International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Deep Learning-based Facial Expression Recognition and Analysis for Filipino Gamers

Juan Raphael Sena, Melvin Cabatuan

Abstract: ***This paper presents a computer vision based emotion recognition system for the identification of six basic emotions among Filipino Gamers using deep learning techniques. In particular, the proposed system utilized deep learning through the Inception Network and Long-Short Term Memory (LSTM). The researchers gathered a database for Filipino Facial Expressions consisting of 74 gamers for the training data and 4 gamer subjects for the testing data. The system was able to produce a maximum categorical validation accuracy of .9983 and a test accuracy of .9940 for the six basic emotions using the Filipino database. The cross-database analysis results using the well-known Cohn -Kanade+ database showed that the proposed Inception-LSTM system has accuracy on a par with the current existing systems. The results demonstrated the feasibility of the proposed system and showed sample computations of empathy and engagement based on the six basic emotions as a proof of concept.***

Index Terms: ***deep learning, gamer facial expression, emotion recognition, Inception network, LSTM.***

I. INTRODUCTION

Recently, gaming has greatly increased in popularity among several different interdisciplinary applications, not only to entertainment, but also to other areas like education [1], training and simulation [2], healthcare [3], and various other domains. Along with the wide acceptance of gaming is the increase of interest in various relevant problems, i.e. game personalization, user experience design, game development issues, as well as public policy issues. One of the recurring themes in game research domain is the analysis of facial expressions [4] and the corresponding basic emotions. However, differing cultures between nations of people affects the emotional responses of an individual due to their differences. Thus, there is a need to provide a facial recognition and analysis system that is specifically suited to a particular subset of the population.

Artificial neural networks (ANN) are algorithms inspired by the human brain, which finds many applications, namely, face detection [5][6], voice recognition [7], electronic communication [8], agriculture/aquaculture [9][10][11], forensics [12], healthcare [13][14][15] and many others. Deep learning is essentially an ANN with many layers. The prospect of facial expressions recognition analysis through deep learning has already been undertaken by a few researchers, e.g. [16] utilized two different deep learning algorithms, Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN) on the Cohn-Kanade+ dataset [17] and undertook comparisons for the feasibility of both systems. In line with the study, current research delves deeper from spatial representations to temporal ones using part-based hierarchical bidirectional recurrent neural network (PHRNN) [18]. As for facial expression recognition and analysis for gamers, a feasibility study was conducted in [19] to determine the usability evaluation of gamers through their emotional responses. Utilizing the evaluation of the six basic emotions during gameplay, the study was utilized to compare the efficiencies of advanced players in completing certain tasks in Call of Duty and comparing those outputs with the results of emotional recognition. The proposed study aims to provide a facial expression recognition system more suited to Filipinos and Filipino Gamers. To solve the problem demonstrated by [16], the system will also provide a Filipino facial expression database for the six basic emotions. In line with that, the facial expression recognition system will apply a spatial-temporal deep learning solution. The system will consist of a Convolutional Neural Network (CNN) architecture known as Inception v3 [20], and a Recurrent Neural Network (RNN) known as Long-Short Term Memory (LSTM) [21]. Lastly, the study also evaluates Empathy and Engagement emotional responses in addition to the six basic emotions.

II. METHODOLOGY

A. Filipino Facial Expressions Database

In order to train the Deep Learning Architecture, this paper utilizes the Cohn-Kanade+ dataset [17] as well as newly gathered dataset of facial expressions for Filipino faces. In this study, the researchers have collected 414 clips of facial expressions from 74 Filipino subjects categorized under six basic emotions. For the data gathering for both database/training and testing, the setup will utilize a white screen background to remain consistent as shown in Fig. 1.

Revised Manuscript Received on 30 July 2019.

* Correspondence Author
Juan Raphael Sena, Electronics and Communications Engineering Department, De La Salle University, Manila, Philippines.
Melvin Cabatuan, Electronics and Communications Engineering Department, De La Salle University, Manila, Philippines.

