

Customer's spontaneous facial expression recognition

Golam Morshed, Hamimah Ujir, Irwandi Hipiny

Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Malaysia

Article Info

Article history:

Received Jul 16, 2020

Revised Jun 4, 2021

Accepted Jun 13, 2021

Keywords:

Customer's emotion

Face detection

Facial expressions

ABSTRACT

In the field of consumer science, customer facial expression is often categorized either as negative or positive. Customer who portrays negative emotion to a specific product mostly means they reject the product while a customer with positive emotion is more likely to purchase the product. To observe customer emotion, many researchers have studied different perspectives and methodologies to obtain high accuracy results. Conventional neural network (CNN) is used to recognize customer spontaneous facial expressions. This paper aims to recognize customer spontaneous expressions while the customer observed certain products. We have developed a customer service system using a CNN that is trained to detect three types of facial expression, i.e. happy, sad, and neutral. Facial features are extracted together with its histogram of gradient and sliding window. The results are then compared with the existing works and it shows an achievement of 82.9% success rate on average.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Golam Morshed

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak, Malaysia

Email: gmriyal@gmail.com

1. INTRODUCTION

Facial expression recognition (FER) systems are currently advanced in terms of human capability in detecting and responding to facial expressions while interacting with each other. Several approaches are developed to detect consumer behavior levels to observe their satisfaction level [1], [2]. These approaches are used to develop applications mainly used in the airport, market, shopping mall et cetera. By observing customer's facial expressions and analyzing them, the provider can offer the best service suitable for the customers. This research is motivated by the customer service complaints received, especially at the government offices, and no study has focused on analysing the sequence of facial expressions during the interaction at the customer service desk.

Consumer's feelings can be observed from their facial expressions. Through face-to-face communication, a service provider can find an affective state by using facial expressions. Many researchers have illustrated that facial expressions are an imperative marker of feeling [3]-[5]. Researchers have discovered that disappointed consumers always undergo negative emotions. Both complaint and non-complaint consumers have experienced the same negative emotions that cause service failure [6]. In addition, the attempt at service recovery always fails if the service provider is unable to discover the negative emotions of the complainer. Seven-expression schema [7], [8] rarely occurs in garments shopping and therefore only three-expression schema is used. One of the examples in this field of work is SentrIBUTE [9], which simplifies emotional analysis to positive, neutral, and negative.

To observe customer interest through a specific product from their expression level, researchers develop robust FER systems using deep learning architecture such as convolutional neural networks (CNN) [10], [11]. CNN-based FER systems are developed to recognize customers' seven kinds of facial

expressions from a different perspective. In consumer science, customer emotional factors are classified into two main categories as positive and negative [12]. While observing a specific product by the customer, positive emotion evokes the customer to accept or purchase it [13]. This positive expression is observed from their happy expression. Besides, negative emotions by the customer more likely to reject the product. This negative expression can be observed in their sad expression. To analyze the customer facial expression by CNN network we choose only three expressions as happy, sad, and neutral.

The existing method for facial expressions has some limitations. The two-dimensional (2D) appearance-based local approach for the extraction of intransient facial features and recognition does not contain enough feature points which results in an incorrect accuracy rate. Customers tend to act faster and change their expressions spontaneously. The effective way to understand the customer's needs is to analyze their spontaneous expressions. Moreover, it is important to analyze the sequence of facial expressions and relate them to the case events. Nonetheless, no study has focused on this area.

The CNN-based FER system solved a lot of problems such as frontal faces, noise, and illumination issues. By implementing dropout and controlling overfitting, the CNN network becomes more robust [14]. The accuracy of the CNN network highly depended on the feature extraction technique and classifier. In most of the CNN-based systems, available classifiers such as support vector machine, bagging, etc are being used with some limitations [15]. To train the CNN network available facial expression dataset is being used. Due to the limitation of those techniques the CNN network accuracy rate was reduced. In the CNN network, the deeper network can help to improve the accuracy rate in the training process [16]. In the training process, each layer will learn from the possible input value from the previous layers. If the first layer captures low-level features, then the next level will also capture low-level features. As a result, each layer of the CNN network is trained with low-level features which will reduce the accuracy level. To overcome this problem, more layers need to be added to the CNN network that will help to learn from various levels of abstractions. In our experiment, we applied multiple layers in the training process to improve accuracy.

This paper presents findings from the spontaneous facial expression classification. Section 2 describes the related works in this field. We present data collection, data pre-processing, and our framework in section 3, followed by the experiment setting, results, and analysis in section 4. Finally, the conclusions are drawn.

