LETTER

A Novel Bimodal Emotion Database from Physiological Signals and Facial Expression*

Jingjie YAN[†a)], **Bei WANG**[††], *Nonmembers, and* **Ruiyu LIANG**[†††], *Member*

SUMMARY In this paper, we establish a novel bimodal emotion database from physiological signals and facial expression, which is named as PSFE. The physiological signals and facial expression of the PSFE database are respectively recorded by the equipment of the BIOPAC MP 150 and the Kinect for Windows in the meantime. The PSFE database altogether records 32 subjects which include 11 women and 21 man, and their age distribution is from 20 to 25. Moreover, the PSFE database records three basic emotion classes containing calmness, happiness and sadness, which respectively correspond to the neutral, positive and negative emotion state. The general sample number of the PSFE database is 288 and each emotion class contains 96 samples.

key words: *bimodal emotion database, physiological signals, facial expression, PSFE*

1. Introduction

In the research domain of affective computing and computer vision, the research of facial expression recognition, speech emotion recognition and physiological signals emotion recognition which are only based on one modality is more and more replaced by the bimodal emotion recognition and multimodal emotion recognition which are based on two modalities, three modalities or even more modalities in recent years [1]–[8]. No matter for facial expression recognition, speech emotion recognition, physiological signals emotion recognition or bimodal emotion recognition and multimodal emotion recognition, the establishment of the corresponding emotion database is the most primary and important step [9]. We can not conduct the emotion recognition research smoothly and deeply in case of lacking the dependable emotion database [10]–[14].

Manuscript received December 2, 2017.
Manuscript revised March 11, 2018.
Manuscript publicized April 17, 2018.

[†]The author is with the Jiangsu Provincial Key Laboratory of Image Processing and Image Communication, College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China.

[††]The author is with the Key Laboratory of Child Development and Learning Science (Ministry of Education), Research Center for Learning Science, Southeast University, Nanjing 210096, China.

[†††]The author is with the School of Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China.

*This work was partly supported by the National Natural Science Foundation of China (NSFC) under Grants 61501249 and 41601601, the Natural Science Foundation of Jiangsu Province under Grant BK20150855, the Research Foundation of The Ministry of Education and China Mobile under Grant MCM20150504 and the NUPTSF under Grant NY214143.

a) E-mail: yanjingjie1212@163.com

DOI: 10.1587/transinf.2017EDL8264

As so far, many emotion databases which are only based on one modality such as the JAFFE facial expression database [2], [15], [16], the Berlin speech emotion database [17], [18] and the Multi-PIE facial expression database [19], [20] have already been established over the last two decades. Base on those emotion databases using one modality, the research of facial expression recognition, speech emotion recognition and physiological signals emotion recognition all become the research priorities in affective computing domain [2]–[4], [18], [21]–[25].

In the wake of the increase of the emotion database based on one modality, a few researchers attempt to focus on the establishment of bimodal emotion database and multimodal emotion database [9], [26]–[29]. Compared to the emotion database based on one modality, establishing a dependable bimodal emotion database and multimodal emotion database is more difficult and complicated, because the channel of the emotion data recording is increased to two, three or even more, and the data recording of each channel should be coinstantaneous [9], [26], [27]. As so far, a few effective bimodal emotion databases and multimodal emotion databases such as FABO and DEAP have been established on the basis of those emotion databases using one modality in last several years [9], [26]–[34].

Although a few bimodal emotion databases and multimodal emotion databases have been established, the bimodal emotion database from physiological signals and facial expression is also rare, and the existing bimodal emotion databases and multimodal emotion databases are mainly based on europeans and americans, and there is rarely bimodal emotion database from physiological signals and facial expression which is based on asian. Therefore, in this paper, we establish a novel bimodal emotion database from physiological signals and facial expression, which is named as PSFE and is based on asian. The physiological signals and facial expression of the PSFE database are respectively recorded by the equipment of the BIOPAC MP 150 [35], [36] and the Kinect for Windows [14], [37] in the meantime. The PSFE database altogether records 32 subjects and records three basic emotion classe containing calmness, happiness and sadness, which respectively correspond to the neutral, positive and negative emotion state.

