Physical Violence Detection with Movement Sensors

Liang Ye[1,2,*], Le Wang[1], Peng Wang[1,3], Hany Ferdinando[2,4], Tapio Seppänen[5], and Esko Alasaarela[2]

1 Communication Research Center, Harbin Institute of Technology, Harbin 150080, China
2 Optoelectronics and Measurement Techniques Laboratory, Department of Electrical Engineering, University of Oulu, Oulu 90570, Finland
3 China Electronics Technology Group Corporation, Nanjing 210012, China
4 Department of Electrical Engineering, Petra Christian University, Surabaya 60236, Indonesia
5 Department of Computer Science and Engineering, University of Oulu, Oulu 90570, Finland
E-mails: yeliang@hit.edu.cn, 1659412561@qq.com, wphitstudent@163.com, hferdina@ee.oulu.fi, tapio@ee.oulu.fi, and esko.alasaarela@ee.oulu.fi

Abstract. With the development of movement sensors, activity recognition becomes more and more popular. Compared with daily-life activity recognition, physical violence detection is more meaningful and valuable. This paper proposes a physical violence detecting method. Movement data of acceleration and gyro are gathered by role playing of physical violence and daily-life activities. Time domain features and frequency domain ones are extracted and filtered to discribe the differences between physical violence and daily-life activities. A specific BPNN trained with the L-M method works as the classifier. Altogether 9 kinds of activities are involved. For 9-class classification, the average recognition accuracy is 67.0%, whereas for 2-class classification, i.e. activities are classified as violence or daily-life activity, the average recognition accuracy reaches 83.7%.

Keywords: physical violence detection, activity recognition, movement sensor

1 Introduction

In recent years, movement sensor techniques have developed very rapidly. Benefit from this, activity recognition based on movement data becomes more and more popular. Commonly used movement sensors are accelerometers and gyroscopes, and the corresponding data are acceleration and gyro, respectively.

adfa, p. 1, 2011.

Existing activity recognition research work mainly focuses on the recognition of daily-life activities. For example, Cheng, *et al*. [1] recognized daily-life activities of standing, sitting, walking, turning left, turning right, going upstairs, going downstairs, jogging, and jumping. Nakano, *et al*. [2] recognized daily-life activities of walking, walking upstairs, walking downstairs, sitting, standing, and lying. Hegde, *et al*. [3] recognized daily-life activities of lying down, sitting, standing, walking driving, descend stairs, ascend stairs, and cycling.

In 2014, Alasaarela [4] argued that activity recognition with movement sensors can also be used for school violence detection, i.e. with wearable movement sensors such as smartphones embedded with accelerometers and gyroscopes, one can detect school violence events. As members of his research group, Ye, *et al*. [5] designed their first experimental classifier FMT (Fuzzy Multi-Threshold) with some simple activities. Later they [6] involved more kinds of activities and different ages of actors and actresses, and designed a more compatible Instance-Based classifier. Besides movement features, physiological features such as ECG (electrocardiogram) [7, 8] and HRV (heart rate variability) [9] were also used for school violence detection, but they are not considered in this paper.

As a continuation, this paper improves the authors' previous work by involving more features and designing a more proper classifier. Besides, more kinds of activities are tested compared with the authors' previous work. The remainder of this paper is organized as follows: Section 2 describes the extracted movement features; Section 3 describes the classifier and the training method; Section 4 shows the simulation results; Section 5 draws a conclusion.

2 Movement features extraction

Movement data of physical violence and daily-life activities are gathered by role playing. Nine kinds of activities are acted, namely beating, pushing, pushing down, walking, running, jumping, falling down, playing, and standing. The first three kinds of activities are physical violence events, and the remaining six are daily-life activities. A Butterworth filter is used before feature extraction to remove high frequency jitters.

The authors' previous work [5, 6] only extracted time domain features of the activities, but this paper extract both time domain features and frequency domain ones. The extracted time domain features are given in Table 1, whereas the extracted frequency domain features are shown in Table 2.

