important role in detecting school bullying and for this reason it is a widely studied topic in many aspects. This paper focuses on voice and motion recognitions, since they can best describe an individual's state. The related techniques are speech emotion recognition, mental stress recognition, and activity recognition correspondingly.

The remainder of this paper is constructed as follows: Section 2 shows some existing school bullying prevention systems; Section 3 describes detailed pattern recognition techniques used for school bullying detection; Section 4 presents a discussion on future work and provides a conclusion.

2. School Bullying Prevention Systems

There are already several solutions for school bullying prevention systems developed for consumer devices such as smartphones. Some examples are presented in the following:

The ICE BlackBox is a personal security application. When a bullying event occurs, the user is able to press a button to activate the ICE BlackBox. Then the system will record the audio, video, and GPS location, and send the information to the ICE BlackBox secure servers. The application will also send text or email messages to the user's family or friends to alert them. However, it is often not practical for the user to activate the application manually when bullying occurs. Furthermore, it is difficult for the user to aim the camera towards the bullies in the event of physical violence, and the presence of the camera may intensify bullying.

In the application TipOff, school bullying as well as crimes can be reported to the server by text. However, the TipOff also requires manual activation.

The application Bully Block blocks unwanted calls or SMS messages from bullies, just like the incoming call firewall in many mobile phones. The Bully Block can record voice during a bullying call, and also keeps a list of blocked calls and SMS messages. The user can set an auto-response to the blocked numbers. But if the bullies change their phone numbers, the block list should be manually updated.

In terms of user friendliness, an anti-bullying system should be aware of what is happening autonomously, and perform corresponding actions (e.g. send SMS messages to parents or friends) automatically. A smartphone is equipped with a 3-axis gyro, an accelerometer and a microphone, so it is fairly capable of detecting a school bullying event autonomously, in combination with pattern recognition technology.

3. Pattern Recognition in School Bullying Detection

A smartphone has several kinds of built-in sensors, so there can be several ways to measure the user's state, e.g. by speech, movement, etc. And then the smartphone can recognize the user's emotion, mental stress, activity, and so on. Since it is not easy for the victim to aim a camera towards the bullies, image recognition by the camera is not taken into consideration.

A common pattern recognition flow chart is given in Fig. 1, which can be used in speech emotion recognition, mental stress recognition, and activity recognition. After acquiring data from sensors, preprocessing is needed to filter out noise [2]. Then extraction of necessary features from the signals and the classification is performed. The selected features and classifiers will be discussed in detail in the following subchapters.

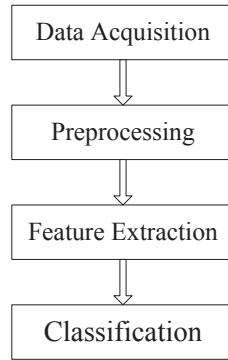


Fig. 1: A common pattern recognition procedure.

3.1 Speech Emotion Recognition

Emotion is a significant feature that can be used to detect school bullying. A victim being bullied may be sad, frightened, or angry. From the victims speech, a smartphone can acquire his/her emotion, and thus recognize the situation. Speech emotion recognition has been well studied in recent years. Generally, there are six basic emotions which are happiness, sadness, anger, fear, surprise, and disgust [3]. Most work was done with these emotions.

The most common voice features include pitch [4-9], energy [4, 5], intensity [6], formants [7], and MFCCs (Mel Frequency Cepstral Coefficients) [4, 8, 10, 11].

(a) Pitch. Pitch is one of the most popular features used for speech emotion recognition. It is a very sensitive factor that responds to the auditory sense. Pitch is also known as a fundamental frequency and refers to the periodic time of a wave pulse generated by air compressed through the glottis from the lungs [3]. However, pitch is also the most gender-dependent feature [12]. Yoon [13] had taken gender differences into consideration while applying speech emotion recognition.

(b) Energy. Energy refers to the volume or intensity of the speech, and contains valuable information. Energy can be used to differentiate sets of emotions in combination with other vocal features, but this feature alone is not sufficient for basic emotion recognition. Scherer [14] stated that fear, joy, and anger have a high energy level, whereas sadness has a low energy level.

(c) Intensity. Intensity is described as the power or energy of the voice. The intensity of a voice can be physically noticed through the pressure of sounds or a subjective level of noisiness.

(d) Formants. Formant is an important feature that reflects the quantitative characteristics of a vocal track [15]. The location of vocal tract resonances depends on the shape, and the physical dimensions of the vocal tract are measured in the frequency domain. Subsequently the resonances tend to form the overall spectrum where speech scientists refer to them as formants. Each formant is characterized by its center frequency and its bandwidth [16]. Formants can be used to differentiate improved articulated speech from speech that is slackened. The formant bandwidth during slackened articulated speech is gradual, whereas the formant bandwidth during improved articulated speech is narrow with steep flanks.

