e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703007 |

Human Activity Recognition

Dr K Shailaja[1], R Joshua Rahul[2], B Jahnavi[3], S Sai Yashwanth[4], G V Charan[5]

Assistant Professor, Department of Computer Science and Engineering, Anurag University, Telangana, India[1]

Student, Department of Computer Science and Engineering, Anurag University, Telangana, India[2]

Student, Department of Computer Science and Engineering, Anurag University, Telangana, India[3]

Student, Department of Computer Science and Engineering, Anurag University, Telangana, India[4]

Student, Department of Computer Science and Engineering, Anurag University, Telangana, India[5]

ABSTRACT: Machine learning research is heavily focused on human activity detection since it has various applications in a variety of fields, including security, entertainment, ambient supported living, and health management and monitoring. Researchers' interest in human daily activities is seen from studies on human activity recognition (HAR). As a result, the general architecture of the HAR system and a description of its key elements are described in this work.

Human activity recognition (HAR) is a field of study that aims to identify the actions performed by humans using sensors such as accelerometers, gyroscopes, and magnetometers. A project that can detect emotions, sit or stand position, and has an exercise rep counter would be a complex HAR system.

One possible approach to building such a system is to use a combination of machine learning algorithms such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Support Vector Machines (SVMs).

To detect emotions, the system could use facial recognition algorithms that analyze facial expressions and map them to different emotions. To detect sit or stand position, the system could use a data TensorFlow library to determine the orientation of the body. Finally, to count exercise reps, the system could use a combination of computer vision algorithms and sensors to track the movement of the body.

I. INTRODUCTION

Human Activity Recognition is one of the active research areas in computer vision for various contexts like security surveillance, healthcare, and human computer interaction. In this paper, a total of thirty-two recent research papers on sensing technologies used in HAR are reviewed. The review covers three areas of sensing technologies namely RGB cameras, depth sensors and wearable devices. It also discusses the pros and cons of the mentioned sensing technologies. The findings showed that RGB cameras have lower popularity when compared to depth sensors and wearable devices in HAR research.

Identifying human exercises requires foreseeing human exercises dependent on sensor information. Lately, it has become known for its numerous product applications that utilize current PC equipment. It records data on exercises like strolling, climbing steps, venturing down, sitting, standing, and resting.

With the increasing demand of security defense, anti-terrorism investigation and disaster rescue, human activity classification and recognition have become a hot research topic. This review is centered around acknowledging actual human exercises dependent on the understanding of camera sensor information which likewise incorporates one-dimensional time-series information.

The human face has peculiar and specific characteristics; therefore, it becomes difficult to understand and identify facial expressions. It is easy to identify the facial expression of a particular person in any image sequence. Presently, the research on facial expression is on the factors i.e, sad, happy, disgust, surprise, fear and angry. This paper aims to detect faces from any given image, extract facial features (eyes and lips) and classify them into 7 emotions (happy, fear, anger, disgust, neutral, sadness, neutral).

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703007 |

II. LITERATURE SURVEY

Human Activity Recognition:

• Ong Chin Ann Faculty of Engineering, Computing & Science Swinburne University of Technology Kuching, Malaysia
• Lau Bee Theng Faculty of Engineering, Computing & Science Swinburne University of Technology Kuching, Malaysia

Human activity recognition is an ability to interpret human body gesture or motion via sensors and determine human activity or action [6]. Most of the human daily tasks can be simplified or automated if they can be recognized via the HAR system [7], [8]. Typically, HAR systems can be either supervised or unsupervised [9]. A supervised HAR system requires some prior training with dedicated datasets while an unsupervised HAR system is being configured with a set of rules during development. HAR is considered as an important component in various scientific research contexts i.e. surveillance, healthcare, and human computer interaction (HCI).

