WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI Yoga Coach

Mrs. Madhu[1], Neeraj G K[2], Ravi R V[3], Vishwas J S[4], Vishal K[5]

[1]Guide & Assistant Professors, Dept of CSE Jyothy Institute Of Technology Bangalore India madhu.g@jyothyit.ac.in
[2]Dept of CSE Jyothy Institute Of Technology Bangalore India neerajgk56@gmail.com
[3]Dept of CSE Jyothy Institute Of Technology Bangalore India ravireddyv195@gmail.com
[4]Dept of CSE Jyothy Institute Of Technology Bangalore India vishwasjs2806@gmail.com
[5]Dept of CSE Jyothy Institute of Technology Bangalore India Vishalvishnu717@gmail.com

ABSTRACT—

The objective of this project is to develop a model known as HAR or Human Activity Detection that would take video input from different sources, like security cameras and recorded videos, and detect actions performed in them by humans.

The project begins with an examination of the models and practices already in use for carrying out this task, together with an evaluation of their advantages and disadvantages. This project will use a lot of techniques, such as sensors (tri-axial accelerometer and gyroscope), modeling techniques like CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks), DNNs (Deep Neural Networks), and a few image processing techniques. The objective is to design a model that can recognize a group of actions that it was trained on throughout the development stage. The created model aids in overcoming the drawback of earlier models that depended on sensors. By getting rid of the sensors, we also get rid of the expenses they entail and the mistakes they could introduce. Our group used logistic regression/logistic classification to find a collection of activities.

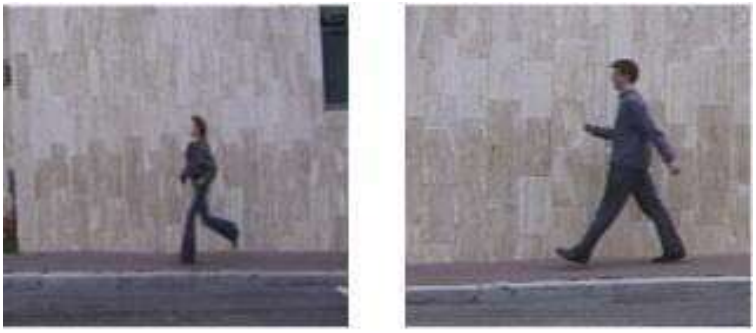
The result will be shown if the similarity to the action is greater than the threshold selected. It compares the action done in the videos given as input to the trained model used to detect these actions.

Keywords—Human Activity Recognition(HAR), Deep Learning, Machine Learning.

I. Introduction

The difficult task of identifying human actions in a video stream has recently attracted a lot of interest from the computer vision research community. Analyzing a human action involves more The difficult task of identifying human actions in a video stream has recently attracted a lot of interest from the computer vision research community. Analyzing a human action involves more than just demonstrating how various bodily parts move; it also involves describing the aim, emotion, and thoughts of the actor. As a result, it is now an important part of understanding and analyzing human behavior, which is important in many fields, such as surveillance, robotics, healthcare, video searching, and human-computer interaction. Video data differs from static images in that it incorporates temporal information, which is crucial for recognizing actions. Video data also incorporates naturally occurring data augmentation, such as jittering for video frame classification. Than just demonstrating how various bodily parts move; it also involves describing the aim, emotion, and thoughts of the actor. As a result, it is now an important part of understanding and analyzing human behavior, which is important in many fields, such as surveillance, robotics, healthcare, video searching, and human-computer interaction. Video data differs from static images in that it incorporates temporal information, which is crucial for recognizing actions. Video data also incorporates naturally occurring data augmentation, such as jittering for video frame classification.

(i) Gestures: Waving and Bend



(ii) Actions: Running and Walking

(iii) Interactions: Pickup phone call and Hugging

Fig. 1: Different classes of human activities

II. LITERATURE SURVEY

Understanding the gaps in the current or existing ecosystem resulted from accessing, analyzing, and exploring various papers. A number of researchers have invested their valuable time in the domains mentioned above in order to produce their beautiful work. Many of the new researchers and students have used these publications as a model and replicated the current methodology to create a more significant study. Our work also aims to imagine the ongoing excavation. We discovered the technique procedure used by the knowledgeable research outline after studying numerous publications. We carefully examined the researchers' manifesto as part of the literature review, and it is quoted verbatim below.