In this experiment, the *y*-axis is the vertical direction, so the horizontal combined vector means the combination of the *x*-axis and the *z*-axis. The combined vector means the combination of all the three axes. MAD is the Median Absolute Deviation, and MAD=median($|x_i$-median$(X)|$), where $X=\{x_1, x_2, ..., x_n\}$. *VarDir* describes the change of horizontal movement direction, and $Area_y$ is the accumulation of movement jitter in the vertical direction [6].

Table 1. Extracted time domain features

Feature	Meaning	From
$Mean_y$	Mean of the *y*-axis	Acceleration
$Mean_{Hori}$	Mean of the horizontal combined vector	Acceleration
$Mean_{Gyro}$	Mean of the combined gyro	Gyro
MAD_y	MAD of the *y*-axis	Acceleration
MAD_{Hori}	MAD of the horizontal combined vector	Acceleration
MAD_{Gyro}	MAD of the combined gyro	Gyro
Max_y	Maximum of the *y*-axis	Acceleration
Max_{Hori}	Maximum of the horizontal combined vector	Acceleration
Max_{Gyro}	Maximum of the combined gyro	Gyro
Min_y	Minimum of the *y*-axis	Acceleration
Min_{Hori}	Minimum of the horizontal combined vector	Acceleration
Min_{Gyro}	Minimum of the combined gyro	Gyro
$Max_{diff(y)}$	Maximum of the differential of the *y*-axis	Acceleration
$Max_{diff(Hori)}$	Maximum of the differential of the horizontal combined vector	Acceleration
$Mean_{diff(y)}$	Mean of the differential of the *y*-axis	Acceleration
$Mean_{diff(Hori)}$	Mean of the differential of the horizontal combined vector	Acceleration
$Max_{diff(Gyro)}$	Maximum of the differential of the combined gyro	Gyro
$Mean_{diff(Gyro)}$	Mean of the differential of the combined gyro	Gyro
ZCR_x	Zero cross rate of the *x*-axis	Acceleration
ZCR_y	Zero cross rate of the *y*-axis	Acceleration
ZCR_z	Zero cross rate of the *z*-axis	Acceleration
$VarDir$	Variation of the horizontal movement direction	Acceleration
$Area_y$	Accumulation of movement jitter of the *y*-axis	Acceleration

Table 2. Frequency domain features

Feature	Meaning	From
Max_{fy}	Maximum of the *y*-axis	Acceleration
Max_{fHori}	Maximum of the horizontal combined vector	Acceleration
Max_{fGyro}	Maximum of the combined gyro	Gyro
Min_{fy}	Minimum of the *y*-axis	Acceleration
Min_{fHori}	Minimum of the horizontal combined vector	Acceleration
Min_{fGyro}	Minimum of the combined gyro	Gyro
MAD_{fy}	MAD of the *y*-axis	Acceleration
MAD_{fHori}	MAD of the horizontal combined vector	Acceleration
MAD_{fGyro}	MAD of the combined gyro	Gyro
$Mean_{fy}$	Mean of the *y*-axis	Acceleration
$Mean_{fHori}$	Mean of the horizontal combined vector	Acceleration
$Mean_{fGyro}$	Mean of the combined gyro	Gyro
$Energy_{fy}$	Energy of the *y*-axis	Acceleration
$Energy_{fHori}$	Energy of the horizontal combined vector	Acceleration
$Energy_{fGyro}$	Energy of the combined gyro	Gyro
$Center_{fHori}$	Main lob center frequency of the horizontal combined vector	Acceleration
$Center_{fy}$	Main lob center frequency of the *y*-axis	Acceleration
$Center_{fGyro}$	Main lob center frequency of the combined gyro	Gyro

The frequency domain features are extracted by FFT (Fast Fourier Transform). The maximum or minimum of the frequency means the frequency with the maximum or minimum amplitude. The horizontal combined vector and the combined vector have the same meanings with those of the time domain feature vectors.