2. BACKGROUND

Consumers always show their reactions while buying a product or experiencing a service. Usually, their reaction reflects satisfaction or dissatisfaction. In face-to-face customer service, it is hard to detect consumer feelings of like or dislike if they come from different countries and cultures [17]. An *et al.* [18], when showing dissatisfaction towards a product or service, the consumer who comes from eastern cultures shows negative emotions rather than raise their voice. Thus, the importance of understanding consumer expressions has received significant attention from researchers intending to reduce the failure rate of the company.

Lu *et al.* [19] proposes a real-time automated recommender system for garments sales. The video-based recommender (VAR) system is developed to observe the customer's emotions while trying the garment product in front of a mirror. With the help of computer vision techniques and customer preference, the system can predict the customer desired product based on their facial expressions also recommending the customer preference. The first step of this system is to capture the customer's reactions while trying or observing a garment product in front of the mirror, using a low-cost camera. The next step is to compare customer preference with the focal customer from an already known database. The last step is to recommend the customer various products based on their reviewing. The VAR system uses viola-jones [20] face detectors to detect a face in each frame. To achieve the best result, only frontal face image frames are selected and processed to the next steps. After detecting the face from each frame, the face is normalized in illumination and size which will be sent for feature extraction. The support vector machine (SVM) [21] classifier is used to classify facial expressions. The expression falls into three schemas as positive, negative, and neutral. The volunteers for this experiment were asked to enact each type of facial expression being happy, unhappy, and neutral three times while looking at the camera. Then they were asked to read the given sentence to capture their facial reaction naturally. The given sentence was purposely structured to trigger a specific reaction from the volunteers. In the last step, volunteers were asked to make a particular expression again. All the steps are recorded as a video clip. Each video clip was then extracted frame by frame and the best image of each expression was manually picked. The accuracy rates for positive, neutral, and negative are 88%, 81%, and 68% respectively.

Le and Veal [22] developed a customer facial expression system by using Kinect sensors. They proposed two models with two Kinect sensors V1 and V2. Those sensors can track facial appearance through a three-dimensional (3D) mask. The emotion labeling is conducted of the recording of customer facial

expressions every 5 seconds from their recorded video. If the customer expressions are not matched with their expected facial expression then they level this expression as “Others”. To evaluate the sensor performance four kinds of classifiers are being used such as Bagging, J48, random forest, and random committee. The highest success rate for both Kinect V1 and Kinect V2 is achieved using random committee with 93.09% and 87.44%.

Nakano and Kato [23] developed a system to estimate customer expectation and satisfaction level using a three-dimensional CNN. From the customer's facial expressions and body motions, the system evaluates the expectation and satisfaction level of the customer. They introduced perception correction hypotheses (PCH) to evaluate the marketing from consumer satisfaction level. PCH conduct voting for customers' expectation and satisfaction level by the histogram. The histogram uses marketing evolution as a class and customers number as a degree. The experiment of the proposed system provided 78.8% accuracy for customer expectation and 69.7% is obtained from the satisfactory result.

Moulay *et al.* [24] used a patterns of oriented edge magnitudes (POEM) descriptor and SVM classifier to recognize a customer's facial expression. They focus on customer facial expressions on a specific product. The approach system is evaluated using Japanese female facial expressions (JAFFE) [25] data set. JAFFE data set consists of 213 images with seven basic expressions. The faces are taken from 10 Japanese models. Based on the different facial expressions, the dataset is divided into 3 classes: neutral, satisfied, and unsatisfied. The result obtained from this experiment is 98.39% for neutral, 98.5% for not satisfied, and 97.05% for satisfied.

Sandeep developed a model to detect customer facial expression by using neural network [26] on a VGG-16 architecture. The VGG-16 architecture [27] can process large-scale image recognition and extract features. S. K. Ramani [26], the CNN network is considered as the baseline of the system and it uses the Google cloud platform vision application programming interface (API) for facial expression detection which could also help to minimize the computational time for processing the images. To capture the corresponding expression a camera is attached with Rasberry Pi which records the customer's facial expression every five seconds. The input images are used to train the vision API and CNN model to classify images into seven expression categories. The softmax function is worked as an output layer in this network. After training the network, responses of the model will be stored in an array. Since weight is considered as a time-dependent function, the final weight will be used as the final score for the feedback. The score is then normalized in a range of -1 to 1. If the value is obtained as 1 then the result is considered as happy, if -1 then the person not happy with the service provided. On the other hand, if 0 is then considered neutral.

