Video Game Review using Facial Emotion Recognition

R Deva Dharshni[1], Sreedevi M. G[2], Sreelakshmi M. G[3], Sreelekshmi Raj[4], Neethu M. S[5]

[1, 2, 3, 4, 5] Dept of Computer Science and Engineering

[1, 2, 3, 4, 5] LBS Institute of Technology for Women, Thiruvananthapuram, Kerala.

Abstract—**Video games are designed keeping in mind a specific target audience. Each video game aims to evoke a particular behavior and set of emotions from the users. In video gaming, during the testing phase, users are asked to play the game for a given period and their feedback is incorporated to make the final product. Use of facial emotion detection can aid in understanding which emotions a user is going through in real-time as he is playing without analyzing the complete video manually. Such product feedback can be taken by analyzing a live feed of the user and detecting his facial emotions. While feelings of frustration and anger are commonly experienced in advanced video games, making use of facial emotion detection will help understand which emotions are experienced at what points in the game. It is also possible that some unexpected or undesirable emotions are observed during the game. Taking a written feedback from the user who has experienced the game can be inefficient. This is because it can often be difficult to put an experience into words. Moreover, users may be unable to remember what exactly they went through emotionally across different parts of the game. Facial Emotion detection is a practical means of going beyond the spoken or written feedback and appreciating what the user is experiencing. When feedback is taken in this format, it becomes genuinely non-intrusive when it comes to user experience.**

Keywords—HAAR Cascade, CNN, FER2013, Matplotlib.

I. INTRODUCTION

Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time. Machine learning algorithm can detect emotions by learning what each facial expression means and applying that knowledge to the new information presented to it. Understanding contextual emotion has widespread consequences for society and business. In the public sphere, governmental organizations could make good use of the ability to detect emotions like guilt, fear, and uncertainty. Disney plans to use facial recognition to judge the emotional responses of the audience. Facial emotions are important factors in human communication that help us understand the intentions of others. According to different surveys, verbal components convey one-third of human communication, and nonverbal components convey two-thirds. Among several nonverbal components, by carrying emotional meaning, facial expressions are one of the main information channels in interpersonal communication. Organizations and businesses can use facial emotion to understand their customers and create products that people like. This paper focuses on how facial emotions of a person can be used in reviewing a game and this information can help the game developer to provide a product that people like.

For facial emotion recognition we use Convolutional Neural Network (CNN) and FER2013 dataset to train and test the model

II. LITERATURE REVIEW

As a part of literature review, the existing systems are compared with each other and various methods used in facial emotion recognition are identified. Several methods are used in identifying the face, extracting the facial features such as eyes, nose, lips, etc. and identifying the various emotions. These methods use different methodologies which includes machine learning algorithms such as Classification and Regression Tree (CART) and Support Vector Machine (SVM) to classify the facial emotions.

[Shivam Gupta, 2018] presents a fully automatic recognition of facial emotions using machine learning algorithm (Support Vector Machine) to classify eight different emotions. Cohn-Kanade Database (CK) and the Extended Cohn-Kanade (CK+) dataset are used. The dataset is divided into training and classification dataset. Face detection is done using Haar filter in OpenCV. In real-time, facial landmarks approach is used to detect emotions. Support Vector Machine (SVM) is used to classify the facial emotions.

[Fuzail Khan, 2018] proposes a framework in which after the initial facial localization is performed, facial landmark detection is done using the Sobel operator and the Hough transform followed by Shi Tomasi corner point detection. The input feature vectors are formulated using Euclidean distances and are trained into a Multi-Layer Perceptron (MLP) neural network in order to classify the expression being displayed. Dataset used is KDEF database.

[Binh. T. Nguyen, 2017] presents a potential approach to detect human emotion in real time. For any face detected in camera, we extract the corresponding facial landmarks and examine different kinds of features and models for predicting human emotion. The experiments show that our proposed system can naturally detect human emotion in real time and achieve an average accuracy about 70.65%.

[Octavio Arriaga, 2017] proposes a general convolutional neural network (CNN) building framework for designing real-time CNNs. The model is validated by creating a real-time vision system which accomplishes the task of face detection, gender classification and emotion classification

simultaneously in one blended step using our proposed CNN architecture. For emotion detection FER2013 dataset is used.

[Ju Jia, Yan Xu, Sida Zhang, Xianglong Xue, 2016] uses random forest as the classifier based on the features extracted from improved principal component analysis (2DPCA). Drawbacks of Traditional PCA such as difficulty in estimating the covariance matrix and computational prohibition to get eigen vectors are eliminated by using improved PCA that is 2DPCA. JAFFE Database is a static expression database, but facial expression is a series of action. Due to these reasons, RF show high accuracy.

