IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HUMAN ACTIVITY RECOGNITION USING OPENCV

[1]B Venkata Ramana, [2]D Lakshmi Prasanna, [3]M Tejasree, [4]C Yasaswi

[1]Professor, [2]B.Tech IV, [3]B.Tech IV, [4]B.Tech IV
[1,2,3,4]Information Technology,
[1,2,3,4]Vignan's Institute of Engineering for Women, Visakhapatnam,India

Abstract: Human Action Recognition is an imperative research area in the field of computer vision due to its numerous applications such as person surveillance, human to object interaction, etc. Human Action Recognition is based on a pre-trained CNN model for feature extraction. Convolutional neural networks (CNN) is a technique of deep learning. Most convolutional neural networks used for recognition task are built using convolution and pooling layers followed by a few number of fully connected layers and identifying similar patterns in an interval to recognize the action by providing accuracy of 79-90% based on the task.

***Index Terms* – Image capturing,Segmentation,Action Recognition,Captioning and speech output.**

I. INTRODUCTION

Human Action Recognition(HAR) has always been an important factor in social communication. Human activity and action recognition are all clues that facilitates the analysis of human behavior. HAR is always a major challenge for any fields of Applications. The Human Actions which are recognized in the videos are based on the analysis of a sequence of video frames by using computer to automatically find human actions without manual operations. Human Action Recognition is an area computer vision research and Applications. The goal of Human Action Recognition is to identify and understand the actions of people in videos and export corresponding tags which can be achieved through automated analysis or interpretation of ongoing events and their text using video data input.

II. MOTIVATION

Understanding the human activity and their interaction with the surrounding objects is a key element for the development of intelligent system. Human action recognition is a field that deals with the problems generates in the integration of sensing and reasoning, to provide context aware data that can be confer the personalized support across an application.

In the human action recognition system , still there are various issues which need to be addressed like as battery limitation of wearable sensors, privacy concern regarding continuous monitoring of activities, difficulty in performing HAR (Human activity recognition) in real time and lack of fully ambient systems able to reach users at any time.

III. KEY FUNCTIONS

Pre-trained: Human action recognition can be done using pre-trained model.
Feature Extraction: Similar patterns are identified based on the image frame captured.
Segmentation: Different classes related to action identified are segmented.
Captioning: The label of the action will be displayed are segmented.
Voice output: The speech output is given from the system based on the action identified.

IV. DESCRIPTION

The HAR data is made with a perspective of Recognizing Human Activities. The human activity recognition model was trained on Kinetic dataset which contains 400 actions. Though the retracing of Spatiotemporal 3D CNN the actions are recognized. The ResNet_34 kinetics model is used for determining the human actions through video classification of kinetic dataset. The activity recognition is done using automated analysis of ongoing events in the input file by providing captioning of the actions and activity label converted to speech.

V. METHODOLOGY

Methodology of Human Activity Recognition include several steps of processing from taking the input, identifying the similar patterns, comparing the frames with the Kinetic dataset, recognizing the actions and providing the context and speech of the action to the video frames.

5.1Capturing the frames

The human actions that are being performed in a video input are divided into frames at certain intervals of time. These frames are captured and taken as input to the CNN model to identify similar patterns by pooling them into certain classes of actions.

5.2 Dataset

A kinetics dataset which consists of 400 human activities is used for prediction and comparison of the input data. Kinetics dataset are taken from youtube recordings. The activities are human focuses and cover a wide scope of classes including human-object communications, for example mowing lawn, washing dishes, humans Actions e.g. Since the dataset is huge and downloading each clip would be a waste of time given that we already have pre-trained models by the original author. It will be easy and provides accurate results when worked on the pre-trained model than to train and tune it separately.

5.3 Recognition of the action

The kinetic dataset is used by the ResNet_34 3D model to compare similar patterns in the input data frames that are captured in the intervals. The similar patterns can be identified by CNN trough pooling layer by layer. The identified actions are categorized into classes of human activities. The recognition of the data input can be done by Resnet_34 model by video classification of 3D kernels. Segmentation of the actions are the classes which are identified by the model.

5.4 Context and Speech output of the action

Through the programming in python the captioning of the activity that is identified by the model can be displayed in the video while execution of the input file. Simultaneously, the speech of the activity that is captioned will be produced.