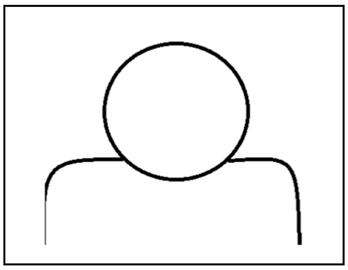


Fig. 1. Guideline of the face's position for training data.

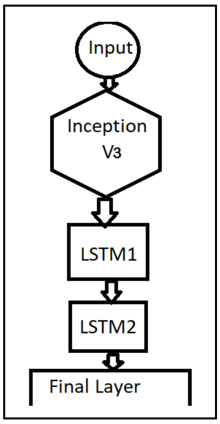


Fig. 2. Block diagram for the whole architecture of the proposed system

The background should not conflict the setup, and it would be ideal to isolate the face. The face position must remain at 0 degrees from the camera to keep the standard position similar to the clips utilized in the Cohn-Kanade Dataset [22]. The gathered database is a set of video clips taken using Macbook Pro's 720p FaceTime HD camera and Online Broadcaster Software (OBS) to capture video in mp4 format with a 1120x720 resolution at 30 frames per second with a duration ranging from 1-2 seconds. The clips' emotional responses were obtained from posed expressions based on Action Units(AUs) on Facial Action Coding Systems [23]. These video clips were used to train the deep learning model. The implementation of the trained model in the proposed system will also take place on the same data-gathering machine.

B. Deep Learning Architecture

The proposed system contains two parts: 1) Spatial Representation via Inception v3 and 2) Temporal Representation via Long-Short Term Memory, as shown in Fig. 2. Inception v3 is utilized via the Deep Learning technique known as Transfer Learning [24]. Transfer learning is a technique which repurposes an existing graph and retrains it for other usages. Inception V3 [20] is a high-performing Convolutional Neural Network which utilizes dimensional reduction. The Inception architecture is based on the problem at how an optimal local sparse structure in a convolutional neural network for images can be approximated given dense components. The concept is to discover the optimal local construction and repeat it spatially. It is a layer-by-layer construction, which utilizes clusters with high correlation of correlation statistics of the previous layer. The Inception architecture is a combination of all the layers with their output filter merged into the input of the next stage. In order to avoid patch-alignment issues, the filter sizes are limited to 1x1, 3x3, and 5x5. From lower units, these correlated units would be first concentrate on local regions. The convolutions will be concentrated on 1x1 initially until the ratio of 3x3 and 5x5 increases as it reaches the higher layers. The purpose of the Inception network for the system is to find the proper significant features per frame of an image sequence. It is to represent the spatial features of each frame and thus only outputs a characterization of an image rather than the whole sequence.

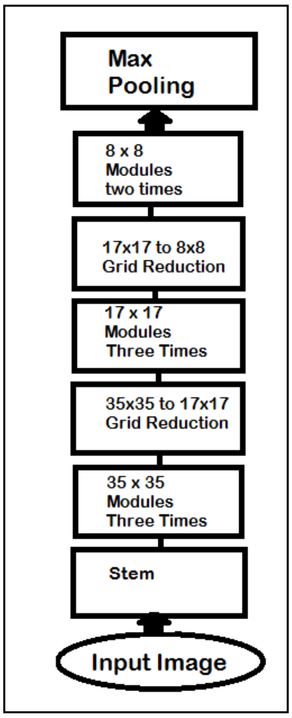


Fig. 3.Simplified Inception V3 network flow

A Recurrent Neural Network (RNN) known as Long-Short Term Memory or LSTM [21] was utilized to understand the relationship of the features per frame with respect to the previous frames as it covers the temporal representation of the system. The usage of LSTM is a type of RNN that can process relationships between frames from farther than t-1. The system utilizes two of such networks in order to uncover deeper relationships between the frames.

The LSTM utilizes a memory cell that contains gates that regulate and control the output and maintained state of the system. Then, longer duration relationships can be uncovered to understand the corresponding Action Units (AUs) and Facial Expression Label.