[Fatima Zahra Salmam, Abdellah Madani, 2016] new extraction method is presented based on the geometric approach which consists of calculating six distances in order to measure parts of the face that better describe a facial expression. Decision tree is applied on two databases (JAFFE and COHN), in order to have a facial expression classifying system with 7 possible classes; this system uses as input the six distances (using Euclidian, Manhattan or Minkowski distance) for each face.

III.   DATASET AND METHOD

A. Dataset

In our proposed system, we use FER2013 as the dataset. The database was created using the Google image search application programming interface (API) and faces have been automatically registered. Faces are labeled as any of the six basic (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral) expressions as well as the neutral. The resulting database contains 35,887 images, most of them in wild settings. Fer2013 consists of 48x48 pixel gray-scale images of faces. Fer2013.csv file much efficient in training. Training file contains two columns, "emotion" and "pixels". The "emotion" column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The "pixels" column contains a string surrounded in quotes for each image. The contents of this string are space-separated pixel values in row major order. Testing file contains only the "pixels" column. The training set consists of 28,709 examples. The test set consists of 3,589 examples.
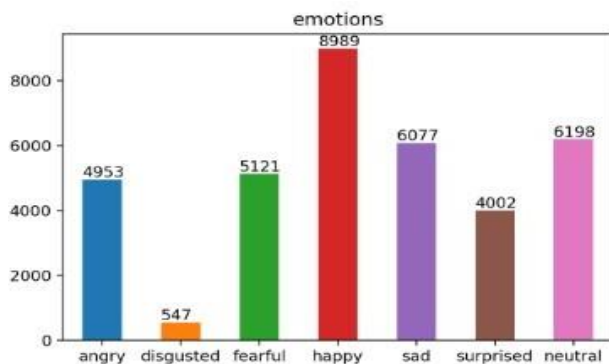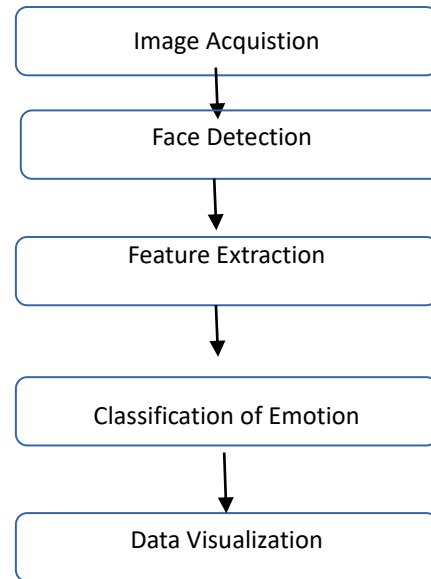

Fig 1: Emotion Distribution FER2013 Dataset

B. Steps

a. Pre-processing

 The dataset consisted of a number of images represented as strings of 2304 space separated numbers which were then converted to a 48*48 matrix. Each number represented a pixel value. The original data of 35,887 images was split into a training set of 28,709 images and a testing set of 7,178 images- an 80:20 split. Generally, when it comes to deep learning, data is the biggest factor. The bigger the training set, the better the output. If there is less training data, there is a lot more variance in the final outputs due to a smaller set to train on. Bearing that in mind, having a testing set of 20% of the total images may be seen as excessive. However, to prevent overfitting it is necessary to have a sizable testing set as well.

b. Image Acquisition

 A web-cam is used to record the user's facial expressions while playing the video game. Images used for facial expression recognition are frames taken from the video. The web-cam on the computer records the real-time video and faces will be detected in each frame of the webcam video and then the further processing will be done on those detected faces.

c. Face Detection

  The faces will be detected in each frame of the webcam video and then the further processing will be done on those detected faces. In our proposed system we use HAAR Cascade classifier to detect face in each frame. Then after that, the image is processed by converting it to grayscale and resize to 48x48 resolution, this improves the efficiency of classifier

d. Feature Extraction

  The face which is extracted from the frame in face detection phase is then converted into a list of pixels. After that, pixel values are normalized by dividing each value by

255. Normalized pixel values are given to the classifier for prediction.

e. Classification of Emotion

In our proposed system, we use CNN (Convolutional Neural Network) for classifying each frame to one of seven emotion categories which include neutral, happy, sad, fear, surprise, angry and disgust. CNN is very effective in image recognition and classification compared to a feed-forward neural network. This is because CNN allows to reduce the number of parameters in a network and take advantage of the spatial locality. Further, Convolutional neural networks introduce the concept of pooling to reduce the number of parameters by down sampling. Applications of Convolutional neural networks include image recognition, self-driving cars and robotics. CNN is popularly used with videos, 2D images, spectrograms, Synthetic Aperture Radars.