2. Collecting Device of the PSFE Database

In our PSFE database, we utilize the equipment of Kinect for Windows to collect emotion data of facial expression [14],

Fig. 1 The picture of the equipment of the Kinect for Windows.

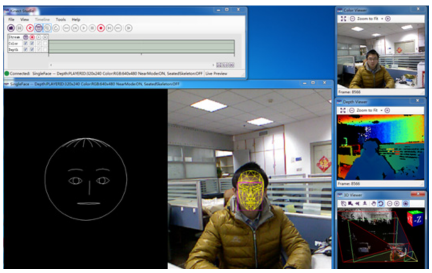

Fig. 2 The picture of collecting window based on the Kinect for Windows SDK.

[37]. The Kinect for Windows is a effective video collecting equipment which is created by the Microsoft company. It contains three cameras which can record two types of video data. The detailed hardware parameters of the Kinect for Windows and the related parameters of the collected video can see [14], [37]. Figure 1 is the picture of the equipment of the Kinect for Windows. Moreover, the Kinect for Windows also provides the Kinect for Windows SDK [38] to deal with the collected videos. Moreover, the Kinect for Windows (as shown in Fig. 2) also provides the Kinect for Windows SDK [38] to deal with the collected videos.

Moreover, we utilize the equipment of BIOPAC MP 150 [14], [35], [36] to collect emotion data of physiological signals. The BIOPAC MP 150 is a effective physiological signals collecting equipment which is created by the BIOPAC company. It can record multiple physiological signals such as electrocardiosignal (ECG), skin impedance signal and temperature signal [4], [14], [30], [31], [35], [36]. In our experiment, we record the ECG signal as physiological signals. Because the ECG signal is one of the most important physiological signals which can obviously reflect the variation of emotion according to the literature of [34]. The facial expression is also a important way to reflect the variation of emotion. So the ECG signal and facial expression are relevant and also may be complementary. The detailed hardware parameters of BIOPAC MP 150 and the introduction of ECG, skin impedance signal and temperature signal can see [14], [37]. Moreover, the BIOPAC company also provides the corresponding collecting software of AcqKnowledge 4.1 [35], [36] to deal with the collected physiological signals. How to use the AcqKnowledge 4.1 can see the website of [35], [36].

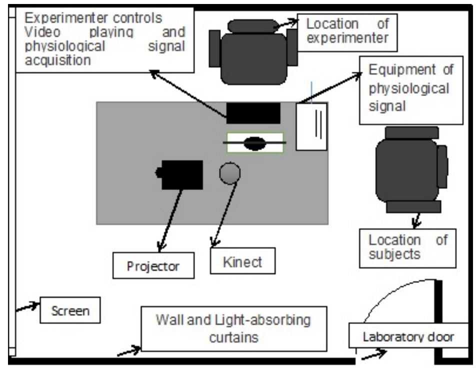

Fig. 3 The environment arrange of the laboratory in detail.

3. Collecting Experiment Design of the PSFE Database

The PSFE database is recorded in the closed laboratory to avoid various interferences. To reduce the affect of illumination, the wall of the laboratory hangs window curtain which can absorb light. The environment arrange of the laboratory in detail is shown in Fig. 3.

We select 32 experiment subjects which include 11 women and 21 man from the undergraduate and graduate student in our college, and their age distribution is from 20 to 25. In the experiment, we utilize three types of movie clip as the method of emotion elicitation [4], [14], [30], [31], [39]. we respectively select the happy movie clip, the sad movie clip and the calm movie clip to elicit the accordingly positive emotion, negative emotion and neutral emotion. The time duration of each movie clip is about three minute. In the procedure of collecting emotion data, each movie clip is projected to the wall and the corresponding audio is emitted by the sound equipment [14], [39].

4. Collecting Result of the PSFE Database

For the ECG data, as the voltage amplitude of ECG between different individuals exits outstanding difference, so the original ECG data has to be normalized to reduce the experiment effect by the different individuals.

