e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:07/July-2023 Impact Factor- 7.868 www.irjmets.com

SIMULATION OF HUMAN ROBOT INTERACTION THROUGH AVATAR

Shetty Shivani*[1], Dr. A. Sharada*[2], Mrs. Bhageshwari Ratkal*[3]

*[1]Student, Department Of Computer Science Engineering, G. Narayanamma Institute Of Technology And Science, Hyderabad, Telangana, India.

*[2]Professor, Department Of Computer Science Engineering, G. Narayanamma Institute Of Technology And Science, Hyderabad, Telangana, India.

*[3]Asst. Professor, Department Of Computer Science Engineering, G. Narayanamma Institute Of Technology And Science, Hyderabad, Telangana, India.

DOI : https://www.doi.org/10.56726/IRJMETS43128

ABSTRACT

Human-Robot Interaction (HRI) systems are extremely important in society. The quality of communication between robots and humans should be assessed to support the usefulness of robots in performing jobs that require communication, particularly for social robots. The suggested system simulates the HRI using Avatar. Avatar is defined as a user-created digital representation that reflects the user's presence in a metaverse in order to facilitate human-robot interaction. Representation encompasses not just the appearance of avatars and their surroundings, but also the methods in which avatars and their surroundings interact. The usage of interactive avatars in the medical field has the potential to give benefits, as bedridden patients feel more comfortable interacting with the avatar, and caretaker intervention is decreased. This project aims to simulate Human-Robot Interaction by taking it to a new level of interaction through Avatar, where it is used across various domains.

Keywords: Deep Learning, Human-Robot Interaction, Avatar, Facial Expression Recognition, Convolutional Neural Network.

I. INTRODUCTION

The field of Human-Robot Interaction (HRI) is concerned with the design, comprehension, and assessment of robotic systems in which humans and robots interact through communication. Communication between a human and a robot can take numerous forms, but these forms are heavily influenced by whether the human and the robot are close to each other. Although robot technology was mostly developed in the mid and late 20th centuries, it is important to remember that the concept of robot-like behavior and its consequences for humans has existed for ages in religion, mythology, philosophy, and literature. Human-robot interaction ranges from healthcare robots that allow elderly patients to have more social interaction with robots that supervise and monitor infants and babies.

Many researchers argue that more research on human involvement in virtual worlds is needed. Avatars are a crucial part of virtual worlds because they allow for user interaction and connection with the system. Avatars can take many different forms, such as reflecting the user in human forms or allowing the user to construct avatars of their choice. In comparison to robots, avatars enhance interaction with people by providing a realistic experience. It is vital to use an avatar where a human is needed in real life.

II. LITERATURE SURVEY

[1] Title: Interaction System Based on an Avatar Projected on a Pyramidal Display

Authors: David Loza Matovelle, Samuel Marcos, Eduardo Zalama, Jaime Gomez Garcia Bermejo

Description of work:

The interaction machine primarily based completely on a three-dimensional digital head projected onto a pyramidal display, which is referred to as an avatar in the virtual world for the user to experience a realistic sort of interaction. The proposed machine makes use of a social robotic behavioral structure already evolved, which permits to exchange traits among our robot realizations and the 3D avatar. It has two subsystems designed for the structuring of the avatar in the three-dimensional space. The back projection subsystem tasks a three-dimensional avatar onto a pyramidal structure. The expression generator subsystem carries out the avatar animations the usage of form keys and bones, following the Facial Action Coding System (FACS). In order to assess the expressiveness of the system, units of experiments were performed: one to research the avatar's gestural ability, that is, its functionality to carry out expressions that are recognized through way of means of an observer, and a second test to degree the emotion.

[2] Title: Identification and Validation of Biomarkers for Autism Spectrum Disorders

Authors: Eva Loth, Will Spooren, Lindsay M. Ham, Maria B. Isaac

Description of work:

A novel approach for recognizing facial expressions in children with ASD is introduced during playtime. Children are monitored while playing or using tablets or computers, as their facial expressions are tracked by researchers. An Active Shape Model (ASM) tracker was implemented, which tracks 116 facial landmarks via webcam 8 input, and the tracked landmark points are used to extract face expression features. Using a Support Vector Machine (SVM) based classifier, we have the ability to enhance our system by recognizing seven expressions rather than simply six expressions as most facial expression recognition systems do, and the system is applied to the FER2013 dataset.