The below Mini_Xception architecture is used for classifying emotion on this data-set. The below architecture was proposed by Octavio Arragia [3].
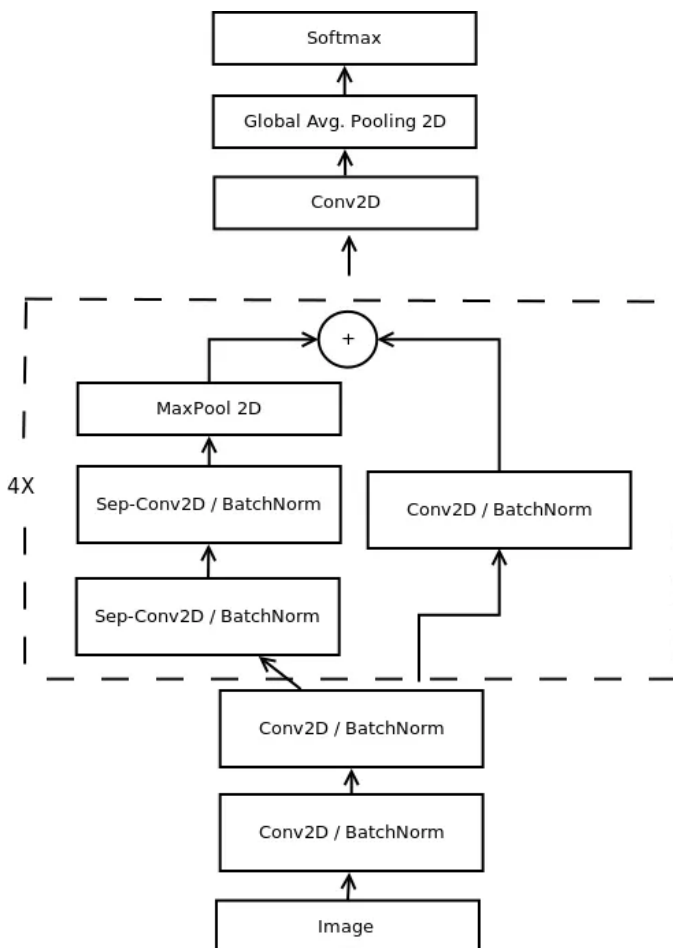

Fig 2: Mini_Xception architecture for emotion classification

Mini_Xception architecture is a fully-convolutional neural network that contains 4 residual depth-wise separable convolutions where each convolution is followed by a batch normalization operation and a ReLU activation function. The last layer applies a global average pooling and a soft-max activation function to produce a prediction. Modern CNN architectures such as Xception leverage from the combination of two of the most successful experimental assumptions in CNNs: the use of residual modules and depth-wise separable convolutions. There are various techniques that can be kept in mind while building a deep neural network and is applicable in most of the computer vision problems. Below are few of those techniques which are used while training the CNN model below:

1. Data Augmentation: More data is generated using the training set by applying transformations. It is required if the training set is not sufficient enough to learn representation. The image data is generated by transforming the actual training images by rotation, crop, shifts, shear, zoom, flip, reflection, normalization etc.

2. Kernel-regularizer: It allows to apply penalties on layer parameters during optimization. These penalties are incorporated in the loss function that the network optimizes. Argument in convolution layer is nothing but L2 regularization of the weights. This penalizes peaky weights and makes sure that all the inputs are considered.

3. Batch Normalization: It normalizes the activation of the previous layer at each batch, that is, applies a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1. It addresses the problem of internal covariate shift. It also acts as a regularizer, in some cases eliminating the need for Dropout. It helps in speeding up the training process.

4. Global Average Pooling: It reduces each feature map into a scalar value by taking the average overall elements in the feature map. The average operation forces the network to extract global features from the input image.