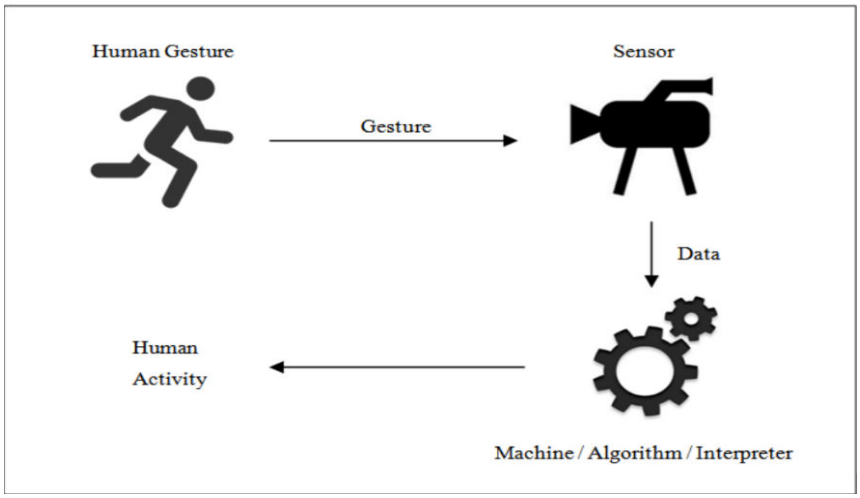

fig 2.1.1 General structure of HAR model

A 2-D Convolutional Neural Network refers to a three-dimensional operation. This means that the dataset will require to work on an image on three different axes and will have three different attributes to define it. A convolution focuses on the multiplication of sets of weights assigned to the input. The weights are referred to as a filter. Once the filter is applied many times, a feature map is obtained, whose value is passed through a 2D convolutional operation to receive an output.

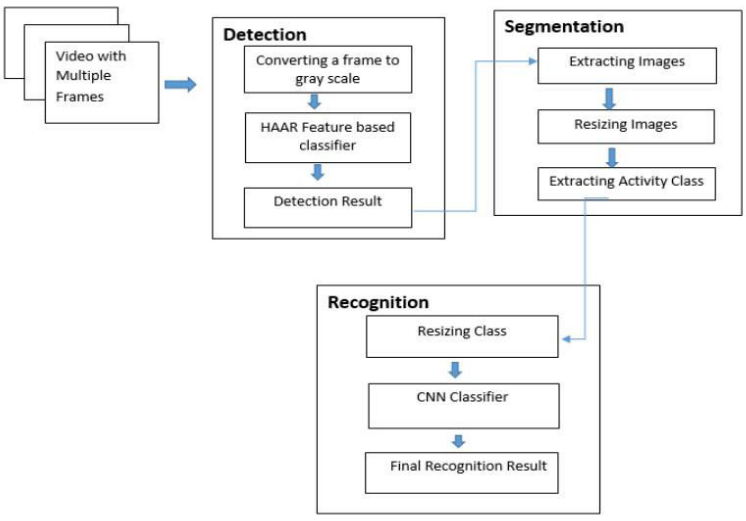

fig 2.1.2 Proposed Human Pose Detection and Recognition System Architecture

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703007 |

A. Human Pose Detection The detection intends to prepare all frames from an input video in real-time because continuous frames may hold human or not that is obliged to develop a proper human activity recognition system. Pose activities may not be recognized in delivered frames due to various reasons, for example, the human may be hidden or partially hidden by other objects or mostly shaded when they are out of focus. We trained our HAAR Feature-based classifier using only human images to detect the human poses from the input videos or images.
B. Human Pose Segmentation Human segmentation is essential for so many reasons such as extracting the two or many people in one single frame, removing the background, etc. The human segmentation is done by the result of the HAAR feature-based classifier where we get (X, Y, H, W), (X, Y) as starting coordinate and H height of the human and W is the width of the human. After deriving the human, we resize the image into 64 X 64 resolution.

III. METHODOLOGY

Develop a neural network model capable of accurately recognizing and classifying various human activities such as walking, exercising, and emotion detection which contributes to applications in fitness tracking, healthcare monitoring, and interactive user interfaces.

Recognize Activity Type: Successfully identify different types of exercises being performed by individuals, such as squats, push-ups, lunges, etc., based on the sensor data from wearable devices.

- Standing
- Sitting
- Idle
- Exercising

Mediapipe python library can be used to estimate human pose landmarks which helps in recognizing human activity.

3.1 Mediapipe

MediaPipe is an open-source framework for building multimodal machine learning pipelines. It is developed by Google AI and is designed to be cross-platform, customizable, and performant. It can be used to build a wide range of applications, including:

- Computer vision: Face detection, hand tracking, pose estimation, object tracking, scene segmentation, and more.
- Audio: Speech recognition, sound classification, music transcription, and more.
- AR/VR: Creating realistic virtual worlds and overlays.
- IoT: Developing smart devices and appliances with machine learning capabilities.

MediaPipe is based on a graph of modular components, which can be combined to create complex pipelines. Each component performs a specific task, such as decoding a video stream, running a machine learning model, or drawing a visualization. MediaPipe also includes several pre-built solutions for common tasks, such as face detection and hand tracking.
The MediaPipe Pose Landmarker task lets us detect landmarks of human bodies in an image or video. You can use this task to identify key body locations, analyze posture, and categorize movements. This task uses machine learning (ML) models that work with single images or video. The task outputs body pose landmarks in image coordinates and in 3-dimensional world coordinates.

