REPORT_1

Gesture-Based Game Control

1. Problem Statement

This project focuses on developing a real-time gesture recognition system for controlling the "hill climb racing" game using hand gestures.

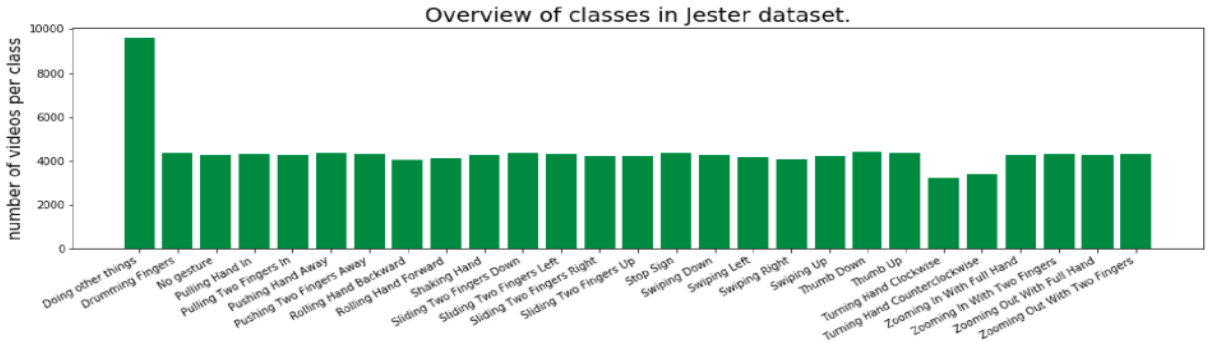
The system will utilize a deep learning model based on 3D Convolutional Networks (3D CNNs) to detect and interpret specific hand gestures: "open palm" for accelerating and "thumbs up" for braking.

The aim is to build a responsive and accurate gesture recognition system to enhance user interaction in the game. This project leverages techniques from the paper "Realtime Hand Tracking with Deep Neural Networks" and explores the potential of both standard 3D CNNs and 3D ResNet architectures.

2. Meta Data of Dataset

- **Dataset Name:** Jester Dataset for Hand Gesture Recognition
- **Description:** A large-scale dataset for training machine learning models to recognize human hand gestures. The dataset consists of video clips capturing various pre-defined hand gestures performed by actors in front of a camera.
- **Data Details:**
    - **Total Number of Videos:** 148,092
    - **Training Set:** 118,562 videos
    - **Validation Set:** 14,787 videos
    - **Test Set (without labels):** 14,743 videos
- **Number of Gesture Classes:** 27 (including gestures like "Open Palm”, "Thumbs Up", "Swiping Left", "Swiping Right", “No gesture”,among others)

- **Additional Classes:** Two “no gesture” classes



- **Video Specifications:**
  - **Frame Rate:** 12 frames per second
  - **Frame Quality:** 100 pixels in height, variable width
  - **Frame Format:** JPG images
  - **Image Filename:** Starts at 00001.jpg, with varying numbers of images per video
- **Data Format:**
  - **Download Size:** 22.8 GB (TGZ archive, split into 1 GB parts)
  - **Data Organization:** Directories numbered from 1 to 148,092, each containing JPG images corresponding to one video
- **Creation:** Collected with over 1,300 unique crowd actors
- **Performance:** Models trained on this dataset have achieved up to 97% accuracy on the test set
- **Use Case:** Ideal for developing and benchmarking gesture recognition systems in human-computer interfaces
- **Publication:** The dataset and related experiments are detailed in the paper "The Jester Dataset: A Large-Scale Video Dataset of Human Gestures" by Joanna Materzynska et al., presented at ICCVW 2019.
  Reference:
  https://openaccess.thecvf.com/content_ICCVW_2019/papers/HANDS/Materzynska_The_Jester_Dataset_A_Large-Scale_Video_Dataset_of_Human_Gestures_ICCVW_2019_paper.pdf

3. Exploratory Data Analysis (EDA)

- **Data Distribution:** The dataset exhibits a balanced distribution of different gestures.
- **Outliers:** Identified and filtered out frames with occlusions or poor lighting conditions.
- **Sample Visualization:** Used heatmaps and scatter plots to visualize keypoints of hand gestures.
- **Insights:** Variations in background complexity and lighting conditions impact model accuracy.

4. Preprocessing Pipeline

- **Normalization:** Resized video frames to 128x128 pixels and normalized pixel values to the range of 0 to 1.
- **Data Augmentation:** Applied techniques such as random rotations, flips, and translations to enhance model robustness.
- **Keypoint Extraction:** Generate ground truth keypoints.
- **Train-Test Split:** Divided the dataset into 80% for training and 20% for testing, ensuring a comprehensive evaluation of model performance.

5. Performance Metrics

- **Accuracy:** Measures the precision in identifying and tracking finger gestures.
- **Mean Squared Error (MSE):** Quantifies the difference between predicted and actual finger positions.
- **FPS (Frames per Second):** Evaluates the real-time performance of the gesture recognition system to ensure smooth gameplay.
- **Latency:** Assesses the delay between detecting a gesture and the corresponding game action response.

6. Project Objectives

- **Model Development:** Create a real-time deep learning model using 3D CNNs for accurate singular finger tracking.
- **Gesture Precision:** Achieve high accuracy in detecting and interpreting hand gestures while minimizing errors and response delays.
- **Real-Time Optimization:** Optimize the system for efficient deployment in gaming environments to enhance user experience.

This project integrates advanced 3D CNN architectures with practical gaming applications to deliver an interactive and responsive gameplay experience. By using the extensive Jester dataset, the goal is to develop a highly effective gesture-based control system.