ISSN 1870-4069

Study of Spontaneous and Acted Learn-Related Emotions Through Facial Expressions and Galvanic Skin Response

Andres Mitre-Ortiz, Hugo Mitre-Hernandez

Center for Research in Mathematics,
Human-Centered Computing Lab,
Quantum: Knowledge City, Zacatecas, Mexico
{andres.mitre, hmitre}@cimat.mx

Abstract. In learning environments emotions can activate or deactivate the learning process. Boredom, stress and happy –learn-related emotions– are included in physiological signals datasets, but not in Facial Expression Recognition (FER) datasets. In addition to this, Galvanic Skin Response (GSR) signal is the most representative data for emotions classification. This paper presents a technique to generate a dataset of facial expressions and physiological signals of spontaneous and acted learn-related emotions –boredom, stress, happy and neutral state– presented during video stimuli and face acting. We conducted an experiment with 22 participants (Mexicans); a dataset of 1,840 facial expressions images and 1,584 GSR registers were generated[1]. A Convolutional Neural Network (CNN) model was trained with the facial expression dataset, then statistical analysis was performed with the GSR dataset. MobileNet's CNN reached an overall accuracy of 94.36% in a confusion matrix, but the accuracy decreased to 28% for non-trained external images. The statistical results of GSR with significant differences in confused emotions are discussed.

Keywords: facial expression recognition, GSR, MobileNet, learn-related emotions, CNN.

1 Introduction

Emotions as stress [7] or happy [23] can appear during deep learning activities; other emotions as relaxation and boredom can deactivate the learning activity [23]. Consequently, emotion recognition is a useful tool to adjust the learning environment.

Behaviors (e.g., facial expressions and movements) and physiological data (e.g., signals from the brain, heart, skin, muscle, or eyes.) are non-verbal expressions of the human body, which can give relevant information about the emotions

[1] Hispanic Facial Expressions and Galvanic Skin Response (HFEGSR) dataset available at: https://github.com/andresmitre/Study-of-Spontaneous-and-Acted-Learn-Related-Emotions-Through-FER-and-GSR

of a person; these outputs are more notorious in video stimuli [24, 20, 14] than image stimuli [21, 15] from the International Affective Picture System (IAPS) or sound stimuli from the International Affective Digitized Sounds (IADS)[1] because of the simultaneous perception activities. The information generated from stimuli is collected in datasets.

There are several acted [17, 16] or spontaneous [17, 19] FER datasets, but no for all emotions related to learning as stress and boredom, such emotions can be found in physiological signals datasets –e.g., Electrocardiogram (ECG), Heart Rate (HR) and Galvanic Skin Response (GSR)– GSR being the most representative signal in boredom [9] and other negative emotions [3] –e.g. stress– for correctly classification in emotion recognition. In this paper we introduce: (i) a technique in a controlled experiment using video stimuli to produce GSR and FER datasets of the next basic emotions related to learning: happy, stress, and boredom; (ii) a classifier of spontaneous and acted emotions with our FER dataset; and (iii) a GSR dataset analysis of significant differences in spontaneous and acted emotions.

The paper is organized as follows: in section 2 related works are presented; in section 3 the experiment is detailed; section 4 discusses the results; section 5 describes the contributions and future work.

2 Related Works

CNN, has a high computational cost due to its network depth and the need of large dataset, where tuning hyperparameters is difficult as it is slow to train the CNN and there are numerous parameters to configure [22]. Concerning emotional classifiers, Hierarchy CNNs in [13] tested the accuracy on multiple FER datasets: 61.6% for SFEW2.0 [6], 72.72% FER-2013 [10], 87.71% TFD [27] and 95.38% GENKI4K [30]. On the other hand, CNN's classifiers such as MobileNet [26] and ShuffleNet [31], reduce computational cost sacrificing accuracy. MobileNet is built primarily from depth-wise separable convolutions used in inception models [12] to reduce the computation in the first few layers. In [11], MobileNet achieved a 79.4% with 5.60 millions less Mult-Adds (computation) compared to FaceNet [25] who reached a 83% accuracy. MobileNet allows using low-cost technology (less processing) in educational environments.

In the physiological process, the GSR signal has better classification accuracy than other physiological signals in discrimination of happiness or neutral states [5], boredom [9] and stress [28]. Features collected from GSR signal may contribute to the accuracy of learning emotions classification with FER datasets.