Emotion Detection: Accurately predict the emotion displayed by the user.

- Happy
- Sad
- Angry
- Surprise
- Fear
- Disgust
- Neutral

To predict emotion, we can perform classification using the Convolutional Neural Network (CNN) model.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703007 |

3.2 CNN Classifier

A convolutional neural network (CNN) classifier is a type of machine learning model that is used to classify images and videos. CNNs are inspired by the structure of the human visual cortex, which is the part of the brain that is responsible for processing visual information. CNNs are made up of a series of layers, each of which performs a specific task. The first layer of a CNN is typically a convolutional layer, which extracts low-level features from the input image. These features can include things like edges, corners, and textures.The next layer of a CNN is typically a pooling layer, which reduces the size of the feature maps from the previous layer. This is done by merging neighbouring pixels together. Pooling layers help to reduce the number of parameters in the network and make it more robust to noise. The convolutional and pooling layers are typically stacked on top of each other multiple times, with each layer extracting higher-level features from the previous layer. The final layer of a CNN is typically a fully connected layer, which classifies the input image into one of several categories.

CNN classifiers are trained on large datasets of labelled images. During training, the network learns to associate the features that it extracts from the images with the corresponding labels. Once the network is trained, it can be used to classify new images by extracting the features from the images and then using the fully connected layer to predict the label.

CNN classifiers are used in a wide range of applications, including:

- Image classification: Classifying images into different categories, such as cats, dogs, and cars.
- Object detection: Detecting objects in images and videos, such as people, faces, and traffic signs.
- Scene segmentation: Segmenting images into different regions, such as the sky, the ground, and the objects in the foreground.
- Medical image analysis: Detecting and diagnosing diseases in medical images, such as X-rays and MRI scans.

We employ the image classifier Convolutional Neural Network (CNN) to make emotion predictions based on the extracted features. CNN is a powerful and efficient machine learning algorithm known for its speed and accuracy. It's particularly well-suited for handling large datasets and complex feature spaces.

3.3. Data Sources

In contrast, our approach capitalizes on the images scrapped from movies and tv shows available on various entertainment platforms like Netflix. We then classified the image scraps into different classes.

3.4. Data Volume and Availability

Our dataset is well classified and pre-processed to maintain the same resolution and quality throughout the dataset. We have also drastically reduced the image size which enabled us to train our model on a large test dataset containing 3000+ images.

3.5. Accuracy and Generalization

Our approach benefits from the structured nature of text data, which can be analysed with a higher degree of accuracy and consistency. It is better suited for recognition of various human activities.

3.6. Speed and Efficiency

Our approach is computationally efficient and faster as we have used Adam optimizer. It can provide near-instant personality predictions, enhancing user experience and accessibility.

Data Collection

Our approach capitalizes on the images scrapped from movies and tv shows available on various entertainment platforms like Netflix. We then classified the image scraps into different classes.

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703007 |

IV. RESULTS

Evaluation

In this project, we created a system that can recognize human activities, like walking or running, using a special type of computer program called a Convolutional Neural Network (CNN). We tried different methods, like using TensorFlow and Keras, as well as a tool called Mediapipe, to see which one worked best for making predictions and classifying these activities. Our findings are justified by a thorough evaluation of multiple machine-learning models, focusing on accuracy and F1-score as performance metrics. The consistently superior performance of the CNN model across these metrics indicates its reliability in personality prediction.

epoch	precision	recall	f1-score	support
1	0.03	0.03	0.03	111
2	0.13	0.11	0.12	1024
3	0.23	0.25	0.24	1774
4	0.17	0.17	0.17	1233
5	0.17	0.19	0.18	1247
6	0.10	0.10	0.10	831

4.2 Accuracies:

Accuracy	**83%**
Macro Average	86%
Weighted Average	83%

V. CONCLUSION

In conclusion, our project has successfully demonstrated the feasibility of predicting human activity types from a live camera source. The CNN model emerged as the most effective in terms of accuracy and F1-score. This suggests that our approach offers a robust solution for human activity recognition.

REFERENCES

1. Human activity recognition: A review by OC Ann· 2014
2. Analysis of Human Activity Recognition using Deep Learning by L Khattar· 2021

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703007 |

3. Human Activity Recognition Using Deep Learning by P Singh· 2022
4. Human activity recognition with smartphones by RM Gomathi· 2022
5. Human Activity Recognition with Videos by A Jain · 2021
6. Human Activity Recognition Using Smartphone Sensor by MA Khatun· 2020
7. Human Action Recognition From Various Data Modalities by Z Sun · 2022
8. Human Activity Recognition System from Different Poses by M Atikuzzaman · 2020

Impact Factor
7.521

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com