To observe the performance of the proposed model, The M&M initiative (MMI) [28] database containing a series of videos of people showing emotions is used. 20 images are selected from each video by changing the timeframe to a single image by repeating over the variances by applying the highest filter from reducing the filter size until 20 frames are generated. Average accuracy of 86.6% is obtained. 20 most representative images are selected by minimal changes in the timeframe as a single image by iterating over the differences using a maximum filter with decreasing filter size until the input image frames are obtained.

Lu *et al.* [19], a conventional computer vision technology is being used which is not capable of dealing with erroneous inputs and parameters. In the system, the customer preferences are calculated by matching the region of interest and grouped them into similar region expressions. Due to the limitation of the system, it only analyzes 10 frames for each second of 30 frames-per-second recording which produces erroneous inputs in this step. The errors produced in this stage cannot be resolved as a CNN network. Le and Veal [22], Moulay *et al.* [24], the systems are developed using a multiclass classifier. Since the systems are developed using a multiclass classifier, it requires more computational time in the training process. Nakano and Kato [23], a PCH classifier is introduced in the training process with a low accuracy rate. The model in [26] can process a large scale of data by using 3rd party API which is not customizable and needs to pay for each time training.

3. RESEARCH METHODS

3.1. Data collection and pre-processing

Images for three types of facial expressions (neutral, happy and sad) are collected in real-time. The main camera is placed in front of the subjects to enable the real-time recording of the facial expressions of the human face while they are observing a specific product. A digital camera is aimed directly at the subject's face to observe their expression. The camera is paired with a linux-based operating system with 8GB RAM and Intel i5 Processor. The image sequences are recorded 30 frames per second. A total number of 53 subjects participated in the experiment and recorded their three types of facial expressions. Each video is recorded between 10-12 sec. Then, a video containing a specific product slideshow is shown to the subjects as shown in Figure 1. The subject's spontaneous expression is captured while they observed the products.

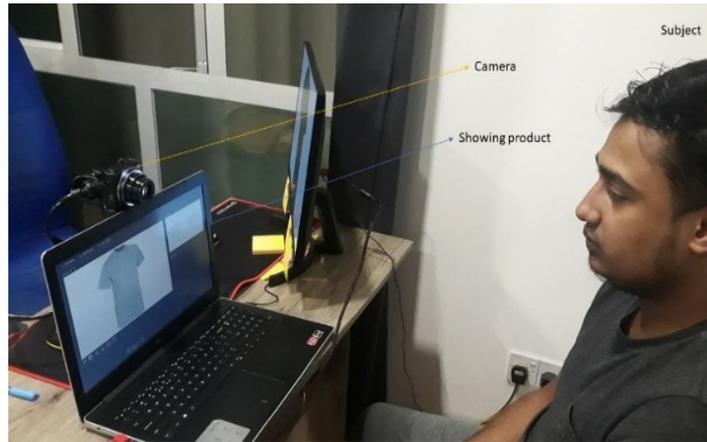


Figure 1. Data gathering setup

The whole experiment was conducted with medium illumination and posed of the frontal profile so that the system can be evaluated in different conditions. In this experiment, only spontaneous expression was recorded which refers to low-intensity facial expressions. Our proposed model is developed using two different classifiers which are SVM [29] and Softmax [30]. These models are trained using the FER2013 dataset and collected real-time facial expression images in happy, sad, and neutral categories as shown in Figures 2 and 3. The FER2013 dataset consists of blurry images and some of the images are not aligned properly [31].



Figure 2. Data set examples–sad face



Figure 3. Data set examples–happy face

CNN is one of the class of artificial neural networks which have proven an effective and more efficient term in the area of image recognition and classification. In our experiment, we propose two models as CNN-SVM (model A) and CNN-Softmax (model B) as shown in Figures 4 and 5. Both models consist of a convolutional layer, pooling layer, rectified-linear fully (also known as ReLU layer), connected layer, and classification layer [32]. The only difference in the classification layers where model A uses SVM and model B uses SoftMax as the classifier.