(e) MFCC. MFCC is a very common group of features used in automatic speech recognition, which converts the basic features (pitch, power, etc.) into 12 MFCC features. MFCC

has been proved to be more informative than Linear Prediction Coefficient (LPC) features in terms of reducing Cross Validation (CV) error of multi-class classifiers [10].

There is a large number of classifiers for speech emotion recognition, e.g. K-Nearest Neighbor (K-NN) [4], Maximum Likelihood Bayes (MLB) [5], Artificial Neural Network (ANN) [6], MultiLayer Perceptron (MLP) [7], Hidden Markov Model (HMM) [11], Gaussian Mixture Model (GMM) [17], Support Vector Machine (SVM), Neural Network (NN) [18], Probabilistic Neural Network (PNN) [19], etc.

The databases used for training or recognizing include DES (Danish Emotional Speech) [5], SUSAS (Speech Under Simulated and Actual Stress) [5], EMO-DB (Berlin EMOtional DataBase) [20], ORI (Oregon Research Institute) [21], etc., which are in different languages. EMO-DB is the database most commonly used by researchers [4, 12, 22, 23]. Some researchers also use real-life recordings as the databases [24, 25].

A comparison of some related works is given in Table 1. The average accuracies are either drawn directly from the references, or calculated from the raw data in them. According to Table 1, all the six basic emotions have pretty high recognition accuracies, i.e. the emotions can be clearly distinguished from each other. So when a victim is being bullied, a smartphone can detect his/her sorrow, fear or anger from his/her speech.

In finer detail, different classifiers fit different emotions. Since the experimental conditions and involved databases were different, it is difficult to draw a conclusion as to which classifier is the best from these experiments. But a comparison of the classifiers within the same experiment is possible. GMM gave the best accuracy in comparison to other classifiers in the same experiment.

Some researchers used a combined architecture of classifiers for speech emotion recognition. For example, Chung-Hsien Wu [7] proposed a 2-level classifier. The base-level classifier consisted of GMM, SVM, and MLP, and the average accuracies of these were 73%, 78% and 72%, respectively. After combination with the MDT (Meta Decision Tree) classifier, the average accuracy could reach 80%. In combination with other non-voice recognition methods, some results in Table 1 could be further improved. For example, Chung-Hsien Wu [7] developed a Semantic Label (SL) approach, which recognized keywords that could present an emotion from the speech. By co-operating with his 2-level voice classifier, the final average accuracy arrived at 86%.

3.2 Mental Stress Recognition

Mental stress is another important feature in school bullying detection. When being bullied, the victim may be nervous and have high-level mental stress. But this feature alone cannot recognize a school bullying event, since there are too many other situations that also make people nervous, for example, when playing games, working out a problem, and so on. So, mental stress should be used in combination with speech emotion for school bullying detection. A list of related works is given in Table 2.

The involved features are EEG (Electroencephalography) [2], MFCCs, TEO-PWP (Teager Energy Operator - Perceptual Wavelet Packet) features [19], PD (Pupil Diameter), PPG (Photoplethysmogram) [33], MV (Morphologic Variability) of ECG (Electrocardiography) [34], HRV (Heart Rate Variability) [34],[35], HR (Heart Rate), SC (Skin Conductance), ECG, EMG (Electromyography) [36], PDM (Principal Dynamic Mode) and PSD (Power Spectral Density) features from HR [37], and PCG (Phonocardiography) [38].

Table 1: Speech emotion recognition accuracies.

Author (et al.)	Features	Classifier(s)	Accuracy
A.A. Mohamoud[4]	MFCCs, LPC, energy, and pitch	K-NN (K = 8)	Angry: 80%
B. Vlasenko [11]	MFCCs and log energy	HMM	Average: 81%
C.-H. Wu [7] ^a	Spectrum, formant, and pitch	GMM, SVM, MLP, and MDT	Neutral: 76%
			Happy: 80%
			Sad: 82%
			Angry: 81%
D. Neiberg [8]	Pitch and MFCCs	GMM	Average: 93%
D. Vernendis [5]	Formats, pitch, energy, and spectral features	MLB	DES database: 54%
			SUSAS database: 57%
H. Altun [10]	Pitch, energy, MFCCs, and LPC	Multi-class SVM	Average: 80%
I. Luengo [16]	Energy, pitch, and voice detection	GMM	Average: 92%
J. Sidorova [6]	Formants, intensity, and pitch	ANN	Average: 79%
J. Nicholson [18]	Speech power, pitch, LPCs	NN	Average: 50%
L. He [19]	TEO-PWP and MFCCs	GMM and PNN	Angry: 54%
			Anxious: 57%
			Dysphoric: 54%
			Neutral: 61%
			Happy: 57%
O.W. Kwon [9]	Pitch, energy, formant, and MFCC	Gaussian SVM	Average: 41%
S. Haq [26]	Pitch, energy, duration, and MFCC	MLB	Average: 53%
S. Emerich [22] ^a	MFCCs and 7 statistical moments	SVM	Happy: 80%
			Disgust: 82%
			Neutral: 91%
			Angry: 93%
			Sad: 93%
			Fear: 87%
S. Pathak [23]	LPCs	NN	Sad: 46%
			Angry: 35%
			Happy: 36%
			Fear: 55%
			Boredom: 44%
T. Iliou [12]	Pitch, energy, formant, and MFCC	K STAR (K*)	Average: 74%
T. L. Nwe [27]	LFPC, LPCC, and MFCC	HMM	Average: 78%
T. Yamada [28]	Frequency and bandwidth	NN	Average: 70%
T.-L. Pao [29]	LPC and MFCCs	K-NN	Average: 80%
V.A. Petrushin[30]	Pitch, formants, bandwidth, etc.	K-NN	Average: 70%
Y.-H. Kao [31]	Pitch, energy, formant, and MFCC	SVM	Average: 90%
Z.-Y. Han [32]	F0, energy, duration, model jitter, tremor, pitch, the first 3 formants, bandwidth, spectral energy distribution, Harmonic to Noise Ratio, voice to unvoiced energy ratio, and glottal flow	ANN	Angry: 78%
			Disgust: 91%
			Fear: 75%
			Joy: 80%
			Neutral: 94%
			Sad: 83%
			Surprise: 83%