C. Six Basic Emotions and Softmax Function

The Six Basic Emotions are Joy, Anger, Sadness, Fear, Disgust, and Surprise as defined in [26]. The major concept of the Six Basic Emotions is that they are the emotions that have little differences between cultures and are the basis for the other emotions. This is the basic classes for Facial Expression Recognition.

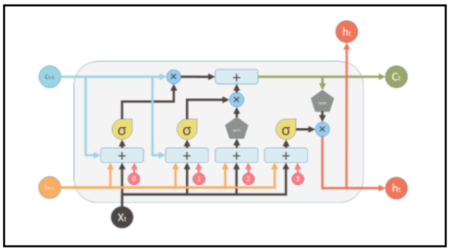


Fig. 4.Memory Cell for RNN where C is memory block and h is outputted predicted label [25]

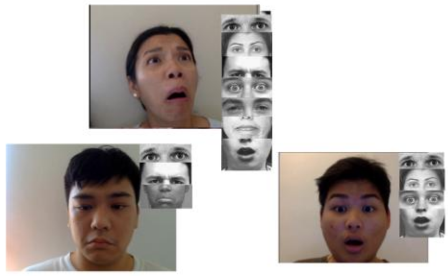
Fig. 5.Sample Facial Expressions of Anger, Disgust, and Joy with their respective action units (AUs).

Given the input faces data and the Architecture, the study first established the ground truth as the basis for these emotions. The ground truth is known as the Action Units (AUs) in the Facial Action Coding System (FACS) [23]. As shown in Fig. 5 and Fig. 7, the various facial expressions of the Six Basic Emotions were defined through these Action Units. The basis of the system's classification all relies on the features extracted by the two connected LSTM Networks and from the second LSTM Layer, it tries to identify the more derivative relationship for each frame. It is here that the system tries to understand which AUs were shown for the system. It is from these AUs that the system will classify the emotional response of the test data. For example, Joy was classified under the AUs 6 and 12 where AU 6 was defined as the Cheek Raiser and AU 12 is the Lip Corner Raiser. After the two LSTM Systems, and a single predicted label for the image sequence was shown, the system finalizes the classification through the final layer of the Network known as the Softmax Function.

The Softmax function is a generalized form in logistic regression for classification of mutually exclusive classes, as in (1).

$$\begin{bmatrix} p(y^i = 1|x^i;\theta) \\ p(y^i = 2|x^i;\theta) \\ \vdots \\ p(y^i = k|x^i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \vdots \\ e^{\theta_k^T x^i} \end{bmatrix} \quad (1)$$

The output of this function will be a vector that will contain six values (one for each emotion) which shows the probability of an emotion, respectively.



Sample Facial Expressions of Fear, Sadness, and Surprise with their respective action units AUs

D. Empathy and Engagement Evaluation

For the application part of the system, four subjects were chosen to conduct gameplay in order to help uncover the empathy and engagement vectors for the emotional system. The setup was made similar to data gathering. The subjects were then made to play Shadowverse while their faces were extracted. The duration of the video extraction will be 30 seconds midway through playing for demonstration. This was conducted to maximize the state of a player's flow within the gameplay [27]. Then, the Engagement and Empathy were evaluated. Engagement is defined as a state of High Valence/Dominance and High Arousal under the Active State Space [28] [29]. Considering that, the output of the classified emotions was based on percentages, this vector is meant to demonstrate the level of engagement one has based on their Six Basic Emotions.

$$\begin{aligned} Arousal &= Surprise + (Anger - Disgust) * .866 + Fear * .525 \\ Valence &= Joy - Sadness - Fear * .525 + (Anger - Disgust) * .525 \quad (2) \\ Engagement &= Arousal * .525 + Valence * .525 \end{aligned}$$

There are two types of Empathy [30]: 1) Situational Empathy: 2) Dispositional Empathy. Situational Empathy are the emphatic reactions in a specific situation whereas Dispositional Empathy is described as a person's stability.