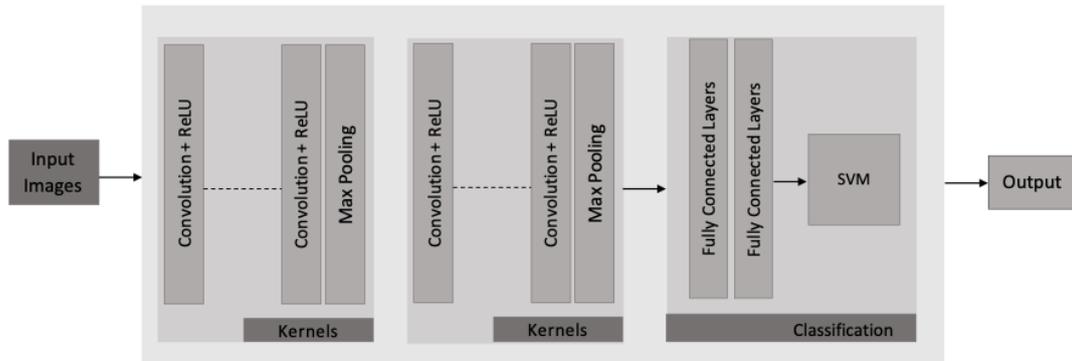


Figure 4. CNN-SVM (model A)

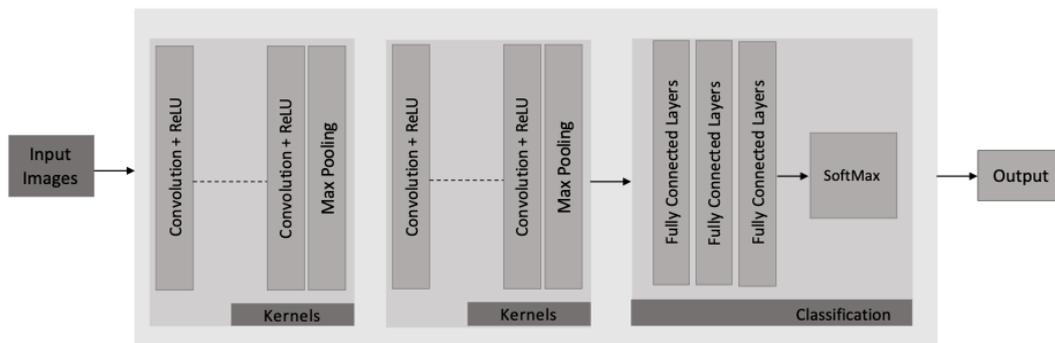


Figure 5. CNN-SoftMax (model B)

Convolutional layers are responsible to calculate the convolution spatially of a group of neurons along with filters that can be learnable [33]. The feature maps are created by calculating the dot product of the filter weights along with the input neurons that are responsible to pass to the next corresponding layer for computations [34]. Each layer is formed based on kernel size, depth, stride, and padding. The kernel size refers to the spatial size of the convolutional filter. Depth represents the depth of the filter; stride refers to the size of the slides across a local region of the input data and padding regulates the edged padded value. The pooling layer is responsible to decrease the spatial dimension of the input [35]. The process is obtained by performing a sliding window on a single depth slice. This technique is used to compute the average or maximum value of the dimensional location. The ReLU layer is responsible to remove the negative value from the input data and replace by the value '0' [36]. In the fully connected layer, the neurons' spatial size is decreased to 1*1 matrix. The classification layer is responsible to measure the predicted probabilities and label them into respective classes. However, our model A uses a multiclass classifier SVM instead of the softmax function developed by Vapnik [37]. The output of the result will be translated into case $y \in \{-1, +1\}$ as in [38].

$$\min \frac{1}{p} \|w\|_2^2 + C \sum_{i=1}^p \max(0, 1 - y'_i (w^T x_i + b))^2 \quad (1)$$

$$S_n(Z) = \frac{e^{z_n}}{\sum_{i=1}^N e^{z_i}} \quad (2)$$

In model B, the softmax function is implemented to classify the multiple classes. In this experiment, the softmax function is used because the softmax is well known normalized function to react with the blurry images rather than other standard normalization. In (2) refers the SoftMax function [39]. Once the CNN network is trained, our proposed system is tested with the trained CNN network and the results are as shown in Figures 6 and 7.