^a. If combined with other non-voice features (e.g. semantic labels [7], facial expressions [22]), the accuracy could be further improved.

Table 2: Mental stress recognition accuracies.

Author (et al.)	Features	Classifier(s)	Accuracy
A. Saidatul [2]	EEG	ANN	Average: 91%
F. Mokhayeri [33]	PD, ECG, and PPG	Fuzzy SVM	Average: 79%
H. Costin[34]	HRV and MV of ECG	Minimum Distance	MV frequency domain: 89% MV time domain: 81% HRV time domain: 81% HRV frequency domain:76%
J. Wijsman [36]	HR, ECG, SC, and EMG	LBN, QBN, K-NN, and FLS	LBN: 78% QBN: 78% K-NN: 76% FLS: 79%
J. Choi [37]	PDM and PSD features from heart rate	K-NN (K=7)	Within-subject: 83% Between-subjects: 69%
J. Choi [35]	HRV	ARMAX	Average: 66%
J. Choi [39]	HRV	ARMAX	Average: 80%
L. He [19]	TEO-PWP and MFCCs	GMM and PNN	High: 97% Moderate: 94% Neutral: 91%

(a) ECG. ECG is a transthoracic interpretation of the electrical activity of the heart over a period of time, as detected by electrodes attached to the surface of the skin and recorded by a device external to the body. An ECG is used to measure the rate and regularity of heartbeats, as well as the size and position of the chambers, the presence of any damage to the heart, and the effects of drugs or devices used to regulate the heart, such as a pacemaker.

(b) EEG. EEG [2] is a highly complex signal. It is one of the most common sources of information used to study brain function and condition. Raw EEG signals have amplitudes of the order of microvolts and contain frequency components up to 100 Hz. Hence, preprocessing is often required which includes filtering, and artifact removals.

(c) EMG. EMG [36] is a technique for evaluating and recording the electrical activity produced by skeletal muscles. EMG is performed using an instrument called an electromyograph to produce a record called an electromyogram. An electromyograph detects the electrical potential generated by muscle cells when these cells are electrically or neurologically activated. The signals can be analyzed to detect medical abnormalities, activation level, recruitment order or to analyze the biomechanics of human or animal movement.

(d) PCG. PCG is a very common technique in heart auscultation. Unfortunately, heart sound is not free from noise, e.g. respiratory sound, and they share the same frequency. So, linear filters have little effect in separating them [40] and a special filter should be used. Independent Component Analysis (ICA) [38] is a candidate to separate them. Both heart and respiratory rates are combined to give information about a users mental stress. Currently, little work has been done on emotion and mental stress detection based on the PCG signal. This idea brings novelty to this research work.

(e) PPG. PPG is an optically obtained plethysmogram, a volumetric measurement of an organ. A PPG is often obtained by using a pulse oximeter which illuminates the skin and measures changes in light absorption [41]. A conventional pulse oximeter monitors the perfusion of blood to the dermis and subcutaneous tissue of the skin. Because blood flow to the skin can be modulated by multiple other physiological systems, the PPG can be used to monitor breathing, hypovolemia, and other circulatory conditions [42].

(f) SC. SC [36] is a method of measuring the electrical conductance of the skin, which varies with its moisture level. The sweat glands are controlled by the sympathetic nervous system, so skin conductance is used as an indication of psychological or physiological arousal. Therefore, if the sympathetic branch of the autonomic nervous system is highly aroused, then sweat gland activity will also increase, which in turn increases skin conductance. In this way, skin conductance can be used as a measure of emotional and sympathetic responses. There has been a long history of electrodermal activity research, most of it dealing with spontaneous fluctuations or reactions to stimuli.