VI. RESULTS

4.1 Results of video input

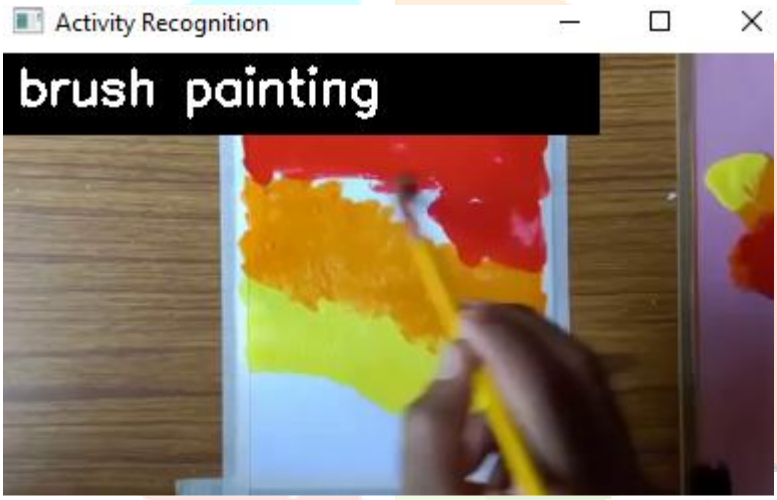

Fig.1

In Fig.1,The video input taken from webcam to detect the activity that is being performed. The result has been obtained through a video file window by providing the action as “brush painting” and also provides voice of the context or the activity that is captioned to the output file.

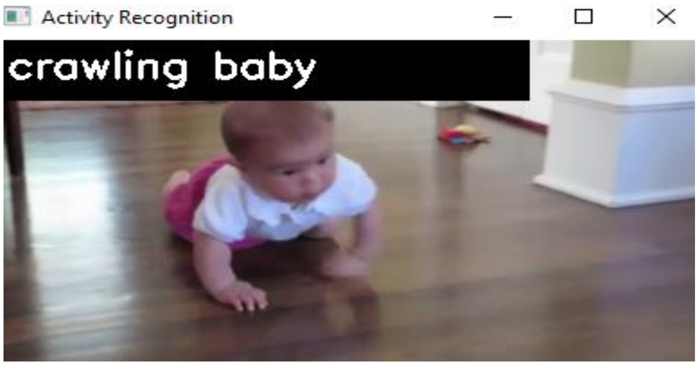

Fig.2

In Fig.2,the action “baby crawling” is identified by providing a video input of baby crawling for 10s. This action is accurate and give the exact output until the video frames of this similar action ends with providing speech output.

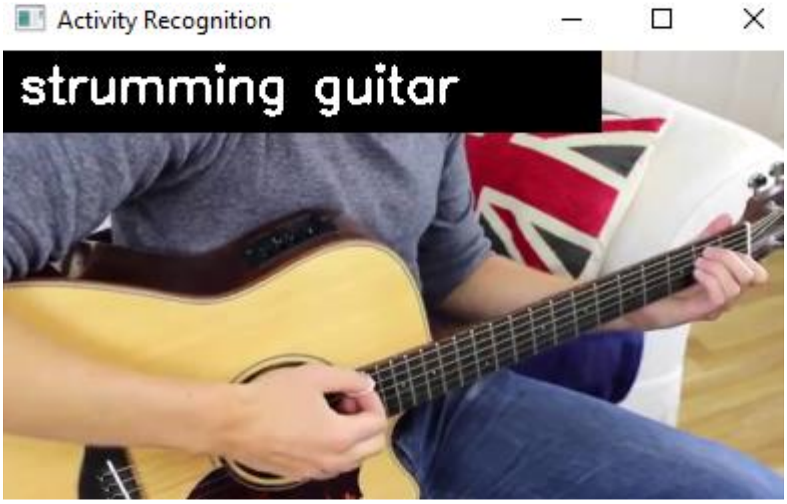

Fig.3

In Fig.3, human to object interaction is identified by the Resnet_34 3D CNN model. It identified the human interacting with the object guitar and provides the captioning as “Strumming guitar” using speech output from the computer. The model provides accurate result for all the video frames in continuous similar patterns.

VII. CONCLUSION

Human Activity Recognition System, we proposed a model trained using Convolutional neural network (CNN) with spatiotemporal three-dimensional kernels on Kinetic data set to recognize almost 400 human activities with satisfactory accuracy level. The designed system can be used to automatically categorizing a dataset of videos on disk, training and monitoring a new employee to correctly perform a task, verify food worker services, monitoring bar/restaurants patrons and ensuring they are well served. We used a dataset covering more than 400 activities in Resnet_34 3D CNN model to make the system more versatile. It is also observed that increasing the number of samples for an activity in the dataset improves the performance and provides more accurate results through speech.

VIII. FUTURE SCOPE

Activity recognition is the basis for the development of many potential applications in health, wellness, or sports

1. HAR can be used for health monitoring which can be achieved by analyzing the activity of a person from the information collected by different devices.
2. HAR is used to discover similar patterns which are the variables that determine which activity the human performs.
3. HAR can be used for robotic automation which makes it easier to train a robot to interact with human and the objects.

IX. REFERENCES

[1] Kay et al.’s 2017 paper, “The Kinetics Human Action Video Dataset”.
[2] Hara et al.’s 2018 CVPR paper, “Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?” by authors Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh
[3] Ainsworth, W.A. and Pell, B. (1989). Connectionist architectures for a text-to-speech system. “Proceedings of Eurospeech'89”, Paris, France, pp. 125–128.