Before handling the emotion data, taking into consideration the emotion elicitation needs some times, so we cut off the data of the front 30 seconds of each facial expression video and ECG signal. So we only keep the data of the remaining 120 seconds to 160 seconds. Moreover, we divide the remaining emotion data into three clips as the part of the initial emotion elicitation period, the interim emotion elicitation period and the later emotion elicitation period [14], [39]. The time duration of each period is about 40 seconds. For each emotion elicitation period, we respectively extract one key facial expression frame and a segment of ECG signal whose time duration is about 40 seconds to conduct the subsequent experiment.

Figure 4 is the facial expression data and the corresponding ECG data of the positive emotion (happiness). From left to right, three images respectively represent the

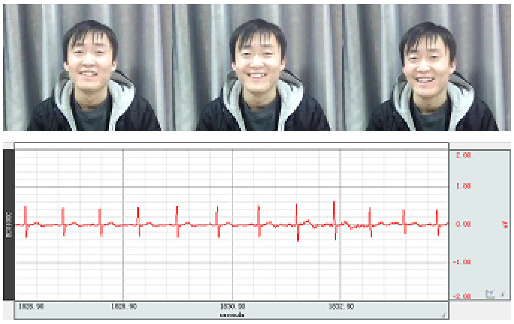

Fig. 4 The facial expression data and the corresponding ECG data of the positive emotion (happiness).

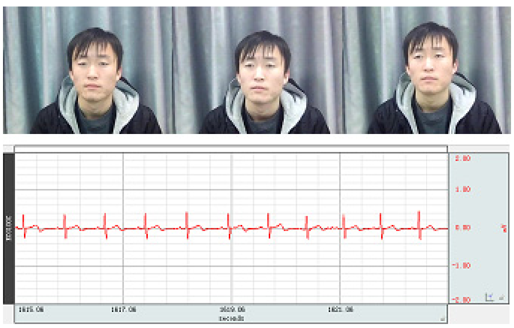

Fig. 5 The facial expression data and the corresponding ECG data of the neutral emotion (calmness).

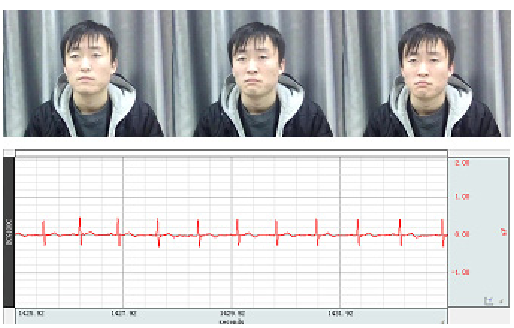

Fig. 6 The facial expression data and the corresponding ECG data of the negative emotion (sadness).

facial expression data of initial emotion elicitation period, the interim emotion elicitation period and the later emotion elicitation period. Moreover, a part of ECG signal is shown under the facial expression images. Figure 5 and Fig. 6 are the facial expression data and the corresponding ECG data of the neutral emotion (calmness) and the negative emotion (sadness) respectively.

The relevant information of the PSFE database is summarized in Table 1 (32 is the number of subjects). Moreover, we compare some other relevant multimodal databases to our PSFE database. The FABO [26] is based on facial expression and gesture,and has nine emotion categories and 23 subjects. The eNTERFACE 05 [29] covers facial expression and speech modality, and has six emotion categories and 42 subjects. The RML [28] is also covers facial expression and speech modality and six emotion categories, but its only has 8 subjects. The DEAP [30] is base on physiological signals and covers five emotion dimensions and 32 subjects. The MAHNOB-HCI [31] has the modality of facial expression, speech and physiological signals, and covers four emotion dimensions and 27 subjects. The GEMEP [33] covers facial expression, gesture and speech modality, and has 18 emotion categories and 10 subjects. The AV+EC [34] has facial expression, speech and physiological signals, and covers two emotion dimensions and 27 subjects. From the above presentation, we can see that the subjects of our PSFE database may be adequate to a certain degree compared to other multimodal databases. But the number of emotion categories of our PSFE database is fewer, and that is because the distinction of different emotion categories for physiological signals is hard compared to facial expression, speech and gesture. In future, we will attempt to add the subjects and emotion categories.