(g) TEO-PWP. TEO-PWP is another approach to extract features from speech signals besides the MFCCs. A detailed feature extraction process can be found in Ling He’s work [19].

Although, with a smartphone it is difficult to acquire some of the features mentioned in Table 2, such as ECG, EEG, EMG, and SC, voice features like MFCCs are sufficient to recognize mental stress. And if the smartphone is near enough to the heart, it can also acquire HR and HRV features from the microphone by measuring the PCG signal.

The classifiers include ANN [2], GMM, PNN [19], Fuzzy SVM [33], Minimum Distance [34], LBN (Linear Bayes Normal), QBN (Quadratic Bayes Normal), K-NN, FLS (Fishers Least Square) [36], and ARMAX (AutoRegressive Moving Average model with eXogenous inputs) [35],[39]. Different classifiers fit different features. Again, GMM outperformed PNN [19] in the same experiment. By using a normalization method [39], Choi improved the recognition accuracy of his previous work [35] by 14%.

Normally there are two kinds of classification experiments, namely within-subject and between-subjects [37]. The difference lies in whether the test subject is in the training dataset. If the test subject is in the training dataset, the experiment is called within-subject; otherwise, it is called between-subjects. This classification is similar in speech emotion recognition. The two kinds of classification experiments are called speaker-dependent and speaker-independent, respectively. Within-subject and speaker-dependent experiments usually have better recognition accuracies, but between-subjects and speaker-independent experiments are more universal.

In Table 2, a higher level of mental stress is easier to recognize. When being bullied, the victims often have high levels of mental stress, so mental stress can be used to detect school bullying in company with emotion recognition.

3.3 Activity Recognition

Emotion and mental stress recognitions from speech can detect verbal bullying, whereas activity recognition can detect physical violence. A smartphone has a built-in 3-axes gyro and an accelerometer, so it can recognize many irregular movements and daily life activities. A list of related works is given in Table 3.

The main features in activity recognition are 2D/3D acceleration data, inclination angle, GPS speed, and their statistic values such as standard deviation, skewness, Signal Magnitude Area (SMA), entropy [43],[44],[46]-[50], as well as KDA features (Kernel Discriminant Analysis) [45].

(a) SMA of 3-axis acceleration. SMA is a viable activity and mobility measure. Peaks occur in the SMA signal when people sit, rise from a chair, or lie down. SMA also helps identify activity intensity changes that can indicate a change of state.

Table 3: Activity recognition accuracies.

Author (et al.)	Features	Classifier(s)	Accuracy
A. Sorvala [43]	3D acceleration and gyro	State Transition Diagram based classifier	Stand→sit: 93% Sit→stand: 87% Sit→lie: 100% Lie→sit: 87% Stand→lie: 97% Lie→stand: 80% Stand→walk: 90%
A.Henpraserttae [44]	3D acceleration	K-NN (K=3)	Lying: 100% Sitting: 96% Standing: 88% Walking: 96% Running: 88% Jumping: 82%
A. M. Khan [45]	KDA features	ANN	Resting: 99% Walking: 95% Walking upstairs: 95% Walking downstairs: 92% Running: 99%
G. Hache [46] ^a	Inclination angle, 3D acceleration (standard deviation, skewness, and signal magnitude area), light intensity, and GPS speed	DT	Walk↔stop: 97% Stand↔sit↔lie: 98% Stair ascent: 30% Stair descent: 83% Ramp walking: 43% Start/stop a car: 66%
G. Hache [47]	2D acceleration data (standard deviation and signal magnitude area)	DT	Stand→walk: 97% Walk→stand: 97% Stand→sit: 98% Sit→stand: 93% Stand→lie: 100% Lie→stand: 100%
M. Muehlbauer [48] R. K. Ganti [49]	3D acceleration and gyro 3D acceleration data (mean, energy, skewness, and entropy) and cepstral coefficients of sound	K-NN (K=3) HMM	Average: 94% Aerobic: 82% Cooking: 100% Desk Work: 53% Driving: 88% Hygiene: 99% Watching TV: 88%
S.-K. Song [50]	3D acceleration and inclination angle	Multi-Layer Perceptron	Running: 91% Stand→sit: 98% Walking: 89% Fall down: 90% Sitting: 92% Lie→sit: 98% Stand→lie: 94%

a. If combined with other non-activity features (e.g. context image [46]), the accuracy could be further improved.

(b) Standard deviation of vertical acceleration. Standard deviation is a well-supported measure for activity classification [51]. Since most daily activities can be classified by changes in vertical axis acceleration, vertical acceleration is used to differentiate between static and dynamic states by comparing the standard deviation.

(c) Skewness of vertical acceleration. The skewness value of the vertical acceleration is a time-domain feature which is used to differentiate walking/running from going up/down stairs [52]. Skewness is a measure of asymmetry of the vertical acceleration about the average acceleration.

(d) Inclination angle. Inclination angle is used to help classify posture [53] and identify postural transition [54].