Most acted [17, 16] and spontaneous [17, 19] FER datasets do not include physiological signals during the data collection procedure. Moreover, GSR datasets [14] offer the video of participants' face without restriction of video recording. An interesting technique to collect FER images was described in [2], which the configured approach with EEG device triggers the facial image collection during high emotional rates. We used this principle to collect FER images during more representative GSR data.

3 Experiment

The objectives of the experiment are: (i) the creation of the FER and GSR datasets of acted and spontaneous emotions related to the learning process. (ii) the emotion classification with a CNN throughout FER. (iii) Find significant differences in the GSR Percentage of Change (P.C) between spontaneous/acted and neutrality.

3.1 Materials and Method

To create the FER and GSR datasets, the materials we used were: Haar Feature-based Cascade [29] –a machine learning based approach where a cascade function is trained from many positive an negative images– for the face recognition developed in OpenCV; DSRL Rebel T3 Cannon Camera in order to capture the facial images; Grove GSR Sensor –measure the resistance of the human skin, the resistance is measured with two electrodes of 1/4"(6 mm) dimension of nickel material attached to an electronic circuit– for the physiological data; and a 22in LCD monitor.

Table 1. Selected scenes from films for audiovisual stimuli.

Emotion	Film title	Scene description
Neutral	The lover	Marguerite gets in a car, gets off and walks to a house
Happy	When Harry met Sally	Sally fakes an orgasm in a restaurant
Stress	Irreversible	Woman raped and brutally beaten
Boredom	Merrified and Danckert	Two men ironing clothes.

An audiovisual stimuli was shown to the participant to obtain the spontaneous emotion, it consisted of a visualization of video clips from films validated by participants in diverse work related emotion. Table 1. describes the film related to the stimuli and a brief description of the scene. The stimuli from Table 1 was validated by 100 participants in neutral and happy stimuli [20], boredom was approved by two studies (study 1: 241 participants, study 2: 416 participants) [18], and stress validated by 41 participants [4]. On the other hand, within the acted emotions, facial expression was imitated by the participants using FACS [8] and others proposals, for an interval of 10 seconds.

3.2 Participants

A sample of 22 subjects of Hispanic ethnicity participated in the study: 9 females and 13 males with a range from 18 to 62 years of age.