Dispositional Empathy was focused in this system for evaluation. Dispositional Empathy can be examined through the stability of a player's emotions and can observe whether a player's emotions change drastically based on the system. Dispositional responses can then be interpreted based on the rate of change of the emotions between a fixed point in time. The equation to obtain Dispositional Empathy is depicted as follows:

$$Dispositional\ Empathy = 1 - \sum \Delta Emotions \quad (3)$$

Emphatic response will be a negative value if Emotional response is shown to change in large scales between examinations of sequences. Positive values refer to the stability of an individual to remain in his/her own emotional state. Provided the emotional states produced by the system, the Dispositional Empathy of an individual can be obtained.

III. RESULTS AND DISCUSSION

The results are divided into two categories: 1) Six Basic Emotion Recognition and Comparison between Datasets; and 2) Empathy and Engagement Results. The first part will tackle the accuracy of the Deep Learning Architecture with respect to three different datasets. The last part shows the feasibility of the evaluation of Dispositional Empathy and Engagement though the data obtained by the four test subjects.

A. Training Data Results

The other dataset that was utilized in comparison will be the Cohn-Kanade + Database [4]. It is one of the widely used databases for Facial Expression Recognition. Its main ground truth is also the Action Units and shares many similarities with the acquired Filipino Database.

TABLE I. COMPARISON BETWEEN FILIPINO DATASET VS. EXISTING DATASETS

Characteristics	Filipino Database	Cohn Kanade Database	Cohn Kanade+ Database
Number of Subjects	74	97	97
Number of Video Clips	452	169	327
Anger Clips	62	27	45
Disgust Clips	62	30	59
Fear Clips	77	23	25
Joy Clips	147	32	69
Sadness Clips	66	26	28
Surprise Clips	38	31	83
Ground Truth	AUs and Emotion Labels	AUs and Emotion Labels	AUs and Emotion Labels
Emotion Response	Posed/ Spontaneous	Posed/ Spontaneous	Posed/ Spontaneou
Resolution	1120x720	640x490	640x490

As depicted in Table 1, the Filipino Dataset has more video clips than that of Cohn-Kanade+ with a lot better resolution and more stable distribution of the Six Basic Emotions. For the specifications of the Deep Learning Architecture, the system runs until the 240th Epoch, Train to Test Split of 80 is to 20, and Batch size of 128.

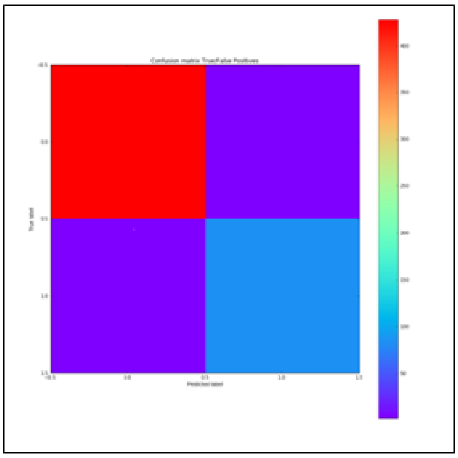


Fig. 6.Confusion Matrix using the Filipino Database

As one can see from the matrix on Fig. 7, the Recognition System properly detects the classifiers very well. The system's strong points are its classification of Fear as well as Surprise. This can be attributed to the necessity of an exaggerated facial response when it comes to those faces and thus helps identify them far better than most of the other emotions. The next two emotions that are identified well are Joy and Disgust. Joy being a positive emotion and thus easier to express in people can help in it is easily identifiable state. What essentially lacks in its identification is due to the potential subdued nature of some expressions of Joy can be in that the corresponding Action Units take place but in a less identifiable manner. Disgust comes in the fact that disgust faces are easy to identify due to their nature, but may find difficulty as the Action Unit Lip Depressor obtains the total depression of the lip, but some cases of Disgust faces only rely on one side of the depression. Even while they are identified correctly, the system has a harder time on Anger and Sadness. This can be attributed to the database itself as the features of Anger through its action units can be confused with the Disgust class. Sadness itself is a difficult emotion to express due to the nature of the emotion. It is not an emotion usually desired to be expressed by individuals and as such may be harder to extract considering it multiple forms of exaggerations or subjugation. As the process of obtaining the videos used for input data, the hardest one to pose for was indeed sadness, which can be attributed to the lack of desire for an individual to express.