There are altogether 41 features (23 time domain features and 18 frequency domain features) extracted for classification. However, since the authors' purpose is to apply the physical violence detecting algorithm on a smartphone for pratical use, the dimension of the feature vector should be as low as possible. In this paper, the authors choose the Wrapper method [10] for feature selection.

The authors firstly designed the entire classification system, including data gathering, data pre-processing, feature extraction, and the classifier. A specific BPNN (Back Propagation Neural Network) [11] works as the classifier. Then the authors use this system to find out the best feature combination with the Wrapper method. In each iteration, Wrapper adds or removes features, and compares the classification results. Finally, Wrapper selects 11 features, including 7 time domain features, namely $Mean_{Gyro}$, Max_{Gyro}, $Max_{diff}(y)$, $Max_{diff(Gyro)}$, ZCR_x, ZCR_y, and $VarDir$, and 4 frequency domain features, namely MAD_{fHori}, MAD_{fGyro}, $Mean_{fHori}$, and $Energy_{fy}$.

3 Classifier design

In the authors' previous work, the first classifier FMT was discarded in the second experiment because it was difficult to find unified thresholds for the actors and actresses of different ages due to strength difference. However, the second Instance-Based classifier could not distinguish the activities of pushing down and falling down very well. Therefore, the authors decide to find a more proper classifier. By comparing the advantages and disadvantages of some commonly used classifiers, e.g. Bayesian, SVM, KNN and KNN-based, the authors finally choose BPNN for this work. BPNN is particularly suitable for solving complex problems with non-linear relationship between the extracted features and classification results. Fig. 1 shows the architecture of a classical BPNN, where p is the input, w is the weight, b is the bias, f is the transfer function, and a is the output.

The parameter setting of BPNN is a key factor which affects the classification performance. Normally, the number of inputs of the network equals to the dimension of the input feature vector. In this paper, it is 11. The number of neurons of the output layer equals to the number of target classes, i.e. 9 in this paper. The number of neurons of the hidden layer should be set larger than the square of the sum of the input and the output dimensions empirically. Other parameters, e.g. the number of hidden layers, the transfer functions of the hidden layers and the output layer, will be set experimentally according to the simulations.

For training BPNN, this paper chooses the Lenvenberg-Marquardt (L-M) method [12]. Compared with other training methods such as the Newton method and the Gradient Descent method, the L-M method can avoid calculating a Hessian matrix which will cause large computational cost. Instead, the Hessian matrix is approximated by

$\mathbf{H}=\mathbf{J}^T\mathbf{J}$ with the gradient $\mathbf{g}=\mathbf{J}^T\mathbf{e}$ where **J** is a Jacobian matrix containing the first order derivative of the training error **e**. In each iteration, $x_{k+1} = x_k - [\mathbf{J}^T\mathbf{J}+\mu\mathbf{I}]^{-1}\mathbf{J}^T\mathbf{e}$.

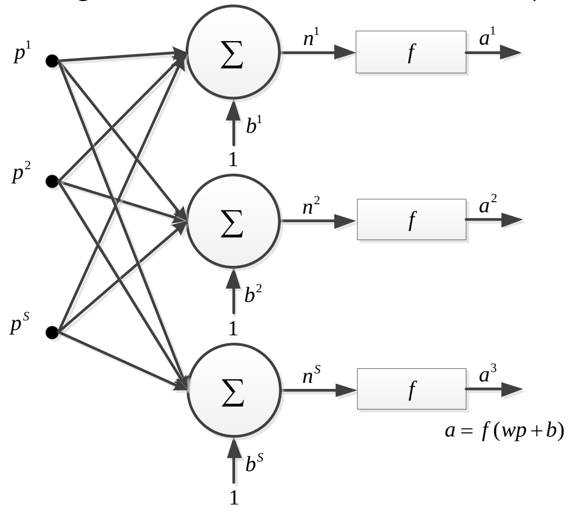

Fig. 1. Architecture of a classical BPNN

4 Simulations

The authors recorded altogether 1160 fragments of the 9 kinds of activities, and the amount of each kind was similar. The movement sensors (accelerometers and gyroscopes) were fixed on the actors' and actresses' waists. Ten-fold cross validation was used for the simulations, i.e. the authors split each kind of activity into 10 groups, 9 of which were used as the training set whereas the remaining 1 as the testing set. Repeat this procedure 10 times and change the testing set each time. The final result was the average of the 10 results.