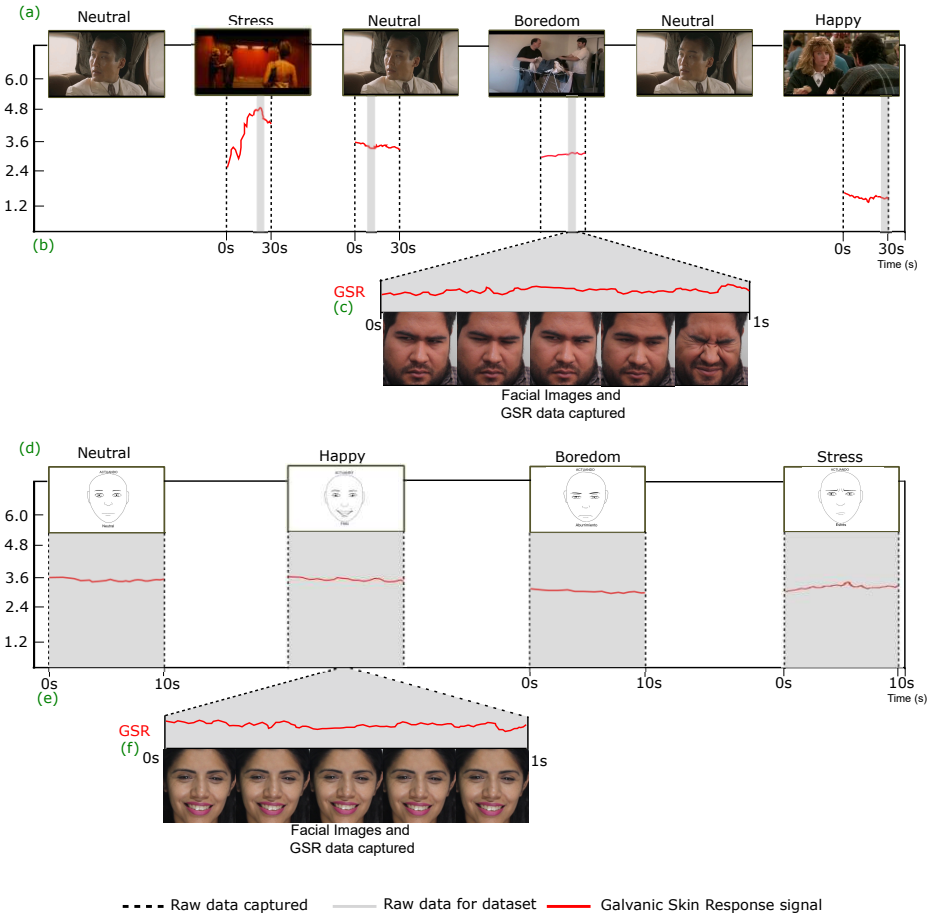

Fig. 1. Procedure of data collection example during spontaneous and acted sessions. Left timeline: (a) films sequence of spontaneous stimuli (neutral, stress, boredom, happy). (b) Timeline procedure for spontaneous stimuli per participant, obtaining the raw data (facial images and GSR) of the session captured within an interval of 30 seconds (gray bars) where the strongest appeared emotion, the GSR signal is noted with the red. (c)(f) Facial images and GSR signal chose for the dataset within a second at 10 FPS. The second was determined according to the biggest difference between the actual emotion recorded, and the baseline (neutral state). Right timeline: (d) acted images sequence (neutral, stress, boredom, happy), the participants imitated the images with the facial expressions described in [8] and other pictorial instructions. (e) Timeline procedure for acted stimuli of 10 seconds per images.

3.3 Procedure

For the creation of the dataset, we proposed a technique that consisted of two sessions: spontaneous and acted. Firstly, the procedure of the experiment

Table 2. MobileNet Body Architecture.

Type/ Stride	Filter Shape	Input Size
Conv / s2	3 x 3 x 3 x 32	224 x 224 x 3
Conv dw / s1	3 x 3 x 32 dw	112 x 112 x 32
Conv / s1	1 x 1 x 32 x 64	112 x 112 x 32
Conv dw / s2	3 x 3 x 64 dw	112 x 112 x 64
Conv / s1	1 x 1 x 64 x 128	56 x 56 x 64
Conv dw / s1	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 128	56 x 56 x 128
Conv dw / s2	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 256	28 x 28 x 128
Conv dw / s1	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 256	28 x 28 x 256
Conv dw / s2	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 512	14 x 14 x 256
5x Conv dw / s1	3 x 3 x 512 dw	14 x 14 x 512
Conv / s1	1 x 1 x 512 x 512	14 x 14 x 512
Conv dw / s2	3 x 3 x 512 dw	14 x 14 x 512
Conv / s1	1 x 1 x 512 x 1024	7 x 7 x 512
Conv dw / s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv / s1	1 x 1 x 1024 x 1024	7 x 7 x 1024
Avg Pool / s1	Pool 7 x 7	7 x 7 x 1024
FC / s1	1024 x 1000	1 x 1 x 1024
Softmax / s1	l—Classifier	1 x 1 x 1000

consisted in providing instructions to the participants about the session and how to position their middle and index phalanges over the GSR electrodes. During the session, participants watched a series of films (stimuli) while raw data was recorded –photos taken at 10 FPS and GSR recording at 10Hz.

Figure 1 illustrates the sequence of spontaneous and acted sessions. Fig. 1(a) shows the spontaneous session; the stimuli consisted in the following sequence: Neutral → Stress → Neutral → Boredom → Neutral → Happy; where at the end of every stimuli, neutrality was induced to the participant to generate the correct desired emotion. Fig. 1(b) indicates the raw data taken in an interval of 30 seconds from the scene where the strongest emotion appeared. Fig. 1(c) facial expression and GSR reading captured at 10 FPS and Hz respectively for spontaneous emotion. For the acted emotions (Fig. 1, bottom timeline), participants were instructed to carefully read the instructions for each emotion by performing the imitation during 10 seconds (Fig. 1(d)). Fig. 1(e) indicates the raw taken in an interval of 10 seconds. Fig. 1(f) facial expression and GSR reading captured at 10 FPS and Hz respectively for acted emotion.

The emotion classification was developed with MobileNet Architecture (Table 2) [11] on TensorFlow; 1,840 images were used for the CNN, the images from the FER dataset were resized at 224 x 224 for the input layer, afterward CNN

classified the images into 8 different classes: “Happy_A”,“Happy_S”, “Stress_A”, “Stress_S”, “Boredom_A”, “Boredom_S”, “Neutral_A” and “Neutral_S” (“_A”, stands for Acted emotion and “_S”, stands for spontaneous emotion). For the external images, a total of 80 images (10 images per class) were used to test the CNN; the images were taken from social networks sites, where users declared their emotional state.