III. METHODOLOGY

The proposed methodology is shown in the figure 1, in which a human face is captured using a webcam and some sample images are provided as input, and pre-processing is done to extract features from facial landmarks. The CNN model is then trained using images from the Facial Emotion Recognition (FER) dataset and the extracted features, and classification for any given image is carried out by identifying the expression from the Avatar dataset. The expression captured from the input human face is detected and from the Avatar dataset, the recognized expression is extracted out of 7 classes as the recognized avatar image output. The project is carried away in two phases, the first phase is the training phase, and the second phase is the testing phase.

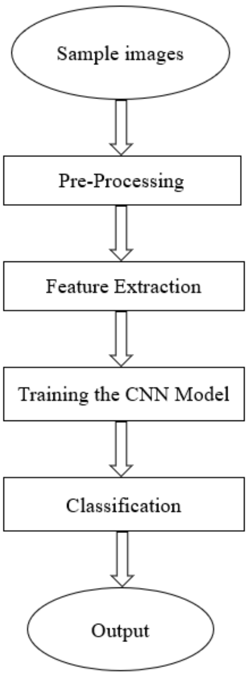

Figure 1: Flow Chart of the Proposed Methodology

IV. MODELING AND ANALYSIS

The Deep Learning model in this proposed system discusses how it is implemented in two phases: training and testing. The training phase includes taking sample images, pre-processing them to extract features using facial landmarks, and training the CNN model. The testing phase includes classification of images.

The Avatar dataset is created using PyCharm application, and by providing values to the identifier types as input, the SVG or PNG files are rendered as a dataset and loaded into the CNN model for Avatar expression recognition. The user expression is recognized by building a CNN model approach of Deep Learning where the given human input face is pre-processed. After feature extraction, the extracted facial expression from input is compared to the 7 different expression classes in the FER2013 and Avatar datasets. Finally, the recognized avatar expression is obtained as a result.

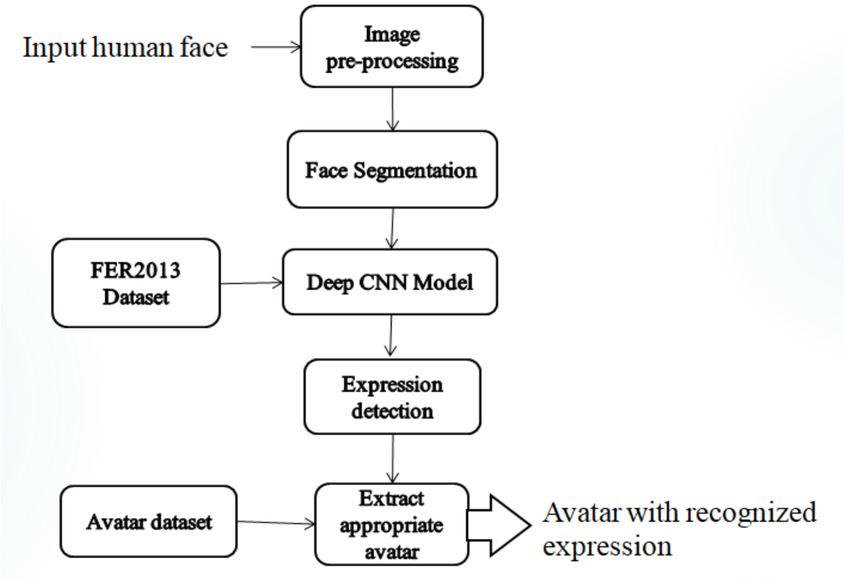

Figure 2: Process Flow of the Implemented CNN Model

Implementation:

- Datasets Used
- Convolutional Neural Network Model Approach for Expression Recognition

Datasets Used:

1. FER2013 dataset used for facial expression recognition:

- Training Samples: 28709 images
- Testing Samples: 7178 images

Facial Expression Recognition Dataset (FER 2013) was used from Kaggle. The data set consists of 2 directories:

- The train directory consists of 7 subdirectories of seven Expressions of Angry, Fear, Happy, Sad, Neutral, Surprise, Disgust. Each contain image files in jpg format. These are the files used for training the network.
- The test directory contains of 7 subdirectories labeled as above each containing image files used to evaluate the model accuracy.