Figure 6. Real-time facial expression experiment- sad face

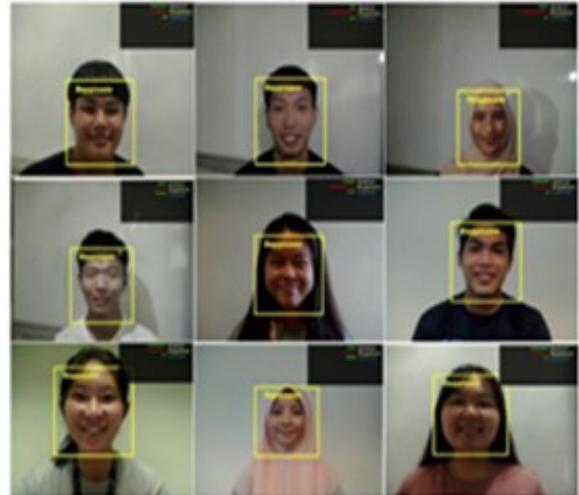


Figure 7. Real-time facial expression experiment- happy face

3.2. Framework

The proposed method is developed using the CNN network and trained using only happy, sad, and neutral images. The first step is image pre-processing which involves data augmentation and normalization as shown in Figure 8. The data augmentation involves scaling and rotation and the normalization process is to reduce the illumination and pose of the input image. The face is detected using Viola-Jones [20] algorithm detection then cropped and normalized as 64x64 pixels. Then the facial parts are extracted and resized as 32x64 pixels. After the preprocessing step, the images are fed to the CNN network to classify the images into three categories as happy, sad, and neutral. The main concept of CNN is that it extracts facial information through its weight and field sharing process. The number of trained parameters is reduced in the sharing process and the propagation of image information through the layers are being calculated by convolution in the training process. In the last stage, the back-propagation algorithm is applied to those images which provide a maximum value of 1 to detect correct facial expression. We took the maximum number 1 for the expected expression and 0 for other expressions. These values are calculated in the final layer of the softmax function. In the testing phase, each input frame will be matched with the given trained CNN model to recognize the facial expression of the customer while observing the product.

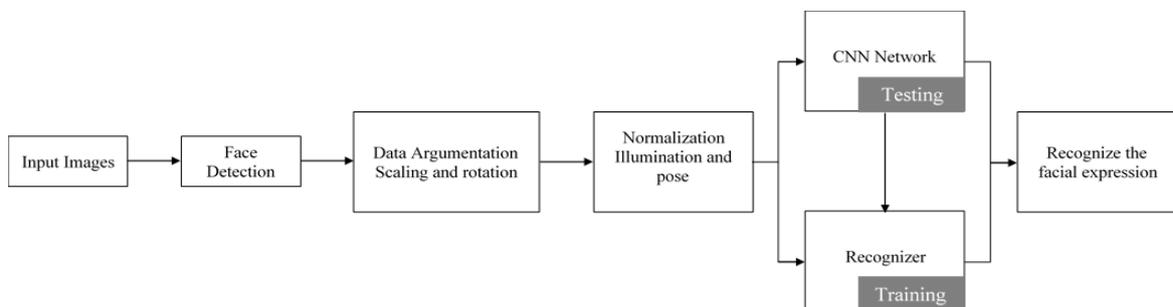


Figure 8. Overall framework

4. RESULTS AND ANALYSIS

The face landmarks and histogram of oriented gradients (HOG) features are extracted and fed into a multi-class SVM and Softmax classifier. In our experiment, we implement facial landmark and HOG feature layers in both models. Table 1 shows the classification results for model A and model B. Based on Table 1, model B gives a higher accuracy result in comparison to model A in all experiments. The training time for the obtained result is 5021 seconds on core i5 and Linux-based operating system. Figure 9 shows the accuracy result of model A and model B. The x-axis represents the experiment type and the y-axis represents the accuracy rate of models A and B. Based on Table 1, model A's accuracy is slightly lower compared to model B by 4.5%. This result differs because in CNN overfitting issue and fully connect layer unable to extract the information from previous layers [30]. Besides, adding more features for each model can improve the accuracy rate. In our model B, one extra fully connected layer was added to improve the accuracy level. Based on Figure 9, we choose model B in the next experiment since it provides more accuracy than model A. The accuracy percentage is calculated by using (3). From Table 2, it can be concluded that the happy expression has the highest success rate which is 89% and the neutral expression which success rate is 84% and the lowest success rate is the sad expression which is 75%.