- Precision = .9953;
- Recall = .8359;
- F - Harmonic Mean = 0.9086

The system itself projects a maximum categorical validation accuracy of .9983 and a test accuracy of .9940.

B. Cross Database Validation Results

Currently, there is no comparison for application of test accuracy concerning Gameplay analysis, so there is no way to compare the test accuracy of the system when it comes to the ground truth for the application. Instead, the research will have its Accuracy compared to cross-database recognition rates. The cross-database validation utilizing the Filipino-trained system and the testing data as Cohn-Kanade+ for the system goes as shown in Table 2.

TABLE II. CROSS-DATABASE VALIDATION FOR INCEPTION-LSTM

Database used for testing	Inception-LSTM
CK+	0.678541

A similar research was utilized to compare the data.
The first comparison is a Facial Expression Recognition using Deep 3D Convolutional Neural Networks. The concept of the research is similar, but the difference lies in the usage of the Neural Networks.

Mahoor [18] utilizes a shallower form of Inception known as 3DInception-Resnet that is also followed by just one LSTM layer.

- Supervised Kernel Mean Matching [31]
- l - p-norm multiclass Support Vector Machines [32]
- Deep Neural Networks [33]
- 3D Convolution Neural Networks [34]

The cross-database analysis (including the cross-database validation for previous research), is depicted in Table III. The results showed that the proposed Inception-LSTM system has accuracy on a par with existing systems. The similarities of the two Inception-based systems, 3D Inception and Inception LSTM, could be due to the similar structures of the network. The difference lies between the deeper spatial learning output and the deeper LSTM layer implemented in the proposed system.

TABLE III. CROSS-DATABASE VALIDATION COMPARISON WITH OTHER METHODS

Database used for testing	Supervised kernel Mean Matching	l - p norm Multiclass SVM	DNN	3D Inception Resnet	Inception -LSTM
CK+	0.56	0.612	0.642	0.6752	0.678541

C. Application on Filipino Gamers

The emotion recognition system's output considers the most probable emotional response of the system based on the facial expressions present in the input system. As such the largest emotional responses based on the ground truth from OpenFace will determine whether or not the Recognition System has been able to detect the most probable emotion.

Basing on the Action Units (AUs) as Ground Truth of each respective Subject and the Output of the system of each respective Subject.

TABLE IV. ACCURACY OF THE PROPOSED SYSTEM VS. GROUND TRUTH

	Test Accuracy for best prediction
Subject 1	0.806683
Subject 2	0.800475
Subject 3	0.794926
Subject 4	0.860092
Subject 5	0.796238
Average	0.811683

Considering the Table IV, as the ground truth emotional spectrum and the recognition system's predicted label only considers the most probable emotional response for each system per second, the average accuracy of the system is for Filipino Gamers is at 0.811683.

The difference between accuracy for the training and the application can be attributed to difference in the data used to train the data as well as the data used to input the system. While the same nationality and culture are considered for both forms of data, the Application Data for Gamers itself, display less posed expressions and ones that are rather more spontaneous. In doing so, the system also misses abrupt emotional responses.Considering the focus of engagement, one can see that the more hardcore gamers exhibit either a larger state of engagement compared to that of those that are not quite involved with the game. Looking back from [28], the facets of Engagement is more reflected on these subjects. The low engagement rates of the gameplay is attributed to low-level challenges, frustrations, and lack of connectivity to the task.For Empathy, the more hardcore gamers produce a lower average dispositional empathy. This can be attributed to the high reactive responses the Subjects have due to being more used to the game and its mechanics. While it is shown that the other subjects express a large amount of reactive response too, Subject 3 and Subject 4 display more varied emotions and more potential action units due to their involvement with the games themselves. Subject 2 also displays a lower average that is reflected from his change between Joy and Anger. Looking back on [30], Empathy is more reflected to that of Subject 2, 3, and 4, which are moderate to hardcore gamers. While even their levels of engagement are different, their emphatic response are similar due to their familiarity and thus attachment to the gameplay. Even with the negative engagement, a hardcore gamer's emotional response reacts more due to being emphatic to the responses conducted by games.