In [1] includes the following "Recent Advances in Video-Based Human Action Recognition using Deep Learning: A Review", In recent years, the study of computer vision and pattern recognition has seen a surge in interest in the area of video-based human action recognition. Many different fields, including surveillance, robotics, healthcare, video searching, and human-computer interaction, are among its many potential uses. Human action identification in videos faces several difficulties, including crowded backdrops, occlusions, viewpoint fluctuation, execution rate, and camera motion. Over the years, numerous strategies have been put up to deal with the difficulties. For research, three different dataset types—single perspective, multiple viewpoint, and RGB-depth videos—are used. This article reviews a number of cutting-edge deep learning-based methods that have been put forth for recognizing human action on the three different kinds of datasets. This review provides information on current trends and potential areas for future work to aid academics. This is because video-based human action identification is becoming more and more popular and has recently seen some development.

In [2] "Deep Learning Approaches for Human Activity Recognition in Video Surveillance", Recognition of human activity in films has drawn a lot of interest in a variety of computer vision applications, including intelligent surveillance, ambient assisted living, and human computer interaction. This study focuses on numerous methods based on deep learning, which is one of the most innovative techniques for Human Activity Recognition. The two main types of neural networks used in deep learning architectures are convolutional and recurrent. Deep learning is able to automatically pick up features from the input modalities. This survey paper presents analysis based on methodology, accuracy, classifier, and datasets. In [3] "Deep Learning Based Human Action Recognition: A Survey", Because of its many potential uses, human action recognition has received a lot of interest. The methods of human action recognition based on deep learning become widespread and evolve at a feverish speed with the rapid rise of computer performance and the Internet. Using color films, skeleton sequences, and depth maps, this study will present a fresh, logical taxonomy and a review of deep learning human action identification techniques. Additionally, a few datasets, useful techniques for action detection deep learning approaches, and the direction of development are all covered. In [4] "Human Activity Recognition System", A challenging task in computer vision is deciphering human activity from videos.

The main feature of intelligent video systems is the automatic identification of human actions in the video sequence and the labelling of such actions. The objective of activity recognition is to recognize an object's actions and intentions after a series of investigations into that object's behavior and its surroundings. The main uses of human activity recognition include robotics, human-computer interaction, fall detection, ambient intelligence, visual surveillance, video indexing, and content-based video analytics.

III. OBJECTIVES

The main objective of our work is the detection of Human Activities using video surveillance techniques. The objective of the project is to develop a fully automated system to positively identify the human activities present in the video which has been recorded actions and differentiate the types of actions. This project's main goal is to find solutions to human-centered issues ranging from security to healthcare by extrapolating from a few basic human behaviors. We may comprehend the procedure for teaching a model to recognize human behaviors by following the steps. However, the objectives can be categorized as below.

- Use the video stream as the input.
- Extract individual frames from the captured video after processing.
- Apply a trained human activity recognition model through the frame.
- Compare the frames and make an activity prediction.
- Categorize the action and add a caption to each shot.
- Play the video through the outputs.

IV. PROPOSED METHODOLOGY

The proposed methodology which we want to implement is the system that is being developed by us is basically based on the latest technologies such as Python 3, Keras, OpenCV, ResNet, and TensorFlow and much more. The main function of this gadget is to process the character frames of the input video in order to predict which hobby is being demonstrated while also systematizing the enter video flow for human activity detection. The frames are captioned after the prediction, which is more accurate and errorless, and the finished prediction is made to display in the developed Graphical User Interface as output. In the actual world, numerous cameras are employed to keep an eye on huge public areas like malls, airports, railroads, and bus stops. For the purpose of researching the issue of processing several views of the same person, some multi-view datasets have been generated. In contrast to single viewpoint streams, these datasets have the advantage of modelling a 3D human body shape from several viewpoints while avoiding occlusion issues.

a) DATASET COLLECTION

Many various kinds of datasets have lately been generated and distributed as a result of the advancement of human action recognition technologies. These datasets are frequently employed for experimental purposes in order to assess the efficacy and performance of new or existing procedures as well as to ensure fair comparison with other approaches. Deep learning may generally be used to many datasets with raw input data. The various dataset types may also have an impact on how complicated the networks are. For instance, single viewpoint data may only need one network to get the desired result, whereas multiple viewpoint data may need numerous networks to produce the same result. A depth camera might offer RGB and depth features using various technologies. We categories the datasets as single viewpoint, multiple viewpoints, depth camera, and RGB camera movies as a result. These datasets provide specialized features, such as joints, 3D body modelling, and gestures, for various research applications. The popular public datasets that deep learning techniques have been successfully used to are reviewed in this section.