(e) GPS data. GPS location coordinates and speed are used in mobility monitoring to complement motion data, improve activity recognition, and provide contextual data [55].

(f) KDA features. KDA is a non-linear discriminating approach based on kernel techniques to find non-linear discriminating features [45].

The classifiers are K-NN [44],[48], ANN [45], DT (Double Threshold) [46],[47], HMM [49], and Multi-Layer Perceptron (a widely used Neural Network classifier) [50]. DT is a commonly used classifier in activity recognition. It uses two thresholds to recognize the state changes, which allows small errors caused by noises, and has a high accuracy. Some researchers also proposed combined architectures for classification. Gaetanne Hache [46] incorporated both movement features and context images to be able to recognize a certain activity. The average accuracy after combination was improved by 13%.

Apiwat Henpraserttae [44] argued that the device orientation and location, as well as the size of training dataset, could affect the recognition accuracy. According to the experimental results, the waist was the best device location.

In the previous work [43] by the authors' research group, a State Transition Diagram based classifier was implemented upon the 3-axis acceleration and gyroscope data to perform the activity recognition. They focused on the state transition rather than the state itself, since the irregular movement is more important in both school bullying detection and elderly care [43],[56],[57]. The results were an average accuracy of 90.5% in seven kinds of state transition, and it would be high enough to detect physical violence in a school bullying situation. The body's irregular movement during physical violence can be treated as special state changes. The accuracy of school bullying detection could be further improved by combining the speech emotion recognition and mental stress recognition together with this activity recognition, which will be the authors' future work.

Overall, in Table 3, state changes and irregular movements can be well recognized. So by activity recognition, a smartphone can know whether the user is suffering from physical violence, e.g. pushing, beating, etc.

4. Discussion and Conclusion

School bullying is a common social problem, which affects teenagers both mentally and physically. It has been studied for a long time. As sensor technology has developed, new approaches to prevent school bullying have been used based on consumer devices. With multiple sensors built in, a smartphone is capable of detecting school bullying by means of speech emotion recognition, mental stress recognition, and activity recognition.

Speech signals are used for emotion recognition and mental stress recognition. From the conversation of the bullies and the victims, a smartphone can recognize their emotions and levels of mental stress. Are they angry? Or are they nervous? Then it can decide whether

a verbal bullying happens. Movement of the body can tell us whether the user is under physical violence, such as being beaten (shaking of the body), or being pushed (irregular movement, unlike walking). So by combining the three types of pattern recognition, a smartphone is capable of detecting most school bullying events. And then the smartphone should send an SMS message to the victim's parents or friends to prevent the bullying. This paper has reviewed some related work and made some comparisons in the three aspects.

In speech emotion recognition, features like MFCCs, pitch, and energy, and classifiers like GMM, PNN, and SVM are discussed. Among all the features, the MFCCs were used most commonly. In practical use, not all the MFCCs were necessary, but a suitable selection of MFCCs sometimes could give better performance. Different classifiers fit different emotions, so it is not easy to draw a conclusion which classifier is best for common use. However, GMM performed best in comparison with other classifiers within the same experiment. Moreover, a combined architecture of classifiers (multi-level classifiers) is suggested to improve the accuracy of emotion recognition. Co-operating with other non-speech recognition methods can also give a higher accuracy. In specific use, when selecting features and classifiers, researchers should choose the ones, which can best recognize the emotions they need, instead of an overall performance. Furthermore, emotions can be classified into groups, instead of single classes. For example, this paper focuses on the detection of school bullying, so angry and sad can be marked as bullying emotions, whereas neutral and happy can be marked as non-bullying emotions. Intra-group recognition is allowed to have high misclassification rates, but inter-group recognition should have low misclassification rates.

In mental stress recognition, features like MFCCs, ECG, EEG, and EMG, and classifiers like GMM, PNN, ANN, and K-NN are discussed. Again GMM outperformed other classifiers in the same experiment. Although features like ECG, EEG, and EMG cannot be acquired by normal smartphones, voice features such as MFCCs are sufficient to recognize mental stress. One of the most possible solutions on a smartphone is the PCG signal. From this signal, heart and respiratory rates can be measured to detect the mental stress, however, advanced linear filters are needed to separate them. Since there are many cases, which can cause high-level mental stress, mental stress recognition cannot be used alone to describe a certain situation, but is often used together with other techniques such as speech emotion recognition. However, this does not mean that mental stress recognition is unimportant. It can be used as an assistant technique in many issues. For example, in school bullying detection, when a boy shouts loudly, this may be classified as either angry or happy (these two emotions have a relative high misclassification rate). But if mental stress recognition is applied together with speech emotion recognition, the misclassification rate may be reduced, for angry usually has a high-level mental stress.