IV. CONCLUSION

Overall, the proposed system has achieved good facial expression recognition results of about 99.83% for the six basic emotions utilizing the Filipino database and an accuracy (67.85%) on a par with the current existing systems using the popular Cohn-Kanade+ database. The study has been able to provide a new facial expressions database, which is among the contributions of this work. Furthermore, the results demonstrated the feasibility of quantifying empathy and engagement based on the six basic emotions recognized by the proposed system. Algorithm-wise, a deep learning model containing the combined Inception v3 framework and 2x LSTM model for video processing and classification has been developed, which can be utilized for other image processing problems. Future research may explore the updated Inception network since the development team behind the Inception Networks are always constantly improving it. Furthermore, the Filipino Dataset can be utilized for more complex emotional recognition tasks. Provided with the tools, it can be possible to identify emotional responses exclusive to Filipino culture. Future applications may utilize the system output as inputs or processing to facial video inputs to other software.

REFERENCES

1. M.A. Cohen, G.O. Niemeyer, and D.S. Callaway, "Griddle: Video Gaming for Power System Education," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3069–3077, 2017.
2. A. R. Darragh, S. Lavender, B. Polivka, C. M. Sommerich, C. E. Wills, B. A. Hittle, R. Chen, and D. L. Stredney, "Gaming Simulation as Health and Safety Training for Home Health Care Workers," *Clinical Simulation in Nursing*, vol. 12, no. 8, pp. 328–335, 2016.
3. A. K. Hall and H. R. Marston, "Games for Health in the Home: Gaming and Older Adults in the Digital Age of Healthcare," *Handbook of Smart Homes, Health Care and Well-Being*, pp. 579–588, 2016.
4. P.M. Blom, S. Bakkes, C.T. Tan, S. Whiteson, D. Roijers, R. Valenti, and T. Gevers, "Towards personalised gaming via facial expression recognition," *In Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.