2. Avatar dataset: Avatar creation and customization is done by using PyCharm software in which we have to install Pyavataaar and import pyavataaar by using the following commands:

- pip install python-avatars
- Import pyavataaar as pa

Importing the pyavataaar module, create a CLI utility to manage the library. Create the feature customization by utilizing the identifiers required for avatar design that support the module that imported into the PyCharm software. By assigning some values to those identifier types face orientation, skin color, hair type, eyes, nose type, mouth type and accessory type. Now avatar images are rendered in svg or jpg file after giving all the necessity features to the main file. Avatar images are imported and loaded to the model as avatar dataset for expression recognition for given input human face.

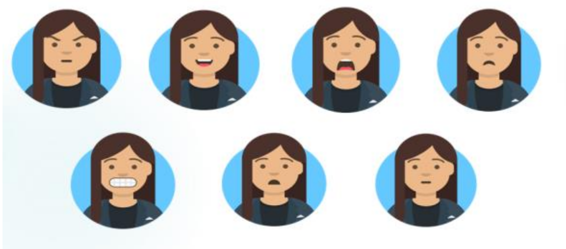

Figure 3: Avatar Images with 7 Facial Expressions

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:07/July-2023 Impact Factor- 7.868 www.irjmets.com

Convolutional Neural Network Model Approach for Expression Recognition:

The Python-based notebook called jupyter is used for building the Convolutional neural network model. Install and import the libraries Numpy, Tensorflow, Keras, CV, Scipy, and Matplotlib. By configuring the paths to the train, validate, and test folders, the data folder that contains images is neatly organized into train, valid, and test subfolders. Using the flow_from_directory () function in the Image Generator class, create batches from the train, test, and valid directories, and then deliver these batches of data to the sequential model using the fit function.

The sequential API allows create models layer-by-layer and by building the model in such way that each input image will pass it through a series of convolution layers with filters of series 32, 64, 128, 256, and 512. The kernel size is set to 3x3 as the input image is below 128x128 pixels and learning rate will be of range 0.1 to 0.0001 by using Adam optimizer. The zero-padding of the edges is achieved by using padding = 'same'.

For max-pooling layer MaxPool2D, which is mainly for reducing the dimensionality of the data is set to a window size of (pool_ size = 2x2). Finally, the output is flattened from the Conv2D layer and pass it to the Dense layer with units = 2 and SoftMax activation and prepare the model for training by using model. compile function.

A CNN model is trained for classification purpose. This model implementation has two phases.

Phase 1 - Training the Model:

- **Step 1**: The steps_per_epoch is calculated based on your training data and by-passing, validation_data = valid_batches, as the model need to pass validation_steps.
- **Step 2**: Getting features for training and Get labels for training and store them using numpy. Getting features for testing and Get labels for testing and store them using numpy. Now store all validated data from dataset and store them to .npy files and load them.
- **Step 3**: Compile and train the model by defining number of features, batch size, number of epochs.
- **Step 4**: Calculate and display the performance metrics Precision, Recall, F1-Score and Test Accuracy of the model using Confusion matrix. Save the 'model.h5' files for classification of the images.

