JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A REVIEW PAPER ON DEVICE AUTOMATION THROUGH VIDEO PROCESSING

[1]Shruthishree S.H, [2]Anuj kumar, [3]Nilesh dubey, [4]Mohit ranjan , [5]Somvrat Shukla

*1 Assistant Professor, Department of Information Science Engineering, FET, Jain University, Kanakapura(T), Ramanagara(D) , Bangalore, India.

*2,*3,*4,*5, B-Tech, Dept. of ISE Jain University: Faculty of Engineering and Technology, Jakkasandra Post, Kanakapura Taluk, Bangalore.

Abstract: In the recent times, machine learning is considered to be the best replacement of manual work. With every passing day, the concepts such as Artificial Intelligence, Video Processing and Image Processing are taking over the maximum attention. PoseNet, being a part of the Machine Learning sphere helped us with Real-Time Human Pose Estimation for one or multiple persons depending on the kind of algorithm we are working with. Referring to the real-time computer vision, we have implemented the programming ideologies mainly from OpenCV or Open Source Computer Vision libraries. The use of ML5 made the concepts of Machine Learning more approachable with the algorithms and models present in the browser. ML5 is a version of less powerful TensorFlow library which can be used in web-based framework like JavaScript. This project is an attempt to familiarize the concept of controlling the electrical appliances by Video Processing and Gesture Control. Performed the software stimulants and coding using the JavaScript and for the final implementation, the PoseNet Technology is also merged with the later.

Keywords: **Automation; Video processing; Pose estimation; PoseNet; ML5; Heatmap; Key points; Tensorflow; Machine learning Algorithm; Classification Algorithm; Gesture recognition; Single pose estimation; Multi pose estimation; Internet of Things(IOT); Firebase Database**

I. INTRODUCTION

This project is solely based on the concept of Video Processing and Gesture Control. So, to speak specifically about the technology used for the aimed purpose is PoseNet. The estimated outputs from the implementation of this technology into work is based on the Human Pose Detection. This is not limited to a single individual but at the same time for multiple people depending upon the algorithm, we are directing it to do. The estimation is based on the vision model which is developed on the basis of the key body joints of the user. This technology is merged with Video Processing using the JavaScript to control the functioning of electrical appliances in a room. Automation means to automate certain function or operation of devices in a home into a centralized home automation system that aims to improve energy management, security and to provide a system its own intelligence so that it can make its own decision. It helps reduction of re do of a particular task. PoseNet, being a part of the Machine Learning sphere helped us with Real-Time Human Pose Estimation for one or multiple persons depending on the kind of algorithm we are working with. Referring to the real-time computer vision, we have implemented the programming ideologies mainly from OpenCV or Open Source Computer Vision libraries.

This project is solely based on the concept of Video Processing and Gesture Control. So, to speak specifically about the technology used for the aimed purpose is PoseNet. The estimated outputs from the implementation of this technology into work is based on the Human Pose Detection. The estimation is based on the vision model which is developed on the basis of the key body joints of the user. This technology is merged with Video Processing using the JavaScript to control the functioning of electrical appliances in a room.

II. REVIEW OF LITERATURE

Existing system:

There are some similar existing systems like Remote controlled appliances such as TV, CD player, Air Conditioner, DVD Player and Music System. Remotes can be also used for control of light, Door Opener etc. A single Universal Remote can control all the devices.We have encountered the use of software connected to databases to create automated systems also known as IoT. We have encountered automation of systems using speech recognition and Natural Language Processing as in Alexa, Siri, Cortana, etc. Also there are noise sensing devices which work on frequency and wavelength of sound produce and if it matches the one stored in device storage any particular operation then that operation is triggered. Also, we encountered the integration of both IoT(Internet of Things) and NLP(Natural Language Processing) to provide an interactive home system where every electrical appliance is controlled wirelessly and voice-controlled. This project shows us the possibility to do the same via video inputs and gesture recognition.

III. MATERIALS & METHODS

Implementation can be divided into 3 parts. These are:
• Data Collection • Data Training • Output

DATA COLLECTION: As the name suggests, data collection deals with the collection of data from the user. The user provides the machine learning model with an input and the gesture to be stored under that input variable. A delay is set before collecting data so that the user can be ready with his pose. The data is then collected. Again, a time delay is specified for collection of data from user. During this time co-ordinates of all points are stored in a file. This file can be downloaded as a .json file.