Table 1 The relevant information of the PSFE database.

Emotion category	The number of sample	man:woman
positive (happiness)	96 (32 × 3)	21:11
neutral (calmness)	96 (32 × 3)	21:11
negative (sadness)	96 (32 × 3)	21:11
In total	288 (32 × 3 × 3)	63:33

Table 2 The experiment result on the PSFE database.

Modality	ECG	Facial	Fusion
Recognition rate	49.63%	83.33%	84.44%

Moreover, we also do the recognition experiment on the PSFE database. The feature extraction of ECG and facial expression is based on the SIFT method [10], [14] and AuBT tool [40] respectively. The fusion strategy is based on the common decision-level fusion [14], [27], and the SVM classifier and 10-fold cross-validation [2], [13], [18] is used for the PSFE database. The experiment result on the PSFE database is shown in Table 2. From the result, we can see that the recognition rate after fusion is high compared with ECG and Facial modality.

5. Conclusions

In this paper, we establish a novel bimodal emotion database from physiological signals and facial expression, which is named as PSFE. The PSFE database altogether records 32 subjects which include 11 women and 21 man, and their age distribution is from 20 to 25. Moreover, the PSFE database records three basic emotion classes containing calmness, happiness and sadness, which respectively correspond to the neutral, positive and negative emotion state. The general sample number of the PSFE database is 288 and each emotion class contains 96 samples. In future, we will attempt to add the subjects and emotion categories.

References

[1] M.H. Siddiqi, R. Ali, A.M. Khan, Y.-T. Park, and S. Lee, “Human fa-

cial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields," IEEE Trans. Image Process., vol.24, no.4, pp.1386–1398, 2015.

[2] J. Yan, W. Zheng, M. Xin, and Y. Yan, "Facial expression recognition based on sparse locality preserving projection," IEICE Trans. Fundamentals, vol.E97-A, no.7, pp.1650–1653, 2014.

[3] M.E. Ayadi, M.S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol.44, no.3, pp.572–587, 2011.

[4] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review," In: Proceedings of the 7th IEEE International Colloquium on Signal Processing and its Applications (CSPA), pp.410–415, 2011.

[5] H. Gunes, M. Piccardi, and M. Pantic, "From the lab to the real world: affect recognition using multiple cues and modalities," Affective Computing: Focus on Emotion Expression, Synthesis, and Recognition, pp.185–218, Vienna: InTech Education and Publishing, 2008.

[6] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.1, pp.39–58, 2009.

[7] Z. Zeng, J. Tu, M. Liu, T.S. Huang, B. Pianfetti, D. Roth, and S. Levinson, "Audio-visual affect recognition," IEEE Trans. Multimedia, vol.9, no.2, pp.424–428, 2007.

[8] G.K. Verma and U.S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," NeuroImage, vol.102, pp.162–172, 2014.

[9] J. Yan, G. Lu, and H. Li, "A Novel Supervised Bimodal Emotion Recognition Approach Based on Facial Expression and Body Gesture," submitted to Mathematical Problems in Engineering, 2017.

[10] J. Yan, W. Zheng, Q. Xu, G. Lu, H. Li, and B. Wang, "Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech," IEEE Transactions on and Multi Media, vol.18, no.7, pp.1319–1329, 2016.

[11] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," Information Fusion, vol.37, pp.98–125, 2017.

[12] R.W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.10, pp.1175–1191, 2001.

[13] J. Yan, W. Zheng, M. Xin, and J. Yan, "Integrating facial expression and body gesture in videos for emotion recognition," IEICE Trans. Inf. & Syst., vol.E97-D, no.3, pp.610–613, 2014.

