Navajyoti, International Journal of Multi-Disciplinary Research
Volume 5, Issue 1, August 2020

Human Action Recognition

Rakshitha G L[1], Divya G M[2], Senthil Vadivu[3]

1Student, Department of MCA, Jyoti Nivas College, Bangalore, India
2 Student, Department of MCA, Jyoti Nivas College, Bangalore, India
3 Faculty, Department of MCA, Jyoti Nivas College, Bangalore, India

ABSTRACT

Human behavior recognition and analysis is one of the most important subjects in computer vision and has drawn growing interest in many real-life applications due to its increasing demand. The goal of action recognition is to automatically evaluate ongoing actions from an unknown video.This research area leads to different applications such as patient management, fall detection, tracking, and interface between human and device. In this paper the description is made from a series of observations on the behavior of the subject and its surrounding environmental conditions on the actions of one or multiple human subject.

We use a Bottom Up method in our proposed system that first identifies the parts or joints in the picture for one or more humans, then assembles the parts together and associates them with a specific individual.Our model consists of two submodels, one of which is the Deep Neural Network (DNN) which recognizes the behavior for each individual based on joints and the other is the Recurrent Neural Network (RNN) which classifies actions using dynamic sequential joint data.

Visual Geometry Group (VGG) is a pre-trained model that is used to interpret the photo or video content. This model is a very strong model and is useful as a classifier for images and as the basis for new models using image data. VGGNet is used in our model to create feature maps for the image input. Instead, the features are fed into trust maps and fields of part affinity.

Since this project is real-time action recognition based on open pose for training and testing, we used a bottom-up approach that first recognizes the parts of the human being that they are put together and associated with a particular person

Keywords: Human action Detection, Deep Neural Network, Machine Learning, Convolutional Neural Network and Recurrent Neural Network

INTRODUCTION

Recently, automated video recognition of human activity has attracted attention due to its increasing need in different real-life settings, such as security monitoring, health care, child care, etc. This type of technology can generally be used in a surveillance environment to alert the security or related authorities of the crime or dangerous activities. In fact, automated understanding of human actions will help facilitate the entire treatment or rehabilitation process in childcare, the home of old age. Video identification of human activity usually starts with input video and pre-processing such as frame extraction. Segmentation is the method of splitting the image into disjoint bits, so that the segmented sections together form the original image. After the segmentation phase is motion tracking in which the basic idea is to identify the moving objects in the picture. Finally, a specific activity is estimated, and algorithms for activity recognition are used to identify and investigate the actions.

In this paper we present an efficient method for multiperson pose estimation on multiple public benchmarks with state-of - the-art precision.

Conventional skeleton-based action recognition methods focus primarily on the design of hand-crafted characteristics to represent the skeleton.Recurrent neural networks (RNNs), convolutionary neural networks (CNNs), where joint coordinates are represented as vector sequences, pseudo-images and graphs, respectively, are the most commonly used models in deep-learning methods

Since this project is real-time action recognition based on an open pose for training and testing, we used a bottom-up approach that first recognizes the parts of the human being that they are put together and associated with a particular person

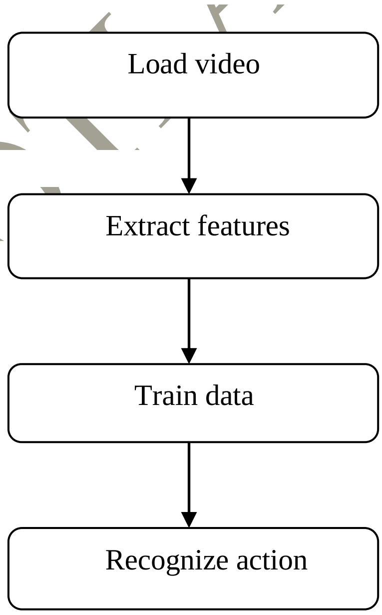
PROPOSED WORK

We want to build a system in this project that can identify human actions using data from the skeletons. We use Bottom Up approaches which involve first detecting the parts or joints in the image for one or more humans, then assembling the parts together and associating them with a specific human being. Because our project is based on Open Pose network, it first extracts features from an image and then fed into maps of trust and fields of part affinity.

Work Flow

The model takes as input a color image of size w × h and produces, as output, the 2D locations of keypoints for each person in the image

Flow diagram



IMPLEMENTATION AND RESULTS

- **Extraction of features**

Firstly, Open Pose network extracts features from an image. The features are then fed into two parallel layered convolution branches

Confidence maps: each map representing a particular part of the human pose skeleton.