4 Results and Discussions

The classification of the trained dataset brings excellent results: training precision: 100%, validation accuracy: 96.5%. On the other hand, the classification for external images with previously trained dataset shows an average poor performance of 28.75%. To evaluate performance of MobileNet, we analyzed classifier based on confusion matrix analysis (fig. 2), where a correct classification was done with an overall accuracy of 94.6% (fig. 2(a)).

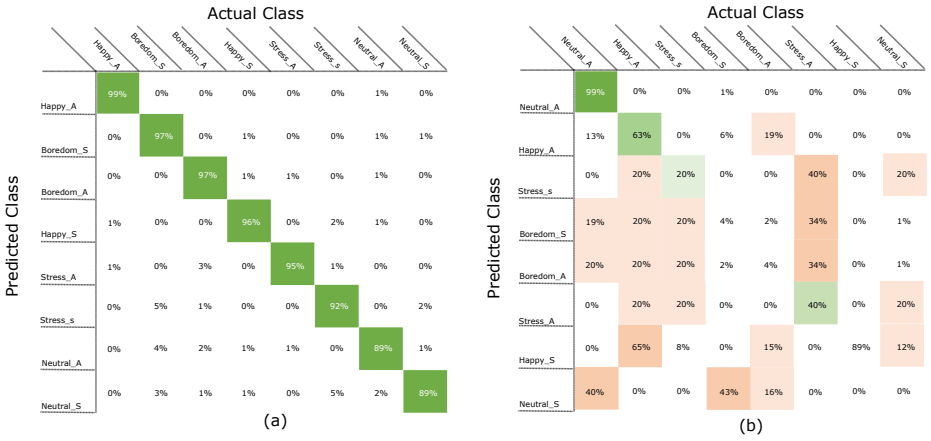

Fig. 2. Confusion Matrix of classifier based on the performance of the CNN. (a) Confusion Matrix of the dataset trained for the CNN, the classification shows an overall accuracy of 94.36%. (b) Confusion Matrix of external images, the classification shows an accuracy of 28.75%. The letters A and S in (a, b) represents the state of emotion, where ”A” stands for acted emotion and ”S” for spontaneous emotion.

Signed Wilcoxon Rank test was used to compare means of change percentage variable of each emotion. For the test with P.C. of spontaneous stress with acted, it was found that Z = -0.141, P > 0.05, no significant differences were found. The test with P.C. of spontaneous happy with the acted one, it was found that Z = 4.46, P = 0.000; P.C. of spontaneous happy (M = 0.196) is greater than happy acted (M = -0.38). Regarding boredom, the result of the test was: Z = -5.167, P = 0.000; P.C. of spontaneous boredom (M = 1.565) is greater than P.C. of

acted boredom (M = 0.006). Therefore GSR can contribute on the recognition of the happy spontaneous confused by a 65% with acted one; in addition the recognition of boredom spontaneous with acted boredom of 2% (fig. 2(b)).

5 Conclusions and Future Work

A classifier of learn-related emotions is a powerful tool for intelligent learning environments, the recognition of emotional states could improve the reinforcement learning of a focus group for education (e.g., educational task and tools) and gaming (e.g., reaction on different levels of immersion). The contribution of this work lies on: i) a new technique to generate a dataset that employs effective stimuli for FER and GSR. ii) Two dataset, FER: 1,840 facial expressions images; GSR: 1,548 registers –baseline difference between emotions and neutrality. iii) Trained model of spontaneous and acted emotions. iv) The possibility of GSR may enhance the classification of the CNN. Future work will consist an including other emotions related to learning (e.g., confusion), as well as incorporating the GSR data into the RGB value channels of the trained images, furthermore, re-train the CNN and contrast results.