5. Depth-wise Separable Convolution: These convolutions are composed of two different layers: depth-wise convolutions and point-wise convolutions. Depth-wise separable convolutions reduce the computation with respect to the standard convolutions by reducing the number of parameters.
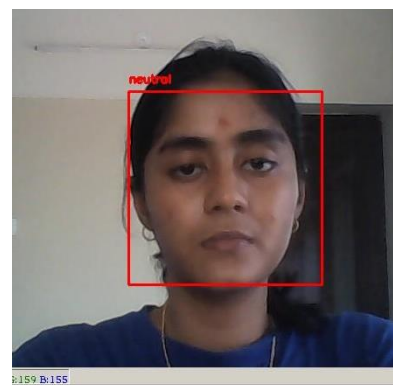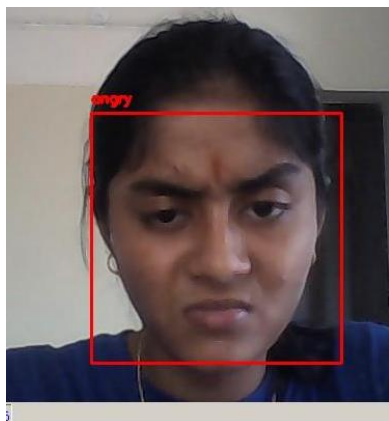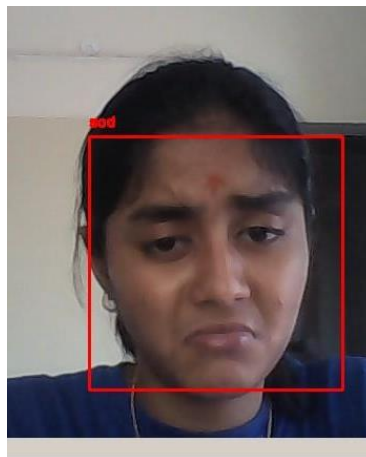
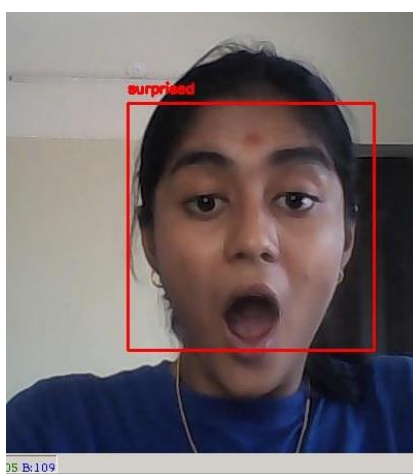
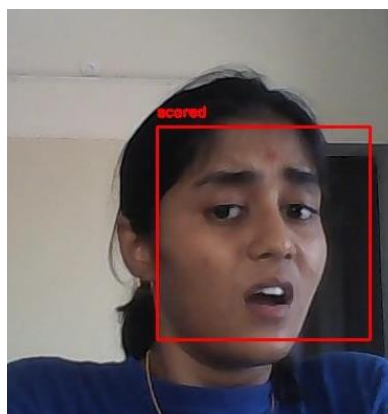
Fig 3a: Neutral

Fig 3b : Angry
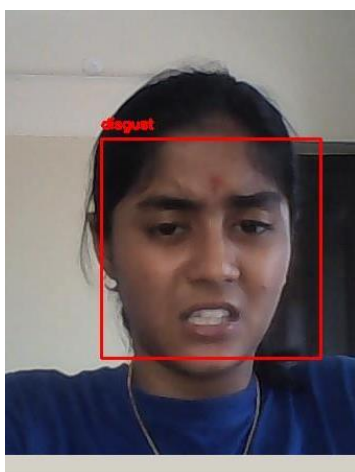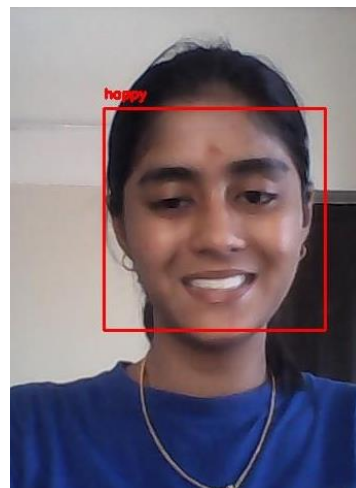

Fig 3e : Sad

Fig 3c: Surprised

Fig 3f : Scared

Fig 3d : Disgust

Fig 3g : Happy

f. Data Visualization

In order to analyze the efficiency of the game using the player emotion, we have to represent them so that we visually analyze it. In our system the detected emotion of the player is stored in separate comma-separated values (CSV) file. Later we use this value to plot graph for each different

