Human Body Recognition

Ankita Mishra, Kapil Verma, Baljender Kaur, Anurag Shrivastava

Department of Computer Science and Engineering,

Babu Banarasi Das Northern Institute of Technology, Lucknow

ABSTRACT

The human activity monitoring system helps to differentiate a person's physical actions such as walking, clapping, shaking hands etc. Activity awareness is the foundation for the development of many potential applications for health, wellness and sports. HAR has a variety of uses because of its impact on health. Helps users improve quality of life in areas such as aged care, daily logging, personal fitness software. Personal Performance Recognition is a field for identifying basic human activity and is currently being used in various fields where important information about an individual's ability to work and lifestyle.

As the famous saying goes "Exercise not only changes our body it changes our mind, our mood, and our attitude". Fitness is a practice today. Everyone wants to be fit, to be beautiful, and to be healthy. But during this epidemic, not everyone can hire a coach or go to the gym. Another option is wearable devices that not everyone can afford. This paper proposed an AI trainer model. The proposed model used by anyone regardless of age and health status. The AI model uses Personal Status Evaluation. It is a popular method and determines the location and posture of the human body. This technique creates important points in the human body and is based on the fact that it creates a virtual skeletal structure in the 2D dimension. Featured is a live video taken from a person's webcam and the output captures location marks or key points in the human body. The AI trainer specifies the calculation and timing of the settings that a person must perform. It also specifies errors and feedback if any. This paper provides a way to use the stop rate that works on the CPU to get the correct points. Based on points touch and other curls (biceps) are calculated. This paper proposes a method that uses OpenCV to use a stand-alone model.

Keywords: ***Pose Estimation, Real Time Body Detection, movement Classification, BlazePose Model, Activity Recognition.***

I. INTRODUCTION

Human Pose Estimation is one of the challenging yet broadly researched areas. Pose estimation is required in applications that include human activity detection, fall detection, motion capture in AR/VR, etc. Nevertheless, images and videos are required for every application that captures images using a standard RGB camera, without any external devices.[1] This paper presents a real-time method of sign language detection and detection using the MediaPipe Holistic pose measurement method. This Complete Framework identifies a wide range of movements — facial expressions, hand gestures, and body language, all of which are excellent examples of sign language awareness. The tests performed included five different signers, who signed ten different names behind nature. The two symptoms, “empty” and “sad,” were best seen by the model.

MediaPipe Pose is a ML solution for high-fidelity body pose tracking, inferring 33 3D landmarks and background segmentation mask on the whole body from RGB video frames utilizing our BlazePose research that also powers the ML Kit Pose Detection API [2]. Current state-of-the-art approaches rely primarily on powerful desktop environments for inference, whereas our method achieves real-time performance on most modern mobile phones, desktops/laptops, in python and even on the web.

With real-time performance of a full ML pipeline that combines positioning and tracking models, each segment should be very fast, using only a few milliseconds per frame. To achieve this, we see that the strongest signal in the neural network is about the location of the torso of the human face (due to its very different features and slight variation in appearance). Therefore, we achieve a fast and lightweight detector with a strong performance (currently for most mobile and web applications allowed) that the head should be visible in our one-man mode. The solution uses a two-step ML pipe detector, which has been proven to work effectively in our MediaPipe Hands and MediaPipe Face Mesh solutions. By using a detector, the pipeline first places a person / area of a region-for-profit (ROI) within the framework. The tracker later predicts the location indicators and the differential mask between ROI using a cut-off ROI frame as input. Note that in video use cases the detector is requested only as required, i.e., in the first frame and where the tracker can no longer detect physical presence in the previous frame. In some frames the pipe simply receives ROI from the local landmarks of the previous frame.

II. LITERATURE REVIEW

Human Pose Estimation from Single Image (Naimat Ullah Khan School of Communication and Information Engineering. Institute of Smart City. Shanghai University, Shanghai, China)[3]. HPE is designed as a DNN-based retrieval component targeted at body parts and is defined as a complication of local organ transplants that are effectively managed by Phase Based Models i.e. PS. These Sector-Based models have limited language as they use local finders that can only show certain variations of interactions between body parts and fail to predict the position in which the joints are more defined or less visible. Therefore, in order to overcome these limitations, comprehensive approaches to HPE were proposed. This section covers three different DPE-based HPE methods; a complete, partial approach and a combination of both methods. Toshev proposed the popular DeepPose, a complete HPE method that uses DNN.

Liangchen Song;Gang Yu;Junsong Yuan;Zicheng Liu; (2021). Human pose estimation and its application to action recognition: A survey . Journal of Visual Communication and Image Representation[4]. Representative methods of video-based methods can be divided into three categories: 3D network-based methods, short-term memory-based methods (LSTM) and dual-based streaming methods. The 3D conversion network integrates frames into video and uses the 3D conversion method to learn in 3D space time structure. 3D convolutions are mathematically expensive, so some functions are exploring how to expand 2D to accommodate 3D architecture. With LSTM-based methods, frames from video are considered as input sequences. In contrast, a convolutional dual-channel network usually uses visual flow information, calculated electronically from a standalone sequence. Also, there are two-channel streaming that does not require visual flow and uses alternatives instead, such as SlowFast Network.