[14] B. Wang, A Study of Bimodal Emotion Recognition Based on Facial Expression and Biological Signals, Master's thesis, Southeast University, Nanjing, China, 2014.

[15] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (KCCA)," IEEE Trans. Neural Netw., vol.17, no.1, pp.233–238, 2006.

[16] M. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," IEEE Trans. Pattern Anal. Mach. Intell., vol.21, no.12, pp.1357–1362, 1999.

[17] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A Database of German Emotional Speech," Interspeech, pp.4–8, 2005.

[18] J. Yan, X. Wang, W. Gu, and L. Ma, "Speech emotion recognition based on sparse representation," Archives of Acoustics, vol.38, no.4, pp.465–470, 2013.

[19] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," Image and Vision Computng, vol.28, no.5, pp.807–813, 2010.

[20] W. Zheng, "Multi-view Facial expression recognition based on group sparse reduced-rank regression," IEEE Transactions on Affective Computing, vol.5, no.1, pp.71–85, 2014.

[21] S. Lalitha, A. Madhavan, B. Bhushan, and S. Saketh, "Speech emotion recognition," In: Proceedings of the IEEE International Conference on Advances in Electronics, Computers and Communications, pp.1–4, 2014.

[22] K.H. Kim, S.W. Bang, and S.R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," Medical and biological engineering and computing, vol.42, no.3, pp.419–427, 2004.

[23] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.12, pp.2067–2083, 2008.

[24] W. Wen, G. Liu, N. Cheng, J. Wei, P. Shangguan, and W. Huang, "Emotion recognition based on multi-variant correlation of physiological signals," IEEE Transactions on Affective Computing, vol.5, no.2, pp.126–140, 2014.

[25] S. Wioleta, "Using physiological signals for emotion recognition," In: Proceedings of the 6th IEEE International Conference on Human System Interaction (HSI), pp.556–561, 2013.

[26] H. Gunes and M. Piccardi, "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior," ICPR, pp.1148–1153, 2006.

[27] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," J. Netw. Comput. Applicat., vol.30, no.4, pp.1334–1345, 2007.

[28] Y. Wang and L. Guan, "Recognizing Human Emotional State From Audiovisual Signals," IEEE Trans. Multimedia, vol.10, no.5, pp.936–946, 2008.

[29] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The eNTERFACE 05 Audio-Visual Emotion Database," In: Proceedings of the IEEE International Conference on Data Engineering Workshops, pp.1–8, 2006.

[30] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," IEEE Transactions on Affective Computing, vol.3, no.1, pp.18–31, 2012.

[31] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," IEEE Transactions on Affective Computing, vol.3, no.1, pp.42–55, 2012.

[32] A. Dhall, R. Goecke, J. Joshi, J. Hoey, and T. Gedeon, "EmotiW 2016: video and group-level emotion recognition challenges," In: Proceedings of the ACM International Conference on Multimodal Interaction, pp.427–432, 2016.

[33] T. Banziger and K.R. Scherer, "Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus," A. Blueprint for Affective Computing A Sourcebook and Manual, 2010.

[34] F. Ringeval, B. Schuller, M. Valstar, S. Jaiswal, E. Marchi, D. Lalanne, R. Cowie, and M. Pantic, "AV+EC 2015: The First Affect Recognition Challenge Bridging Across Audio, Video, and Physiological Data," In: Proceedings of the ACM International Workshop on Audio/visual Emotion Challenge, pp.3–8, 2015.

[35] Reference Manual: MP System Hardware Guide, http://www.biopac.com.

[36] Reference Manual: AcqKnowledge Software Guide, http://www.biopac.com.

[37] Kinect for Windows, http://msdn.microsoft.com/zh-tw/hh367958.

[38] Kinect for Windows SDK, http://msdn.microsoft.com/en-us/library/hh855347.aspx.

[39] J.J. Gross and R.W. Levenson, "Emotion elicitation using films," Cognition and Emotion, vol.9, no.1, pp.87–108, 1995.

[40] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," In: Proceedings of the IEEE International Conference on Multimedia and Expo, pp.940–943, 2005.