In activity recognition, features like 2D/3D acceleration data, inclination angle, and GPS speed, and classifiers like DT and HMM are discussed. Some statistical features of the original signals are also involved in, e.g. standard deviation and signal magnitude area (SMA). SMA is a commonly used feature by many researchers. It is a combination of the three axes of the acceleration, and its variation can indicate the change of peoples state. The advantage of SMA is that it does not depend on the attitude of the sensor. However, the disadvantage of SMA is that it loses the direction information. So it is often used to detect the change of state, instead of static states. The three separate vectors of the 3D acceleration can give out much more information than the SMA, but the complexity is much larger. So a trade-off is often applied in specific use. For example, in school bullying detection, since physical bullying often happens from horizontal direction, the

three vectors can be combined into two, i.e. a horizontal vector and a vertical vector. DT is a commonly used classifier, which allows small errors caused by noise, and gives high accuracies. According to the researchers results, state changes and irregular movements can be well recognized. So, physical violence in a school bullying event, e.g. pushing and beating, can be easily detected.

Truly there are many researchers who are studying these topics, and the authors cannot refer to them all in one paper. In future work, the authors shall appreciate more contributions of these researchers, and give more detailed solutions for school bullying prevention on consumer devices.

Acknowledgments

This work was supported by the Natural Scientific Research Innovation Foundation in Harbin Institute of Technology under Grant No. HIT.NSRIF.2013029, the National Science and Technology Major Project under Grant No. 2012ZX03004003, the 973 Program under Grant No. 2013CB329003, and partly by the China National Science Foundation under Grant No. 61101123.

References

- [1] S. C. Hunter and M. Borg, "The influence of emotional reaction on help seeking by victims of school bullying," *Educational Psychology*, vol. 26, no. 6, pp. 813-826, 2006.
- [2] A. Saidatul, M. P. Paulraj, S. Yaacob, and M. A. Yusnita, "Analysis of EEG signals during relaxation and mental stress condition using AR modeling techniques," *2011 IEEE Int. Conf. Control System, Computing and Engineering*, Penang, Malaysia, Nov. 25-27, 2011, pp. 477-481.
- [3] M. N. Hasrul, M. Hariharan, and Sazali Yaacob, "Human affective (emotion) behaviour analysis using speech signals: a review," *2012 Int. Conf. on Biomedical Engineering*, Penang, Malaysia, Feb. 27-28, 2012, pp. 217-222.
- [4] A. A. Mohamoud, and M. Maris, "An implementation of anger detection in speech signals," *2008 IET 4th Int. Conf. on Intelligent Environments*, Seattle, USA, July 21-22, 2008, pp. 1-4.
- [5] D. Ververidis and C. Kotropoulos, "Fast sequential floating forward selection applied to emotional speech features estimated on DES and SUSAS data collections," in *Proc. XIV European Signal Processing Conference*, Florence, Italy, Sep. 4-8, 2006.
- [6] J. Sidorova, "Speech emotion recognition with TGI+.2 classifier," in *Proc. EACL (Student Research Workshop)*, Athens, Greece, April 2, 2009, pp.54-60.
- [7] C.-H. Wu and W.-B. Liang, "Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels," *IEEE Trans. Affective Computing*, vol. 2, no. 1, pp. 10-21, 2011.
- [8] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using GMMs," in *Proc. INTERSPEECH 2006 - ICSLP. 9th Int. Conf. on Spoken Language Processing*, Pittsburgh, PA, USA, Sep. 17-21, 2006.
- [9] O. W. Kwon, K. Chan, J. Hao, and T. Lee, "Emotion recognition by speech signals," in *Proc. 8th European Conference on Speech Communication and Technology, EUROSPEECH 2003 - INTERSPEECH 2003*, Geneva, Switzerland, Sep. 1-4, 2003, pp. 125-128.
- [10] H. Altun and G. Polat, "Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection," *presented at Expert Syst. Appl.*, pp. 8197-8203, 2009.