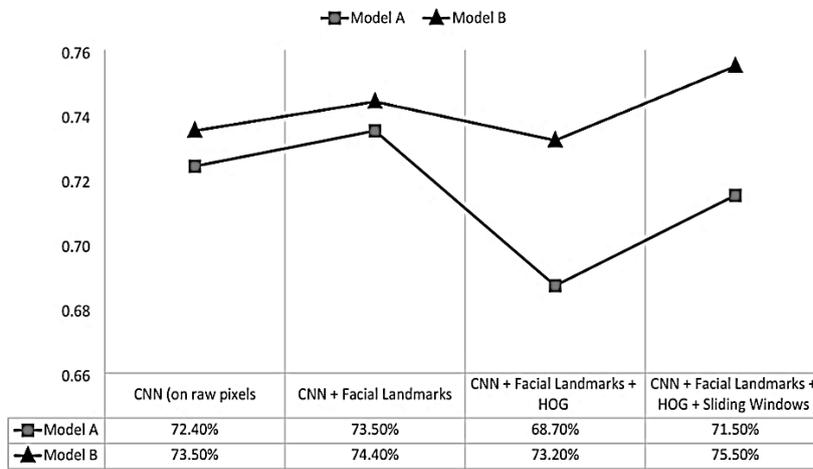


Figure 9. Comparison between model A and model B

Table 1. Classification result for model A and model B

Experiments	Model A	Model B
CNN	72.4%	73.5%
CNN + Facial Landmarks	73.5%	74.4%
CNN + Facial Landmarks + HOG	68.7%	73.2%
CNN + Facial Landmarks + HOG + Sliding Windows	71.5%	75.5%

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Table 2. Confusion matrix for happy, sad and neutral expression

Facial Expression		Predicted	
Happy	Actual	True Negative - 6	False Positive-4
		False Negative- 2	True Positive-41
Sad	Actual	True Negative - 12	False Positive-3
		False Negative- 10	True Positive-28
Neutral	Actual	True Negative - 8	False Positive-5
		False Negative- 3	True Positive-37

Table 3 shows the accuracy level of the customer's three kinds of facial expressions for different proposed models. Based on the table the available classifier is being used to classify the customer facial expression into different categories. Lu *et al.* [19], customer reaction is being observed to capture the

customer preferences and achieved success rates for positive, neutral, and negative are 88%, 81%, and 68% respectively. Nakano and Kato [23], the model is developed using the CNN network and achieved 78.8% accuracy for customer expectation and 69.7% for satisfaction. In their model, the PCH classifier was applied in their training process to obtain the result. Le and Vea [22], Kinect based customized feature extraction technique is used in their model. VGG-16 architecture is being used with a CNN-based feature extraction technique and achieved an average success rate of 86.6% [26]. In our proposed model, a fully multi-connect layer is applied in the training process with softmax function which helps the model to obtain its highest accuracy result from the given input and provide a success rate of 89 % as happy, 75% as sad, and 84% as neutral. Based on these results, it is proven that our proposed model has a better performance compared to the existing studies mentioned in Table 3.

Table 3. Comparison between other models

System	Classifier	Feature extraction	Accuracy (%)
A video-based automated recommender (VAR) system for garments [19].	SVM	Local binary pattern (LBP)	Happy-88%, neutral-81% and sad 68%
Potentiality of 3D convolutional neural networks to estimate customer expectation and satisfaction [23].	Perception correction hypotheses (PCH)	CNN Based feature extraction	Average accuracy level 69.7%
A customer emotion recognition through facial expression using kinect sensors v1 and v2 [22]	Bagging, J48, Random Forest, and Random Committee	Kinect based customized feature extraction. For sensor kinect v1 using C# and kinect v2 using c++	Average accuracy for sensor V1 Bagging-90.45% J48-88.54% Random forest-91.87% Random committee-93.089% Average accuracy for sensor V2 Bagging-83.05% J48-79.38% Random forest-85.32% Random committee-87.44%
Facial expression detection using neural network for customer-based service [26]	Softmax	CNN based feature extraction.	Average accuracy rate 86.6%
Proposed method	Softmax	Face landmark, HOG and sliding window	Happy - 89 %, neutral - 84%, sad - 75%

5. CONCLUSION

The main purpose of this research is to recognize human expression in consumer science and compare it to the existing methods in the facial expression field of study. Currently, the system is limited to recognize only three kinds of facial expressions as happy, sad, and neutral. For future work, our studies will be extended to more expressions. In our experiment, we achieved 89%, 75%, and 84% accuracy results in happy, sad, and neutral expressions which can be improved by using a multi-class classifier in the training process.

ACKNOWLEDGEMENTS

The authors fully acknowledged Universiti Malaysia Sarawak for the support which makes this important study viable and effective.

REFERENCES

- [1] R. Zhi, X. Hu, C. Wang, and S. Liu, "Development of a direct mapping model between hedonic rating and facial responses by dynamic facial expression representation," *Food Research International*, vol. 137, Nov. 2020, doi: 10.1016/j.foodres.2020.109411.
- [2] S. S. Samant, M. J. Chapko and H. Seo, "Predicting consumer liking and preference based on emotional responses and sensory perception: A study with basic taste solutions," *Food Research International*, vol. 100, pp. 325-334, Jul. 2017, doi: 10.1016/j.foodres.2017.07.021.
- [3] S. G. Moore, "Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions," *Journal of Consumer Research*, vol. 42, no. 1, pp. 30-44, 2015, doi: <https://doi.org/10.1093/jcr/ucv003>.
- [4] M. D. Rocklage., and R. H. Fazio, "The enhancing versus backfiring effects of positive emotion in consumer reviews," *Journal of Marketing Research*, vol. 57, no. 2, pp. 332-352, Feb. 2020, doi: 10.1177/0022243719892594.