Part Affinity Fields: represents the association between parts

The detection takes place in three stages :
Stage 0: The VGGNet's first 10 layers are used to construct feature maps for the input file.

Stage 1: A2-branch multi-stage CNN is used where a series of 2D confidence maps (S) of body part positions (e.g. elbow, knee etc.) are predicted by the first branch. Trust maps and Affinity maps for the keypoint-Left Shoulder are given below.

Fig 1: Showing confidence maps for Left Shoulder for the given image

The second branch predicts a set of 2D vector fields (L) of part affinities, encoding the degree of part association.

Fig 2: Showing Part Affinity maps for Neck – Left Shoulder pair for the given image

Stage 2: The maps of trust and affinity are parsed with greedy inference to produce the 2D keypoints for all people in the picture.

Training data

Training is performed on the COCO dataset. The COCO model yields 18 points. In the image below you will see the outputs plotted on a human.

Fig 3: COCO model produces 18 points

Skeleton construction

Because we know the points indices in advance, we can draw the skeleton when we have the key points by simply joining the pairs and forming the bounding boxes.

Fig 4: skeleton is created based on the joints

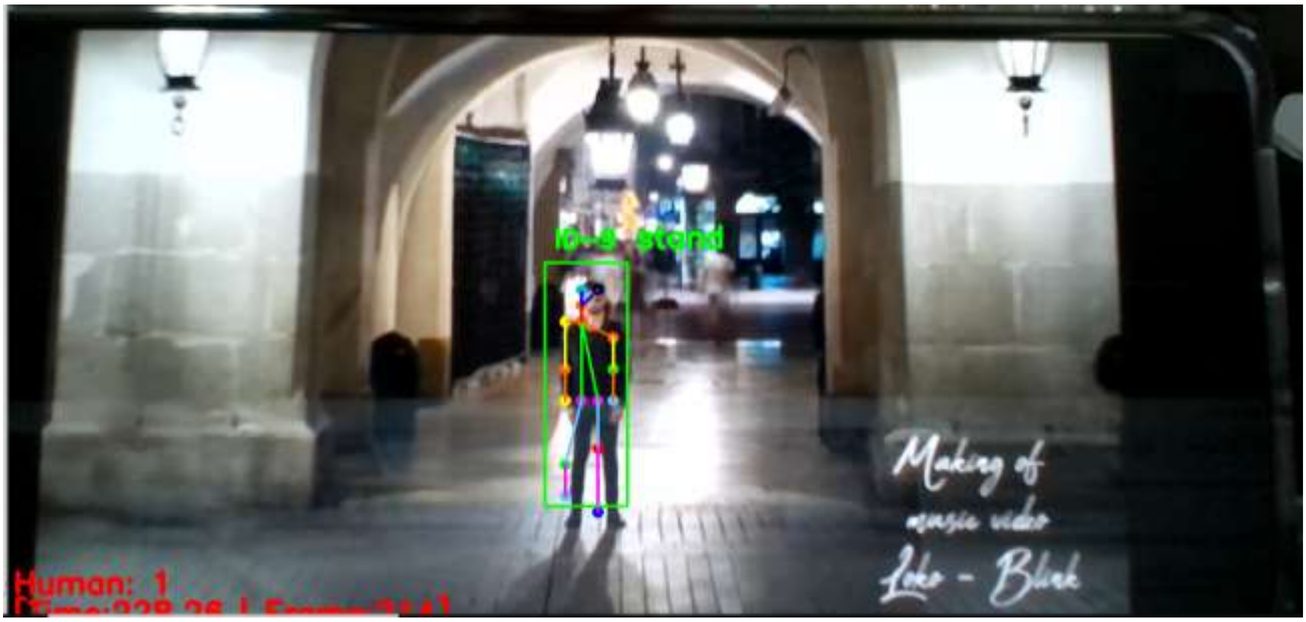
Finally action is recognized based on the skeleton and the joints