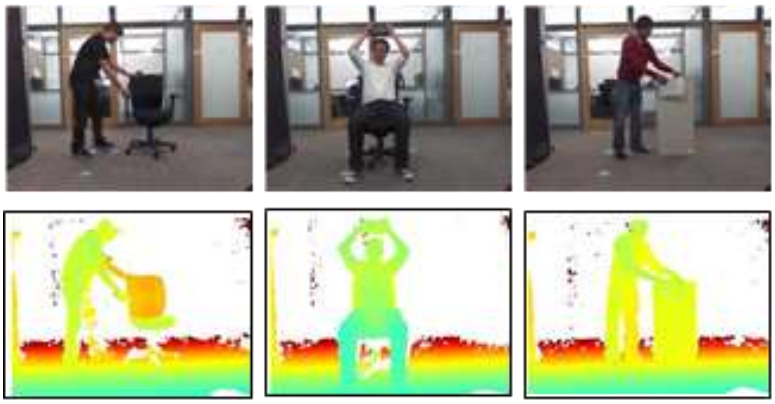


Fig 2- Samples of dataset collection

b) DATA PREPROCESSING

Data pre-processing aim is to improve the quality of the image so that we can analyze it in a better way. Involves certain processing techniques to ensure the video is clearly filtered before being analyzed. Removing noise in the collected videos requires the use of algorithms. Converting the video to processed video is carried out for quality assessment. Some of the process carried out includes image resizing removing unwanted portion and so on.

c) MODEL IMPLEMENTATION

CNN is an algorithm in Machine Learning that is used to deeply analyze an image. Apart from CNN we propose to use a random forest algorithm that works similarly as decision tree algorithm. Also, random forest algorithm is also used to classify and regression problems. To analyze the classical machine learning and modern deep learning technique. Random forest algorithm requires feature extraction however, CNN doesn't require.

The Human Activity detection modules are shown below. Video streams from a camera will be accepted as input by the system. The system can also extract every frame from the visual input. The system should be able to pre-process these input frames that were retrieved and resize them to the necessary threshold size. The system must be able to compare the frames to the weights it has learned. The system should be able to classify the input sequence into different categories after comparing with acceptable accuracy.

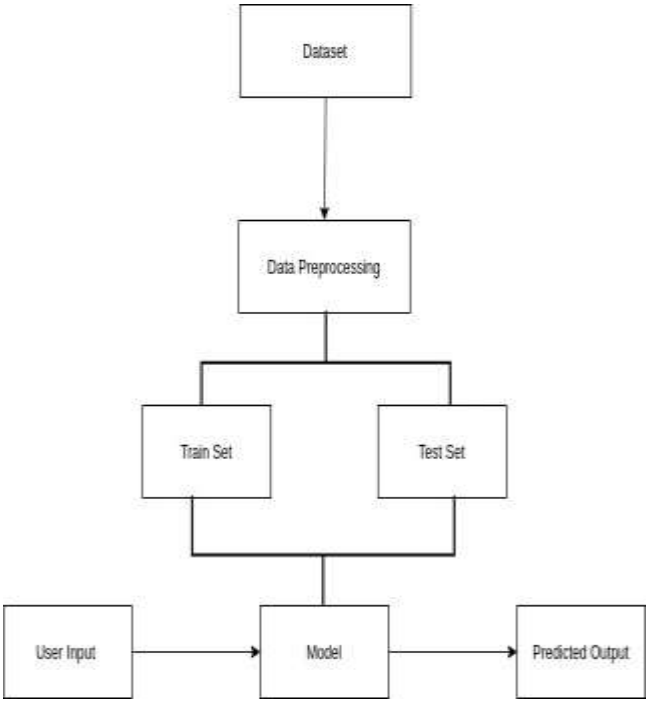


Fig 3- Flow diagram of the entire program

The following requisites include

- Programming Language: Python. Programming language of version 3.O or later
- Hardware: 8 GB RAM (At least)
- 250 GB of Hard Disk Space (At least), 1 GB of VRAM (At least)
- Software: Windows 7 or later, Ubuntu, Mac OS X or later
- Proposed technologies to use: System with Intel or AMD x86–64 processors with four logical cores and AVX2 instruction set support.

V. CONCLUSION

We are using a Human Activity Recognition System in this research. It is a program that recognizes and labels the motions or actions carried out by people. At first, we are conducting a literature review on how to put this application into practice. Next, we are going to examine the application's functional and nonfunctional needs. Then, to help with implementation knowledge, we are going to create a few UML diagrams. Following a requirement and design analysis, we provided a set of input data for the model to be trained using 80% of the data and tested with the remaining 20%. Once trained, the model can identify and label the activity present in each video provided as the input. The architectures of the convolution neural network (CNN) machine learning technique is used to train the model.