- [5] N. Hussain, H. Ujir, I. Hipiny and J. L. Minoi, "3D facial action units recognition for emotional expression," *International Journal of Recent Technology and Engineering*, vol. 8, no. 28, pp. 1317-1323, 2019.
- [6] L. J. Harrison-Walker, "The effect of consumer emotions on outcome behaviors following service failure," *Journal of Services Marketing*, vol. 33, no. 3, pp. 285-302, 2019, doi: 10.1108/JSM-04-2018-0124.
- [7] H. Ujir and M. Spann, "Surface normals with modular approach and weighted voting scheme in 3D facial expression classification," *International Journal of Computer and Information Technology*, vol. 3, no. 5, pp. 909-918, Sep. 2014.
- [8] P. Tarnowski., M. Kolodziej, A. Majkowski and R. J. Rak., "Emotion recognition using facial expressions," in *ICCS*, pp. 1175-1184, Jun. 2017, doi: 10.1016/j.procs.2017.05.025.
- [9] J. Yuan., S. McDonough., Q. You., and J. Luo, "Sentribute: image sentiment analysis from a mid-level perspective," in *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 2013, pp. 1-8, doi: 10.1145/2502069.2502079.
- [10] G. Yolcu., I. Oztel, S. Kazan, C. Oz and F. Bunyai, "Deep learning-based face analysis system for monitoring customer interest," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 237-248, 2020, doi: 10.1007/s12652-019-01310-5.
- [11] H. Zhang., A. Jolfaei, and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing," *IEEE Access*, vol. 7, pp. 159081-159089, Oct. 2019, doi: 10.1109/ACCESS.2019.2949741.
- [12] P. M. A. Desmet, "Faces of product pleasure: 25 positive emotions in human-product interactions," *International Journal of Design*, vol. 6, no. 2, Dec. 2012.
- [13] M. S. Balaji and A. Quazi, "Customers' emotion regulation strategies in service failure encounters," *European Journal of Marketing*, vol. 51, no. 5/6, pp. 960-982, 2017, doi: 10.1108/EJM-03-2015-0169.
- [14] A. D. Gavrilov, A. Jordache, M. Vasdani and J. Deng, "Preventing model overfitting and underfitting in convolutional neural networks," *International Journal of Software Science and Computational Intelligence (IJSSCI)*, vol. 10, no. 4, pp. 19-28, Dec. 2018, doi: 10.4018/IJSSCI.2018100102.
- [15] K. Mohan, A. Seal, O. Krejcar and Yazidi, "A. FER-net: Facial expression recognition using deep neural net," *Neural Computing and Applications*, pp. 1-12, 2021, doi: 10.1007/s00521-020-05676-y.
- [16] T. Vijayakumar., "Comparative study of capsule neural network in various applications," *Journal of Artificial Intelligence*, vol. 1, no. 1, pp. 19-27, 2019, doi: 10.36548/jaicn.2019.1.003.
- [17] C. H. Sampaio, W. J. Ladeira, and F. D. O. Santini, "Apps for mobile banking and customer satisfaction: A cross-cultural study," *International Journal of Bank Marketing*, vol. 35, no. 7, pp. 1133-1153, 2017, doi: 10.1108/IJBM-09-2015-0146
- [18] K. An, M. K. Hui and K. Leung, "Who should be responsible? Effects of voice and compensation on responsibility attribution, perceived justice, and post-complaint behaviors across cultures," *International Journal of Conflict Management*, vol. 12, no. 4, pp. 350-364, 2001, doi: 10.1108/eb022863.
- [19] S. Lu, L. Xiao and M. Ding, "A video-based automated recommender (VAR) system for garments," *Marketing Science*, vol. 35, no. 3, pp. 484-510, 2016, doi: 10.1287/mksc.2016.0984.
- [20] N. T. Deshpande and S. Ravishankar, "Face detection and recognition using Viola-Jones algorithm and fusion of PCA and ANN," *Advances in Computational Sciences and Technology*, vol.10, no. 5, pp. 1173-1189, 2017.
- [21] S. Kumar, S. Singh and J. Kumar, "Multiple face detection using hybrid features with SVM classifier," in *Data and Communication Networks*, pp. 253-265, 2019, doi: 10.1007/978-981-13-2254-9_23.
- [22] H. T. Le and L. A. Vea, "A customer emotion recognition through facial expression using kinect sensor v1 and v2: A comparative analysis," in *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication*, 2019, pp. 1-7, doi: 10.1145/2857546.2857628.
- [23] T. Nakano and S. Kato, "Potentiality of 3D convolutional neural networks to estimate customer expectation and satisfaction," *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, 2017, pp. 1-4, doi: 10.1109/GCCE.2017.8229391.
- [24] M. S. Bouzakraoui, A. Sadiq, and N. Enneya, "A customer emotion recognition through facial expression using POEM descriptor and SVM classifier," in *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, 2017, pp. 1-6, doi: 10.1145/3090354.3090436.
- [25] M. J. Lyons, M. Kamachi, and J. Gyoba, "Japanese female facial expressions (JAFFE)," *Database of digital images*, vol. 3, 1997.
- [26] S. K. Ramani, "Facial expression detection using neural network for customer based service," in *2018 International Conference on Computer, Communication, and Signal Processing (ICCCSP)*, pp. 1-4, Feb. 2018, doi: 10.1109/ICCCSP.2018.8452864.
- [27] A. Caroppo, A. Leone, and P. Siciliano, "Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults," *Journal of Computer Science and Technology*, vol. 35, no. 5, pp. 1127-1146, 2020, doi: 10.1007/s11390-020-9665-4.
- [28] M. F. Valstar and M. Pantic., "Induced disgust, happiness and surprise: an addition to the MMI facial expression database," in *Proceedings of International Conference Language Resources and Evaluation, Workshop on EMOTION*, 2010, pp. 65-70.
- [29] X. Qi, T. Wang and J. Liu, "Comparison of support vector machine and Softmax classifiers in computer vision," in *2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 2017, pp. 151-155, doi: 10.1109/ICMCCE.2017.49.