Then the authors set the parameters of BPNN experimentally: there are 6 neurons in the hidden layer; the transfer function of the hidden layer is logsig, whereas that of the output layer is purelin. So the architecture of this specific BPNN was,

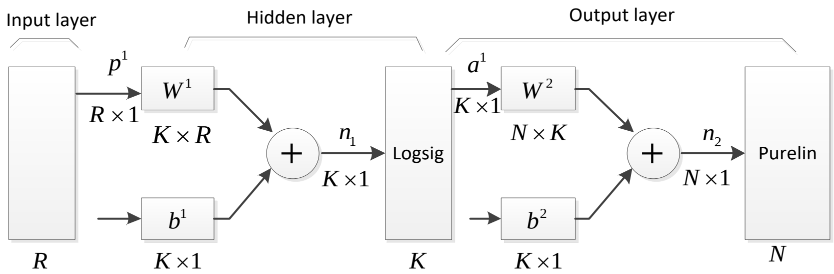

Fig. 2. Architecture of the employed BPNN

Firstly, the 9 kinds of activities were classified into 9 classes. The confusion matrix is given in Table 3.

Table 3. Confusion matrix of 9-class classification (%)

Classified as	Beat	Push	Push down	Walk	Run	Jump	Fall down	Play	Stand
Beat	**50.0**	10.0	13.3	0.0	3.3	0.0	3.3	20.0	0.0
Push	8.3	**68.3**	8.3	0.0	1.7	6.7	6.7	0.0	0.0
Push down	0.0	3.3	**40.0**	10.0	0.0	3.3	10.0	23.3	10.0
Walk	2.5	1.3	7.5	**43.8**	0.0	0.0	0.0	42.5	2.5
Run	0.0	1.4	0.0	0.0	**91.4**	5.7	0.0	0.0	1.4
Jump	0.0	0.0	0.0	0.0	12.5	**87.5**	0.0	0.0	0.0
Fall down	0.0	20.0	6.7	0.0	0.0	0.0	**66.7**	6.7	0.0
Play	4.2	5.0	15.0	5.8	0.0	2.5	4.2	**60.8**	2.5
Stand	0.0	0.0	1.1	1.1	0.0	0.0	0.0	3.3	**94.4**

Then, considering the theme of this paper, i.e. physical violence detection, the 9 kinds of activities are classified into 2 classes, namely physical violence and daily-life activities, respectively. Table 4 shows the 2-class classification confusion matrix.

Table 4. Confusion matrix of 2-class classification (%)

Classified as	Physical violence	Daily-life activity
Physical violence	**71.7**	28.3
Daily-life activity	11.2	**88.8**

Table 4 shows an intuitional result of the proposed violence detecting method, and Table 3 may tell some details of it. Violent activities of beating and pushing down are likely to be misclassified as the daily-life activity of playing, and daily-life activities of falling down and playing are likely to be misclassified as violenct activities of pushing and pushing down, respectively, because these kinds of activities have similar strenghth and suddenness especially when playing contains confrontational actions.

The average accuracy of the proposed method is about 83.7%, which outperforms the authors‘ previous Instance-Based method by 3.7%. The first method FMT is un-comparable because it is hardly possible to find unified thresholds for actors and actresses of different ages due to strength difference.