DATA TRAINING: Data training essentially deals with the training of data provided by the user and building of models from the trained data to meet the needs of the user. The trained model can be used in various tasks. The applications of this project are limited only to our imaginations. OUTPUT: The output part deals with the output concerning the trained model. It predicts the output according to the user defined inputs.

Algorithms/Technology Used:

The Algorithm used here are POSENET which uses ML5 library and TensorFlow. JavaScript is used to write effective code in fewer lines. Also, JavaScript video feed is used to take video input from browser. P5 editor is used storing files and processing code. Database and IOT device like raspberry pi is used to connect the device and database.

Working Mechanism of POSENET:

To begin with, we can say that the basic working structure of PoseNet lies in the fact that the process first detects the user. This is followed by the estimation of parts and then calculating the pose. This kind of process is also known as a top-down approach. In the case of multiple individuals, this process remains unchanged only that the calculations take time in accordance with the number of individuals. we can say PoseNet impliesto a vision model based on the posture of a person through an image or video which will be totally based on the position of body key joints. Talking about the key joints we must know that there are 17 of them. They are as follows:

0- Nose
1- LeftEye
2- RightEye
3- LeftEar
4- RightEar
5- LeftShoulder
6- RightShoulder
7- LeftElbow
8- RightElbow
9- LeftWrist
10- RightWrist
11- LeftHip
12- RightHip
13- LeftKnee
14- RightKnee
15- LeftAnkle
16- RightAnkle

There are two phases of real-time pose estimation: 1. An RGB image (or Red Scale Green Scale Blue Scale) is the input from the user which can be made through a convolutional neural network. 2. No matter whether the decoding algorithm is for pose

detection of an individual or multiple people, the code will decode pose, pose confidence scores, key point positions and key point confidence scores. These are output obtained from the model. Pose at the peak level refers to the object returned by the PoseNet which consists of the key points and also the confidence-level score at that particular instance for each and every person. Confidence Score will give us an idea about the overall confidence which is detected during the time of estimation of pose. It usually ranges between 0.0 to 1.0. The key point position refers to the x and y coordinates in accordance with the original input image. The confidence score at the key point can also be measured using the key point confidence score. It too ranges between 0.1 and 1.0.

Figure:1 Key joints estimation

The TensorFlow can implements the PoseNet in various ways.

1. The detection of the key body joints of the image is detected with the help of heatmaps and offset vectors.
2. The image when scaled through TensorFlow consists of a number of layers and model outputs.
3. The accuracy of the image in terms of the distance between the source and the receiver is also taken into consideration by TensorFlow.

IV. RESULT & DISCUSION

The methods implemented in this project include Video capture in a brower and frame by frame analysis of the video using heatmaps to identify 17 key points in a human body as mentioned earlier. The co-ordinates of each point is stored in a json file, namely the x-axis and y-axis of each point in each frame is stored in the form of a comma separated value(csv) file which is downloaded in the system in the form of json file. This is the first part of the program where the data is collected from the user, stored in a json file and downloaded in the system. The second part includes feeding this json file in a training algorithm where the model, model_meta and model.weights.bin is created , saved and downloaded. Here we also specify the epoch value so as to form a graph which shows how the losses are exponentially reduced during the training of data. The third part includes the output part where the input fed by the user is stored as a variable to the gesture values stored and when the video feed matches the model training set, a specific output is displayed. This displayed output can be stored in a database for further use in IoT related projects.