- [11] B. Vlasenko and A. Wendemuth, "Tuning hidden Markov model for speech emotion recognition," in *Proc. 33rd German Annual Conference on Acoustics*, Stuttgart, Germany, March 2007, pp. 317-320.
- [12] T. Iliou and G. Paschalidis, "Using an automated speech emotion recognition technique to explore the impact of bullying on pupils social life," *2011 Panhellenic Conference on Informatics*, Kastonia, Greece, Sept. 30-Oct. 2, 2011, pp. 18-22.
- [13] W.-J. Yoon and K.-S. Park, "Building robust emotion recognition system on heterogeneous speech databases," *IEEE Trans. Consumer Electronics*, vol. 57, no. 2, pp. 747-750, 2011.
- [14] K. R. Scherer, "Vocal communication of emotion: a review of research paradigms," *Speech Communication*, vol. 40, no. 1-2, pp. 227-256, 2003.
- [15] C. Lili, J. Chunhui, W. Zhiping, Z. Li, and Z. Cairong, "A method combining the global and time series structure features for emotion recognition in speech," in *Proc. 2003 Int. Conf. on Neural Networks and Signal Processing*, Nanjing, China, Dec. 14-17, 2003, vol. 2, pp. 904-907.
- [16] I. Luengo, E. Navas, I. Hernez, and J. Snchez, "Automatic emotion recognition using prosodic parameters," in *proc. INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, Sep. 4-8, 2005, pp.493-496.
- [17] J.-S. Park, J.-H. Kim, and Y.-H. Oh, "Feature vector classification based speech emotion recognition for service robots," *IEEE Trans. Consumer Electronics*, vol. 55, no. 3, pp. 1590-1596, 2009.
- [18] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," *6th Int. Conf. on Neural Information Processing*, Perth, WA, Australia, 1999, vol. 2, pp. 495-501.
- [19] L. He, M. Lech, N. Maddage, S. Memon, and N. Allen, "Emotion recognition in spontaneous speech within work and family environments," *3rd Int. Conf. on Bioinformatics and Biomedical Engineering*, Beijing, China, June 11-13, 2009, pp. 1-4.
- [20] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier and B. Weiss, "A database of German emotional speech," in *Proc. 9th Int. Conf. on Spoken Language Processing (INTERSPEECH-SLP 2005)*, ISCA, Lisbon, Portugal, Sep. 4-8, 2005, pp. 1517-1520.
- [21] B. Davis, L. Sheeber, H. Hops, and E. Tildesley, "Adolescent responses to depressive parental behaviors in problem-solving interactions: implications for depressive symptoms," *Journal of Abnormal Child Psychology*, vol. 28, no. 5, pp. 451-465, 2000.
- [22] S. Emerich, E. Lupu, and A. Apatean, "Bimodal approach in emotion recognition using speech and facial expressions," *International Symposium on Signals, Circuits and Systems*, Iasi, Romania, July 9-10, 2009, pp. 1-4.
- [23] S. Pathak and A. Kulkarni, "Recognizing emotions from speech," *3rd Int. Conf. on Electronics Computer Technology*, Kanyakumari, India, April 8-10, 2011, vol. 4, pp. 107-109.
- [24] W. Kim and J. H. L. Hansen, "Angry emotion detection from real-life conversational speech by leveraging content structure," *2010 IEEE Int. Conf. on Acoustics Speech and Signal Processing*, Dallas, TX, USA, March 14-19, 2010, pp. 5166-5169.
- [25] L. He, M. Lech, N. Maddage, and N. Allen, "Emotion recognition in speech of parents of depressed adolescents," *3rd Int. Conf. on Bioinformatics and Biomedical Engineering*, Beijing, China, June 11-13, 2009, pp. 1-4.
- [26] S. Haq, P. J. B. Jackson, and J. Edge, "Audio-visual feature selection and reduction for emotion classification," in *Proc. Int. Conf. on Auditory-Visual Speech Processing*, Queensland, Australia, Sep. 26-29, 2008, pp. 185-190.
- [27] T. L. Nwe, S. Foo, and L. De Silva, "Speech emotion recognition using hidden Markov models," *Speech communication*, vol. 41, pp. 603-623, 2003.