The outcome of model training looks like this, with verbose set to the highest setting in terms of output messages:

```
Epoch 1/50
448/448 [==============================] - 201s 448ms/step - loss: 1.7057 - accuracy: 0.3093 - val_loss: 1.4492 - val_acc
uracy: 0.4438
Epoch 2/50
448/448 [==============================] - 219s 488ms/step - loss: 1.3974 - accuracy: 0.4597 - val_loss: 1.2971 - val_acc
uracy: 0.4999
Epoch 3/50
448/448 [==============================] - 235s 525ms/step - loss: 1.2685 - accuracy: 0.5181 - val_loss: 1.2096 - val_acc
uracy: 0.5449
Epoch 4/50
448/448 [==============================] - 238s 531ms/step - loss: 1.1925 - accuracy: 0.5470 - val_loss: 1.1720 - val_acc
uracy: 0.5453
Epoch 5/50
448/448 [==============================] - 241s 537ms/step - loss: 1.1373 - accuracy: 0.5708 - val_loss: 1.1185 - val_acc
uracy: 0.5784
```

Figure 4: Output from model training

Phase 2 - Testing the Model:

- **Step 1**: Load the weights which are saved by the model in the Disk model. json and model.h5.
- **Step 2**: Input the image which is pre-processed and cropped to get the facial landmarks and features.
- **Step 3**: Then the image is classified by the CNN model and image of avatar with recognized expression is displayed.

V. RESULTS AND DISCUSSION

Results of Avatar with Recognized Expression:

The recognized expressions of given input image are extracted and is shown in the following figure 5 and 6. The input human face has a happy expression which is identified by using CNN model and recognized happy

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:07/July-2023 Impact Factor- 7.868 www.irjmets.com

expression avatar image is extracted as output. Similarly, for surprised expression is also recognized and extracted as output.

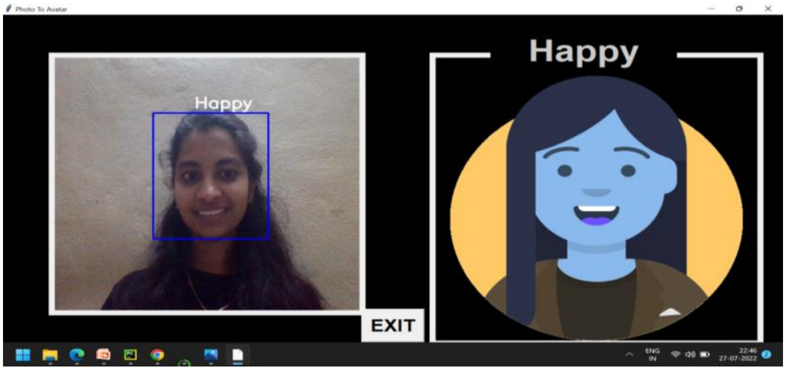

Figure 5: Avatar with Happy Expression

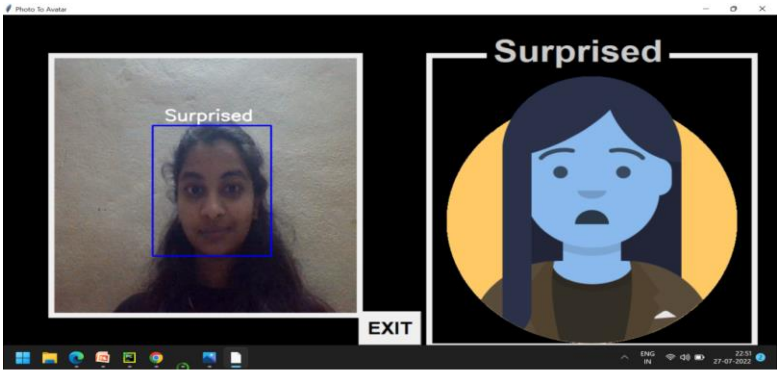

Figure 6: Avatar with Surprised Expression

Tables and Graphs:

The results in the following table show that there is a promising growth in the performance of our implementation with accuracy of 87%, when compared with other machine models.