5 Conclusion

This paper proposed a physical violence detecting method. The movement data were gathered by movement sensors by means of role playing. Both time domain features and frequency domain features were extracted and filtered to describe the differences between physical violence and daily-life activities. A specific BPNN trained with the L-M method worked as the classifier. Ten-fold cross validation was performed for simulation. The average classification accuracy was 83.7%, which showed an improvement compared with the authors’ previous work.

Acknowledgements

This paper was supported by the National Natural Science Foundation of China (61602127), and partly supported by the Directorate General of Higher Education, Indonesia (2142/E4.4/K/2013), and the Finnish Cultural Foundation, North Ostrobothnia Regional Fund.

The authors would like to thank Tuija Huuki, Vappu Sunnari, Seppo Laukka and Antti Siipo from University of Oulu, Finland, teachers Taina Aalto and Pekka Kurttila and principal Maija Laukka from Oulunlahti School, Finland, pupils from 2nd and 6th grades of Oulunlahti School, Tian Han and Zhu Zhang from Harbin University of Science and Technology, China, Yubo Zhang, Jifu Shi and Zhi Xun from Harbin Institute of Technology, China for their assistance to this work.

References

1. Long Cheng, Yiyi Yu, Xinyang Liu, Jinyu Su, and Yani Guan. Recognition of Human Activities Using Fast and Adaptive Sparse Representation Based on Wearable Sensors. 2017 16th IEEE International Conference on Machine Learning and Applications, 2017, pp. 944-949.
2. Kotaro Nakano, and Basabi Chakraborty. Effect of Dynamic Feature for Human Activity Recognition Using Smartphone Sensors. 2017 IEEE 8th International Conference on Awareness Science and Technology (ICAST), 2017, pp. 539-543.
3. Nagaraj Hegde, Matthew Bries, Tracy Swibas, *et al*. Automatic Recognition of Activities of Daily Living utilizing Insole Based and Wrist Worn Wearable Sensors. IEEE Journal of Biomedical and Health Informatics, 2018 (99), pp. 1-9.
4. Liang Ye, Hany Ferdinando, and Esko Alasaarela. Techniques in Pattern Recognition for School Bullying Prevention: Review and Outlook. Journal of Pattern Recognition Research, 2014, 9(1), pp. 50-63.
5. Liang Ye, Hany Ferdinando, Tapio Seppänen, and Esko Alasaarela. Physical Violence Detection for School Bullying Prevention. Advances in Artificial Intelligence, 2014, vol. 2014, pp. 1-9.
6. Liang Ye, Hany Ferdinando, Tapio Seppänen, *et al*. An Instance-Based Physical Violence Detection Algorithm for School Bullying Prevention. 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), Dubrovnik, Croatia, 2015, pp. 1384-1388.
7. Hany Ferdinando, Liang Ye, Tian Han, *et al*. Violence Detection from ECG Signals: A Preliminary Study. Journal of Pattern Recognition Research, 2017, 12(1), pp. 7-18.
8. Hany Ferdinando, Tapio Seppänen, and Esko Alasaarela. Enhancing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction. Proceeding of 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM) 2017, February 24-26, 2017, Porto, Portugal, pp. 112-118.
9. Hany Ferdinando, Liang Ye, Tapio Seppänen, *et al*. Emotion Recognition by Heart Rate Variability. Australian Journal of Basic and Applied Sciences, 2014, 8(10), pp. 50-55.
10. Naoual El Aboudi, and Laila Benhlima. Review on Wrapper Feature Selection Approaches. 2016 International Conference on Engineering & MIS (ICEMIS), 2016, pp. 1-5.

11. Zhuhui Huang, and Wei Wang. Multi-sensor Fusion Based on BPNN in Quadruped Ground Classification. 2017 IEEE International Conference on Mechatronics and Automation (ICMA), 2017, pp. 1620-1625.
12. Wei Bengang, Wu Xinye, Yao Zhoufei, *et al*. A method of optimized neural network by L-M algorithm to transformer winding hot spot temperature forecasting. 2017 IEEE Electrical Insulation Conference (EIC), 2017, pp. 87-91.