- [28] T. Yamada, H. Hashimoto, and N. Tosa, "Pattern recognition of emotion with neural network," in *Proc. 1995 IEEE IECON 21st International Conference on Industrial Electronics, Control, and Instrumentation*, Orlando, FL, USA, Nov. 6-10, 1995, vol.1, pp. 183-187.
- [29] T. Pao, Y. Chen, J. Yeh and J. Burns, "Emotion recognition and evaluation of Mandarin speech using weighted D-KNN classification," *Intl. J. Innovation Computing, Information and Control*, vol. 4, pp. 1695-1709, 2008.
- [30] V. A. Petrushin, "Emotion recognition in speech signal: experimental study, development and application," in *Proc. Int. Conf. on Spoken Language Processing*, Beijing, China, Oct. 16-20, 2000, pp. 222-225.
- [31] Y. Kao and L. Lee, "Feature analysis for emotion recognition from Mandarin speech considering the special characteristics of Chinese language," in *Proc. INTERSPEECH 2006 - 9th Int. Conf. on Spoken Language Processing*, Pittsburgh, USA, Sep. 17-21, 2006.
- [32] Z.-Y. Han, S.-X. Lun, and J. Wang, "A study on speech emotion recognition based on CCBC and Neural Network," *2012 Int. Conf. on Computer Science and Electronics Engineering*, Hangzhou, China, March 23-25, 2012, pp. 144-147.
- [33] F. Mokhayeri, M.-R. Akbarzadeh-T, and S. Toosizadeh, "Mental stress detection using physiological signals based on soft computing techniques," *18th Iranian Conference on BioMedical Engineering*, Tehran, Iran, Dec. 14-16, 2011, pp.232-237.
- [34] H. Costin, C. Rotariu, and A. Pasarica, "Mental stress detection using heart rate variability and morphologic variability of ECG signals," *2012 International Conference and Exposition on Electrical and Power Engineering*, Iasi, Romania, Oct. 25-27, 2012, pp. 591-596.
- [35] J. Choi and R. Gutierrez-Osuna, "Estimating mental stress using a wearable cardio-respiratory sensor," *Sensors, 2010 IEEE*, Kona, HI, USA, Nov. 1-4, 2010, pp. 150-154.
- [36] J. Wijsman, B. Grundlehner, H. Liu, Hermie Hermens, and Julien Penders, "Towards mental stress detection using wearable physiological sensors," *33rd Annual Int. Conf. of the IEEE EMBS*, Boston, Massachusetts, USA, Aug. 30-Sep. 3, 2011, pp. 1798-1801.
- [37] J. Choi and R. Gutierrez-Osuna, "Using heart rate monitors to detect mental stress," *Sixth International Workshop on Wearable and Implantable Body Sensor Networks, BSN 2009*, Berkeley, CA, USA, June 3-5, 2009, pp. 219-223.
- [38] Y.-H. Chen, H.-H. Chen, T.-C. Chen, and Liang-Gee Chen, "Robust heart rate measurement with Phonocardiogram by on-line template extraction and matching," *33rd Annual Int. Conf. IEEE Eng. in Medicine and Biology Society*, Boston, USA, Aug. 30-Sep. 3, 2011, pp. 1957-1960.
- [39] J. Choi and R. Gutierrez-Osuna, "Removal of respiratory influences from heart rate variability in stress monitoring," *IEEE Sensors Journal*, vol. 11, no. 11, pp. 2649-2656, 2011.
- [40] C. Ahlstrom, "Nonlinear Phonocardiographic signal processing," Ph.D. Project, Biomed. Eng. Dept., Linkoping University, Linkoping, Sweden, 2008.
- [41] K. H. Shelley and S. Shelley, "Pulse oximeter waveform: photoelectric Plethysmography," *Clinical Monitoring: Practical applications for anesthesia and critical care*, Ed Lake, C. W. B. Saunders Company, pp. 420-428, 2001.
- [42] A. T. Reisner, P. A. Shaltis, D. McCombie, and H. H. Asada, "Utility of the Photoplethysmogram in circulatory monitoring," *Anesthesiology*, vol. 108, pp. 950-958, 2008.
- [43] A. Sorvala, E. Alasaarela, H. Sorvoja, and R. Myllyla, "Activity classification using a state transition diagram and activity levels," *2012 6th International Symposium on Medical Information and Communication Technology*, La Jolla, CA, USA, March 25-29, 2012, pp. 1-4.
- [44] A. Henpraserttae, S. Thiemjarus, and S. Marukatat, "Accurate activity recognition using a mobile phone regardless of device orientation and location," *2011 Int. Conf. on Body Sensor Networks*, Dallas, TX, USA, May 23-25, 2011, pp. 41-46.

- [45] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis," *2010 5th Int. Conf. on Future Information Technology*, Busan, Korea, May 21-23, 2010, pp. 1-6.
- [46] G. Hache, E. D. Lemaire, and N. Baddour, "Wearable mobility monitoring using a multimedia smartphone platform," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 9, pp. 3153-3161, 2011.
- [47] G. Hache, E. D. Lemaire, and N. Baddour, "Mobility change-of-state detection using a smartphone-based approach," *2010 IEEE International Workshop on Medical Measurements and Applications Proceedings*, Ottawa, ON, Canada, April 30-May 1, 2010, pp. 43-46.
- [48] M. Muehlbauer, G. Bahle, and P. Lukowicz, "What can an arm holster worn smart phone do for activity recognition?," *2011 15th Annual International Symposium on Wearable Computers*, San Francisco, CA, USA, June 12-15, 2011, pp. 79-82.
- [49] R. K. Ganti, S. Srinivasan, and A. Gacic, "Multisensor fusion in smartphones for lifestyle monitoring," *2010 Int. Conf. Body Sensor Networks*, Singapore, Singapore, June 7-9, 2010, pp. 36-43.
- [50] S.-K. Song, J. Jang, and S. Park, "A phone for human activity recognition using triaxial acceleration sensor," *Int. Conf. on Consumer Electronics 2008, Digest of Technical Papers*, Las Vegas, NV, USA, Jan. 9-13, 2008, pp. 1-2.
- [51] C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 3, pp. 136-147, 1997.
- [52] J. Baek, G. Lee, W. Park, and B.-J. Yun, "Accelerometer signal processing for user activity detection," *Lecture Notes in Computer Science*, pp. 610-617, 2004.
- [53] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 156-167, 2006.
- [54] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 711-723, 2003.
- [55] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 20-26, 2008.
- [56] A. Sorvala, E. Alasaarela, H. Sorvoja, and R. Myllyla, "A two-threshold fall detection algorithm for reducing false alarms," *2012 6th International Symposium on Medical Information and Communication Technology*, La Jolla, CA, USA, March 25-29, 2012, pp. 1-4.
- [57] Z. A. Khan and Won Sohn, "Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care," *IEEE Trans. Consumer Electronics*, vol.57, No.4, pp. 1843-1850, 2011.