References

1. Bradley, M.M., Lang, P.J., Bertron, A., Zack, J., Gintoli, S., Axelrad, J., Cason, J., Brollia, T., Hayden, S., Thorne, B., Karlsson, M., Bittiker, A.: The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual. Tech. Rep. 2, University of Florida, Gainesville, Fl (2007)
2. Cabada, R.Z., Estrada, M.L.B., HernÃ¡ndez, F.G., Bustillos, R.O., Reyes-GarcÃa, C.A.: An affective and web 3.0-based learning environment for a programming language. Telematics and Informatics 35(3), 611 – 628 (2018), sI: EduWebofData
3. Colomer Granero, A., Fuentes-Hurtado, F., Naranjo Ornedo, V., Guixeres Provinciale, J., Ausín, J.M., Alcañiz Raya, M.: A comparison of physiological signal analysis techniques and classifiers for automatic emotional evaluation of audiovisual contents. Frontiers in computational neuroscience 10, 74 (2016)
4. Cousijn, H., Rijpkema, M., Qin, S., Marle, H.J.F.V., Franke, B., Hermans, E.J.: Acute stress modulates genotype effects on amygdala processing in humans 107(21), 9867–9872 (2010)
5. Das, P., Khasnobish, A., Tibarewala, D.N.: Emotion recognition employing ecg and gsr signals as markers of ans. In: 2016 Conference on Advances in Signal Processing (CASP). pp. 37–42 (June 2016)
6. Dhall, A., Goecke, R., Lucey, S., Gedeon, T., et al.: Collecting large, richly annotated facial-expression databases from movies. IEEE multimedia 19(3), 34–41 (2012)
7. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction 22(2), 145–157 (2012)
8. Ekman, P., Rosenberg, E.L.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)

9. Giakoumis, D., Tzovaras, D., Moustakas, K., Hassapis, G.: Automatic recognition of boredom in video games using novel biosignal moment-based features. IEEE Transactions on Affective Computing 2(3), 119–133 (July 2011)
10. Goodfellow, I., Yoshua, B., Aaron, C.: Deep learning. MIT Press (2016)
11. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (2017)
12. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
13. Kim, B.K., Roh, J., Dong, S.Y., Lee, S.Y.: Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. Journal on Multimodal User Interfaces 10(2), 173–189 (2016)
14. Koelstra, S., Muhl, C., ..., M.S.I.T., 2012, U.: Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing 3(1), 18–31 (2012)
15. Lang, P., Bradley, M., Cuthbert, B.: Technical report a-8, international affective picture system (iaps): affective ratings of pictures and instruction manual (university of florida, gainesville, fl) (2008)
16. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A.: Presentation and validation of the radboud faces database. Cognition and Emotion 24(8), 1377–1388 (2010)
17. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. pp. 94–101. IEEE (2010)
18. Markey, A., Chin, A., Vanepps, E.M., Loewenstein, G.: Identifying a Reliable Boredom Induction. Perceptual and Motor Skills 119(1), 237–253 (2014)
19. Mavadati, S.M., Mahoor, M.H., Bartlett, K., Trinh, P., Cohn, J.F.: Disfa: A spontaneous facial action intensity database. IEEE Transactions on Affective Computing 4(2), 151–160 (2013)
20. Megías, C.F., Mateos, J.C.P., Ribaudi, J.S., Fernández-Abascal, E.G.: Validación española de una batería de películas para inducir emociones. Psicothema 23(4), 778–785 (2011)
21. Nourbakhsh, N., Chen, F., Wang, Y., Calvo, R.A.: Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference. ACM Transactions on Interactive Intelligent Systems 7(3), 1–20 (2017)
22. Pérez-Espinosa, H., Avila-George, H., Rodriguez-Jacobo, J., Cruz-Mendoza, H.A., Martínez-Miranda, J., Espinosa-Curiel, I.: Tuning the parameters of a convolutional artificial neural network by using covering arrays. Research in Computing Science 121, 69–81 (2016)
23. Rainhard Pekrun, R.P.P.: Control-value theory of achievement emotionsNo Title. In: International Handbook of Emotions in Education, chap. 7, p. 22 (2014)
24. Sauter, D.A., Fischer, A.H.: Can perceivers recognise emotions from spontaneous expressions? Cognition and Emotion 32(3), 504–515 (2018)
25. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)
26. Sifre, L., Mallat, S.: Rigid-motion scattering for image classification. Ph.D. thesis, Citeseer (2014)

27. Susskind, J.M., Anderson, A.K., Hinton, G.E.: The toronto face database. department of computer science, university of toronto, toronto, on. Tech. rep., Canada, Tech. Rep, 3 (2010)
28. Villarejo, M.V., Zapirain, B.G., Zorrilla, A.M.: A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. Sensors (Switzerland) 12(5), 6075–6101 (2012)
29. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 1, I–511–I–518 (2004)
30. Whitehill, J., Serpell, Z., Yi-Ching Lin, Y.C., Foster, A., Movellan, J.R.: The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions. IEEE Transactions on Affective Computing 5(1), 86–98 (2014)
31. Xiangyu, Z., Xinyu, Z., Mengxiao, L., Jian, S.: Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Computer Vision and Pattern Recognition (2017)