user and analyze the graph to review the game performance and further improve the product. Python library Matplotlib provide bult-in function to plot the following graph.
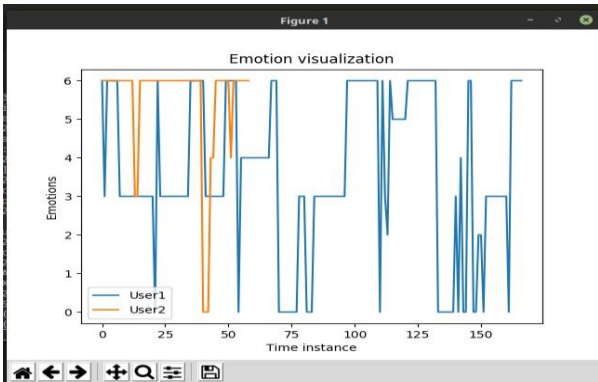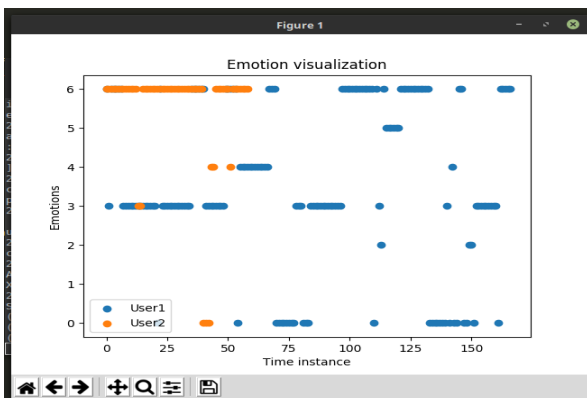

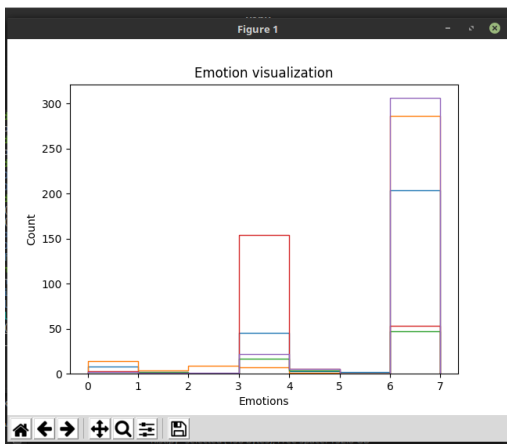Fig 4a : Normal plot of two users

Fig 4b : Scatter plot of two users

Fig 4c : Histogram

The histogram can be used to visualize the overall distribution of emotions. Here [0-1] represents count of emotion angry ,[1-2] count of emotion disgust, [2-3] count of emotion fear, [3-4] count of emotion happy, [4-5] count of emotion sad, [5-6] count of emotion surprise, [6-7] count of emotion neutral. If count of emotions neutral, sad and disgust is high compared to all other count we can say that game is not liked by majority of people and there is need to improvise

the product so that organization may not lose their customers or users.

IV. RESULTS AND DISCUSSION

The CNN model learns the representation features of emotions from the training images. While performing tests on the trained model, it was that model detects the emotion of faces as neutral if the expressions are not made distinguishable enough. The model gives probabilities of each emotion class in the output layer of trained CNN model. Emotion recognition is a complex task, more so when using real-time images. For humans this is difficult because the correct recognition of a facial emotion often depends on the context within which the emotion originates and is expressed. Emotion detected is stored in a separate file and later used to plot graph.

V. CONCLUSION

Nowadays, the feedback system used in game development is in written form which is very much inefficient as it can often be difficult to put an experience into words and cannot be relied upon to do any improvements on the game. The proposed system serves as a solution to this problem with the use of a facial emotion recognition system. Facial Emotion detection is a practical means of going beyond the spoken or written feedback and appreciating what the user is experiencing. When feedback is taken in this format, it becomes genuinely non-intrusive when it comes to user experience. At the same time, such feedback is more reliable than other forms.

REFERENCES

[1] Shivam Gupta, "Facial emotion recognition in real-time and static images", IEEE Proceedings of the Second International Conference on Inventive Systems and Control, 2018.

[2] Fuzail Khan, "Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks", 2018.

[3] Octavio Arriaga, Paul G. Ploger, Matias Valdenegro, "Real-time Convolution Neural Networks for Emotion and Gender Classification, 2017.

[4] Binh T. Nguyen, Minh. H.Trinh, Tan.V.Phan, Hien.D. Nguyen, "An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks", 7th International Conference on Information Science and Technology Da Nang, Vietnami, April 16-19; 2017.

[5] Ju Jia, Yan Xu, Sida Zhang, Xianglong Xue, "The facial expression recognition method of random forest based on improved PCA extracting feature", 2016.

[6] Fatima Zahra Salmam, Abdellah Madani, Mohamed Kissi "Facial Expression Recognition using Decision Trees", 13th International Conference Computer Graphics, Imaging and Visualization, 2016.