Results

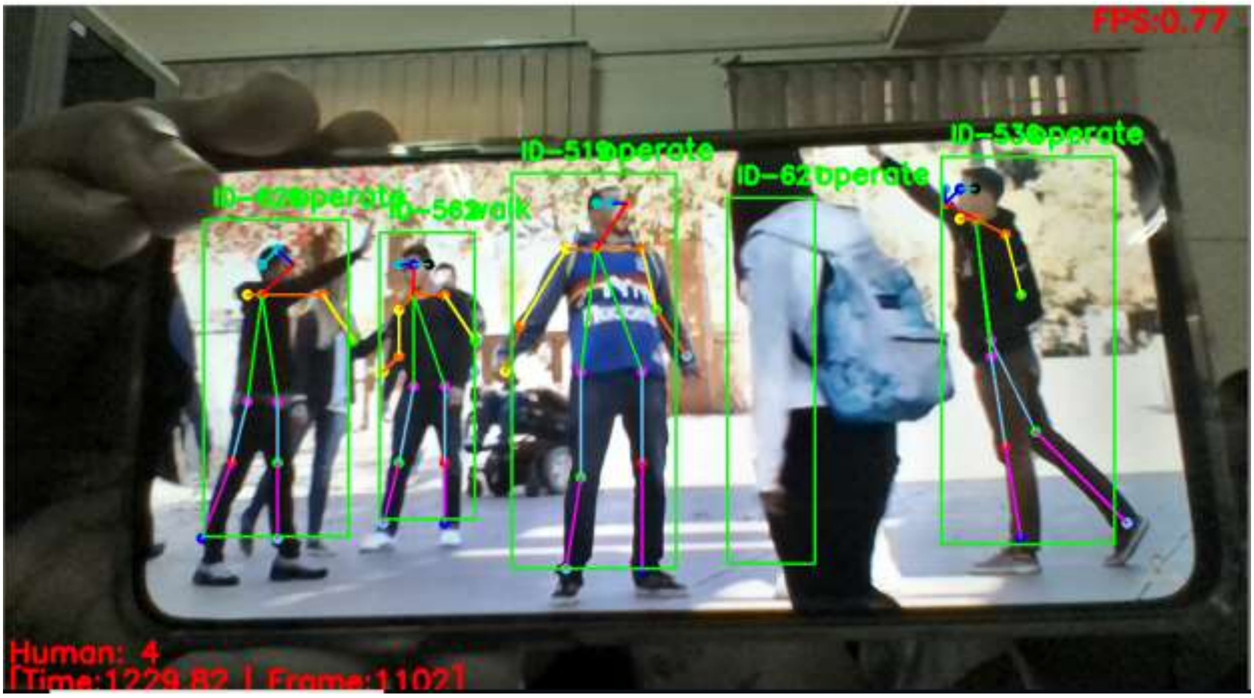
Actions recognized by our system are walk and stand

Walk recognition

Stand recognition



Detecting a human by providing it with a ID



Provides a skeleton

A warning is produced when full body is not being detected inside the bounding box

CONCLUSION

The perception of human activity remains an important topic in computer vision. HAR is the basis for many applications, including video surveillance, health care, and contact between people and computers. Methodologies and technologies have evolved tremendously in recent decades and have continued to evolve up to date. However, challenges still exist when faced with realistic sceneries, in addition to the inherent variation in intraclass and problem of similarity between classes.

REFERENCES

1. Zhe Cao, et al, OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, 30 May 2019
2. Mihai Nan and Adina Magda Florea,et al, Spatio-Temporal Features in Action Recognition Using 3D Skeletal Joints, 21 January 2019
3. Thao Minh Le, et al , A Fine-to-Coarse Convolutional Neural Network for 3D Human Action Recognition,18 August 2018
4. Zhengyuan Yang1, et al, Action Recognition with Visual Attention on Skeleton Images
5. Mengyuan Liu, et al, 3d Action Recognition Using Multi· Temporal Skeleton Visualization, 2017
6. Cho Nilar Phyo, et al , Skeleton Motion History based Human Action Recognition Using Deep Learning, 2017
7. Lei Shi, et al , Skeleton-Based Action Recognition with Directed Graph Neural Networks
8. Deepika Singh , et al , Human Activity Recognition Using Recurrent Neural Networks
9. Hong Liu, et al , Skeleton-Based Human Action Recognition Using Spatial Tempor 3d Convolutional Neural Networks ,2018
10. Raviteja Vemulapalli , et al, Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group ,2014
11. Komang Mandira G A ,et al, Human activity recognition using skeleton data and support vector machine, 2019
12. Suolan Liu, et al , Human Activities Recognition Based on Skeleton Information via Sparse Representation , 08 January 2018