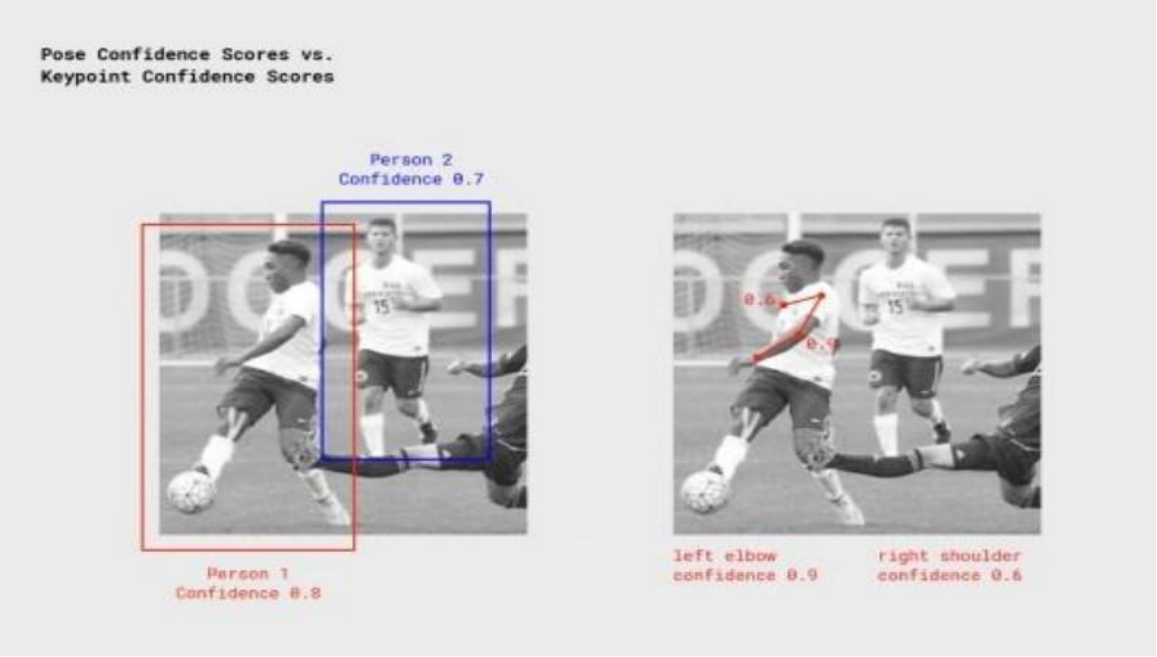

Figure:2 Pose confidence score vs Key point scores

The above-mentioned processes such as the Google-Coral or the TensorFlow makes the process or rather the implementation of the PoseNet Technology a bit complicated to carry out. This is mainly because of the complexity in the processing of the image. Therefore, comparatively the JavaScript is easier to understand to execute as well. The JavaScript allows an easier approach to the demonstration of a model using PoseNet for carrying out the tasks assigned through the algorithms.

V. CONCLUSION

In the end, we can say that we have made a system that can eliminate the anomalies of another system. Automation using video processing uses JavaScript language for code and is run on the p5 editor platform. It uses the Ml5 library to run a browser-based pose estimation machine learning algorithm. For now, it will perform pose estimation on a single body but can be configured to perform the same on multiple entities. To make it more reliable and efficient, we can add a layer of security by implementing a face recognition algorithm with the pose estimation model. We use a relatively new technology known as PoseNet which is a vision model used to estimate the pose of a person in an image or a video by estimating the key body joints. In all, a total of 17 body points are identified, and x and y coordinates of each point are noted.

VII. REFERENCES

[1] Paper on Camera based Smart Surveillance System-Literature Survey published in [IRJET] by Ishan Kokadwar, Anurag Kulkarni, Sayali Khare, Vaibhav Limbhore, Swati Chandurkar e-ISSN: 2395-0056|p-ISSN: 2395-0072, Volume: 07 Issue: 06 | June 2020.

[2] Journal Control of Home Devices based on Hand Gestures published in researchgate.net by Pomboza-Junez Gonzalo and Holgado-Terriza Juan A, Chapter · September 2015 DOI: 10.1109/ICCE-Berlin.2015.7391325.

[3] Literature survey on Home Appliances Control Based on Hand Motion Gesture published in [IJETER] by D. Chaitanya, Y. Aruna Suhasini and G.Nagarjun.

[4] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, Mo Nours Arab, (2008) "Human- Computer Interaction: Overview on State of the Art", International Journal on Smart Sensing and Intelligent Systems, Vol. 1(1).

[5]Mrs. Prajakta Jadhav1, Mrs. Shweta Suryawanshi2, Mr. Devendra Jadhav3 "Automated Video Surveillance eISSN: 2395 - 0056 | p-ISSN: 2395-0072" International Research Journal of Engineering and Technology (IRJET).