- [30] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82-92, Aug. 2019, doi: 10.1016/j.neucom.2019.05.005.
- [31] Giannopoulos, Panagiotis, I. Perikos and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," *Advances in Hybridization of Intelligent Methods*, pp. 1-16, 2018, doi: 10.1007/978-3-319-66790-4 1.
- [32] J. Van Kleef, "Towards human-like performance face detection: A convolutional neural network approach," *University of Twente*, 2016.
- [33] G. Lin and W. Shen, "Research on convolutional neural network based on improved ReLu piecewise activation function," *Procedia computer science*, vol. 131, pp. 977-984, 2018, doi: 10.1016/j.procs.2018.04.239.
- [34] N. Ma, X. Zhang, H. T. Zheng and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 116-131, doi: 10.1007/978-3-030-01264-9 8.
- [35] S. Ren, K. He, R. Girshick, X. Zhang and J. Sun, "Object detection networks on convolutional feature maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, 1476-1481, 2016, doi: 10.1109/TPAMI.2016.2601099.
- [36] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," *arXiv preprint arXiv: 1803.08375*, 2018.
- [37] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.
- [38] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2528-2536, 2016, doi: 10.1109/TMM.2016.2598092.
- [39] L. Chen, M. Zhou, W. Su, M. Wu, J. She and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Information Sciences*, vol. 428, pp. 49-61, 2018, doi: 10.1016/j.ins.2017.10.044.

BIOGRAPHIES OF AUTHORS



Golam Morshed is currently working as a Chief Technology Officer at HMN group of companies and Senior Software Engineer at AccountSoft Enterprise Sdn Bhd. He graduated from Universiti Malaysia Sarawak in 2018 with a bachelor's degree in computer science (Software Engineering).



Hamimah Ujir is currently a Senior Lecturer at FCSIT, UNIMAS. She received her PhD from University of Birmingham, United Kingdom in 2013. Hamimah's research interests lay in the interdisciplinary field of computer vision, with related interests being in computer graphics, image processing, and mathematical methods. Her previous and current works include 3D physical simulation and 3D static and dynamic facial expression analysis. She also conducted research works on academic quality in higher education.



Irwandi graduated with a PhD in Computer Vision from University of Bristol back in 2014. He also holds an MSc (2007) and a BSc (2003) from Universiti Teknologi Malaysia. He is currently serving as the Deputy Dean of Undergraduates at the Faculty of Computer Science & Information Technology, Universiti Malaysia Sarawak. His research interests are, but not limited to, Computer Vision, Pattern Matching and Augmented Reality. More specifically, he is actively pursuing research grants/postgraduates on the topics of Visual Animal Biometric and the use of Augmented Reality (AR) for Assisted Living/Education.