Table 1. Validation for Accuracy

Expression	Precision	Recall	F1-Score
Anger	69.20	65.60	67.29
Disgust	83.40	76.40	79.20
Fear	65.72	56.64	60.80
Happy	89.89	92.04	91
Sad	60.90	62.66	61.70
Surprise	86	81.73	83.90
Neutral	67.90	72.25	77.16

Accuracy = No. of correct predictions / Total No. of predictions

= 0.74+0.85+0.98+0.92+0.89+0.87+0.88 / 7 = 0.87

The accuracy and loss from the model are plotted by graph and the model is trained with the parameter values LR = 0.01, batch size = 16, epochs = 50 and Adam optimizer. The following figure shows the results of one trial after training the model for 50 epochs.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:07/July-2023 Impact Factor- 7.868 www.irjmets.com

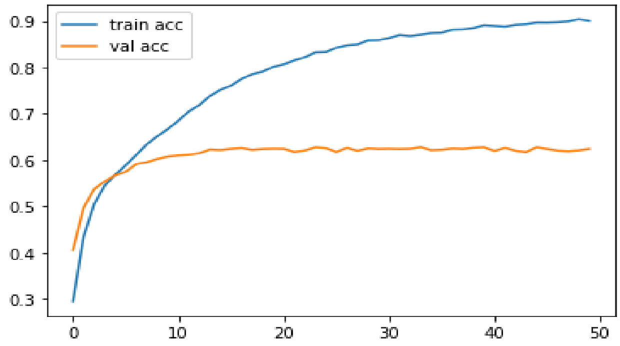

Figure 7: Graph for Accuracy

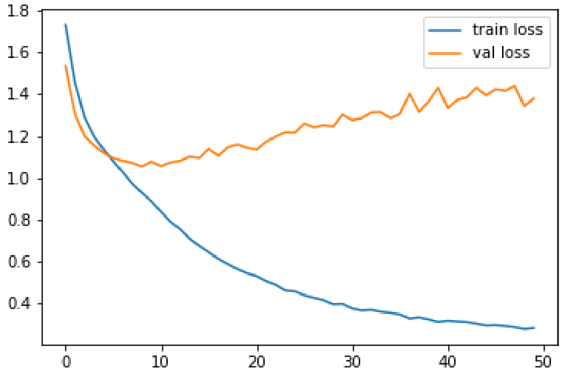

Figure 8: Graph for Loss

VI. CONCLUSION

Interaction with people through avatars, which could provide a real-life experience and enable new levels of interaction in the field of Human Robot Interaction, as well as to overcome the limitations of the existing system. For classification, the model implemented considers seven discrete and distinct expression classes (angry, disgust, fear, happy, surprise, neutral, and sad). These expressions are identified by using an avatar with the same expression as the human input face. In conclusion, the model-based interaction between humans and robots might be more convenient than chatbots.

The findings obtained as a confusion matrix indicate that facial images can be used to recognize expressions with an accuracy of 87%, and the measures of performance of the implemented model are calculated based on Precision, Recall, and F1-score.

VII. REFERENCES

[1] Akriti, A. Krishnam Raju, Suman Deb, “Facial Emotion Detection Using Deep Learning”, 2020 International Conference for Emerging Technology (INCET) Belgaum, India. June 2020.

[2] H. Azevedo, Isaque Elcio de Souza, “A Simulated Environment for Long-term Interactions”, Nucleus of Robotics and Computational Vision (NRVC), 2019.

[3] L. F. Guerrero-Vasquez, Denny’s X. Landy-Rivera “AVATAR: Contribution to Human-Computer interaction processes through the adaptation of semi-personalized virtual agents,” IEEE Biennial Congress of Argentina, June 2018.

[4] Adam Brenner, Celiktutan, H. Gunes, “Personality Perception of Robot Avatar Teleoperator in Solo and Dyadic Tasks”, Humanoid Robotics, May 2017.

[5] Abir Fathallah, Lotfi Abdi, Ali Douik, “Facial Expression Recognition via Deep Learning”, 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications.

[6] Zhentao Liu, Min Wu, Weihua Cao, Luefeng Chey, Jianping Xu, Ri Zhang, Mengtian Zhou, and Junwei Mao, “A Facial Expression Emotion Recognition Based Human-robot Interaction System”, IEEE/CAA JOURNAL OF AUTOMATICA SINICA, VOL. 4, NO. 4, OCTOBER 2017.

[7] H. Lin and H. Wang, "Avatar creation in virtual worlds: Behaviors and motivations,"Computers in Human Behavior, vol. 34, pp. 213- 218, May 2014.