5. W. Yang, L. Zhou, T. Li, and H. Wang, "A Face Detection Method Based on Cascade Convolutional Neural Network," *Multimedia Tools and Applications*, 2019.
6. S.D.F. Hilado, E.P. Dadios, R.C. Gustilo, "Face detection using neural networks with skin segmentation", Proceedings of the 2011 IEEE 5th International Conference on Cybernetics and Intelligent Systems, CIS 2011, pp. 261-265, 2011.
7. S.G.E. Brucal, A.D.M. Africa, and E.P. Dadios, "Female Voice Recognition Using Artificial Neural Networks and MATLAB Voicebox Toolbox," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(1-4), pp.133-138, 2018.
8. L. Villanueva, R.C. Gustilo, "Artificial neural network based antenna sensitivity assignments for chaotic Internet Service Provider network architecture" *International Journal of Engineering and Technology* (UAE), 7(2), pp. 14-17, 2018.
9. R.C. Gustilo, E.P. Dadios, "Machine vision support system for monitoring water quality in a small scale tiger prawn aquaculture", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 20(1), pp. 111-116, 2016.
10. R.C. Gustilo, E.P. Dadios, " Behavioural response analysis using vision engineering (BRAVENet)", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 21(2), pp. 211-220, 2017.
11. A.D.M. Africa, J.C.C.A. Aguilar, C.M.S. Lim, P.A.A. Pacheco, and S.E.C. Rodrin, "Automated aquaculture system that regulates Ph, temperature and ammonia," *In 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management* (HNICEM), pp. 1-6, IEEE, 2017.
12. J. Del Espiritu, G. Rolluqui, R.C. Gustilo, "Neural network based partial fingerprint recognition as support for forensics", *8th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, HNICEM 2015.
13. A.D.M. Africa, J. Velasco, "Development of a urine strip analyzer using artificial neural network using an android phone," *ARPN Journal of Engineering and Applied Sciences*, 12, pp. 1706-1713, 2017.
14. P.J.M. Loresco, and A.D. Africa, "ECG Print-out Features Extraction Using Spatial-Oriented Image Processing Techniques," *Journal of Telecommunication, Electronic and Computer Engineering* (JTEC), 10(1-5), pp.15-20, 2018.
15. A.D.M. Africa, "A Rough Set Data model with presbyopic and hypermetrope parameters for eye vision diagnostics," *In 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1-5, IEEE, 2015.
16. H. Jung, and J. Kim, "Development of deep learning-based facial expression recognition system," *Frontiers of Computer Vision (FCV)*, 2015 21st Korea-Japan Joint Workshop on, 2015.
17. J.F. Cohn, and T. Kanade, "The extended cohn-kanade dataset (ck+): A complete expression dataset for action unit and emotion-specified expression. *In Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis* (CVPR4HB 2010).
18. K. Zhang, Y. Huang, Y. Du, and L. Wang, L., "Facial expression recognition based on deep evolutional spatial-temporal networks," *IEEE Transactions on Image Processing*, 26(9), pp.4193-4203, 2017.
19. F. Iqbal, "Understanding User Interaction in a Video Game by using Eye Tracking and Facial Expressions Analysis" (Master's thesis), 2015.
20. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818-2826, 2016.
21. C.Y. Ma, M.H. Chen, Z. Kira, and G. AlRegib, "TS-LSTM and temporal-inception: exploiting spatiotemporal dynamics for activity recognition," *Signal Processing: Image Communication*, vol. 71, pp.76-87, 2019.
22. T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (Cat. No. PR00580) , pp. 46-53, IEEE, 2000.
23. B. Farnsworth, "Facial Action Coding System (FACS)—A Visual Guidebook. https://imotions. com/blog/facial-action-coding-system/. Accessed on July 2, 2018, 2016.
24. Google. "Image Classification Transfer Learning with Inception v3," [Online]. Available: https://codelabs.developers.google.com/codelabs/cpb102-txf-learning /index.html#0. [Accessed: 2016].
25. C. Olah, "Understanding lstm networks," 2015.
26. P. Ekman, and W. V. Friesen, "Constants across cultures in the face and emotions," *J. Personality Social Psychology*, vol. 17(2), 1971.
27. K. Isbister, "How games move us: Emotion by design," Mit Press, 2016.
28. S. Moss, "Engagement at work," 2016.
29. E. Philips,"Go with the flow: engagement and concentration are key," 2013.
30. K. Stueber, "Measuring empathy," 2013.
31. Y.Q. Miao, R. Araujo, and M.S. Kamel, "Cross-domain facial expression recognition using supervised kernel mean matching," *In 2012 11th International Conference on Machine Learning and Applications*, vol. 2, pp. 326-332, IEEE, 2012.
32. X. Zhang, M.H. Mahoor, and S.M. Mavadati, "Facial expression recognition using lp-norm MKL multiclass-SVM," *Machine Vision and Applications*, 26(4), pp.467-483, 2015.
33. A. Mollahosseini, D. Chan, and M.H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *In 2016 IEEE Winter conference on applications of computer vision (WACV)*, pp. 1-10, IEEE, 2016.
34. B.H. Mohammad Mahoor, "Facial expression recognition using enhanced deep 3D convolutional neural networks," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 30-40, 2017.

AUTHORS PROFILE

Juan Raphael Sena received the Bachelor of Science degree and Master of Science in Electronics and Communications engineering from De La Salle University (DLSU), Manila, Philippines, in 2016.

Melvin Cabatuan, Ph.D. received the Bachelor of Science degree (cum laude) in Electronics and Communications engineering from Cebu Institute of Technology, Cebu, Philippines, in 2004; Master of Engineering from Nara Institute of Science and Technology (NAIST), Ikoma, Nara, Japan, in 2010; and Ph.D. in Electronics and Communications engineering from De La Salle University (DLSU), Manila, Philippines, in 2016. He joined the Electronics & Communications Engineering Department of DLSU in 2011, where he is currently an Associate Professor. His research interest involves computer vision, health informatics, and machine/deep learning applications to various engineering problems. He is an active member of the Institute of Electronics Engineers of the Philippines (IECEP) since 2004, and the IEEE Republic of the Philippines section since 2011.