Deep Learning-Based Human Pose Estimation: A Survey CE ZHENG∗ , University of Central Florida, USA WENHAN WU∗ , University of North Carolina at Charlotte, USA[5] In this survey, we studied a comprehensive overview of the latest methods based on in-depth learning of 2D and 3D HPE. Complete taxonomy and performance comparisons of these methods have been included. Despite the great success, there are still many challenges as discussed in Sections 3.3 and 4.3. We also show you a few promising future directions to improve progress in HPE research. • Familiarity with the HPE domain. For some applications, such as measuring human posture in baby pictures or artwork collections, there is not enough training data with annotations of basic truth. In addition, the data for these applications shows a different distribution from the standard status data. HPE methods trained in the existing standard database may not integrate well into different domains. The latest practice is to bridge the domain gap using GAN-based learning methods. However, the mechanism for effectively transmitting personal position information to close domain spaces has not been improved.

DeepPose: Human Pose Estimation via Deep Neural Networks Alexander Toshev toshev@google.com Google Christian Szegedy szegedy@google.com Google[6] We studied to identify the best of our knowledge, the first use of Deep Neural Networks (DNNs) in the measurement of human status. Our problem-solving constructs such as DNN-based retrieval from shared links and the presented platform for such retreats are useful for capturing context and thinking about the situation in a holistic manner. As a result, we are able to achieve better or better results in a number of challenging educational databases.

AI-Based Yoga Pose Estimation for Android Application Girija Gireesh Chiddarwar, Abhishek Ranjane, Mugdha Chindhe, Rachana Deodhar, Palash Gangamwar Computer Department, SCOE, Vadgaon, Pune-411041, India [7]. Deep Learning Methods proved to be very useful in spatial measurement, compared to any other methods. The various uses of standing measurements have stimulated many advances in the field, both in terms of speed and accuracy. We concluded that PoseNet, by current standards, is the best way to use mobile applications, especially yoga. This method can be easily performed using the TensorFlow-Lite framework. A pre-trained model can be considered to score 17 key points. We use these points to calculate angles and compare them with the correct angle angles calculated by OpenCV. Invalid skeleton and angles are displayed. Therefore, this paper has provided an overview of how posture techniques have evolved over the years, and how they can be used effectively in many applications, such as yoga practice.

III. METHODOLOGY

In this section, we introduce the key components of what we call the Physical Awareness Program. An important preparatory step, which influences all the following design decisions of the automatic pipe is the determination of the appropriate input model (human body / sign) and target (emotion / expression). A live view of one-on-one touch, local gestures, and real-time hand tracking on mobile devices can enable a variety of modern life applications: fitness and game analysis, control of posture and sign language recognition, virtual reality and effects. MediaPipe already offers fast, quick and accurate solutions, but different, for these complex functions. Integrating them all into a consistent real-time end-to-end solution is a complex problem in a different way that requires simultaneous consideration of many interdependent emotional networks. The MediaPipe Holistic Pipeline integrates different body models namely structure, facial features and hand, each developed for its specific background. The stand-alone model, for example, takes low resolution and video frame with adjusted resolution (256x256) as input resolutions. But if one could cut out the regions / parts of the hand and face to move to their proper models, the image correction would be too low to say the right thing. Therefore, we have designed MediaPipe Holistic as a multi-phase pipeline, more precise than any other, which covers a wide range of regions using appropriate regional image editing.

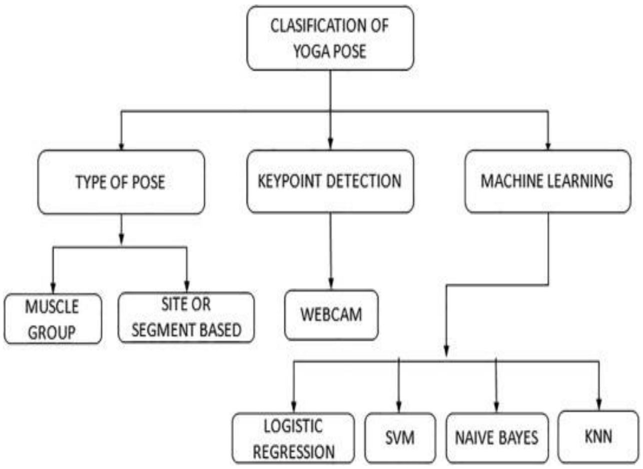
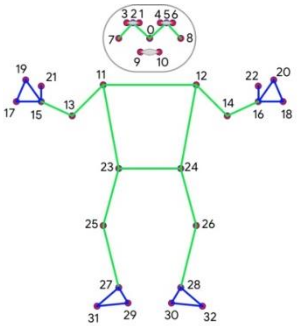


Fig. 1 Classification of yoga poses.

A. Pose Estimation Quality



0.	nose
1.	right eye inner
2.	right eye
3.	right eye outer
4.	left eye inner
5.	left eye
6.	left eye outer
7.	right ear
8.	left ear
9.	mouth right
10.	mouth left
11.	right shoulder
12.	left shoulder
13.	right elbow
14.	left elbow
15.	right wrist
16.	left wrist
17.	right pinky knuckle #1
18.	left pinky knuckle #1
19.	right index knuclke #1
20.	left index knuckle #1
21.	right thumb knuckle #2
22.	left thumb knuckle #2
23.	right hip
24.	left hip
25.	right knee
26.	left knee
27.	right ankle
28.	left ankle
29.	right heel
30.	left heel
31.	right foot index
32.	left foot index

Fig. 2 list of all 32 joints in BlazePose model

In order to evaluate the quality of our models in comparison with other effective publicly available solutions, we use three different verification data sets, representing different specific positions: Yoga, Dance and HIIT. Each photo contains only one person located 2-4 meters from the camera. In order to keep up with other solutions, we tested only 17 key points from COCO topology. However, COCO key points are only found in the area of the ankle and wrist, with no scales and knowledge of the shape of the hands and feet, which are essential for practical performance such as strength and dance. The inclusion of additional key points is essential for the subsequent use of background-based modeling models, such as those of the hands, face, or feet.

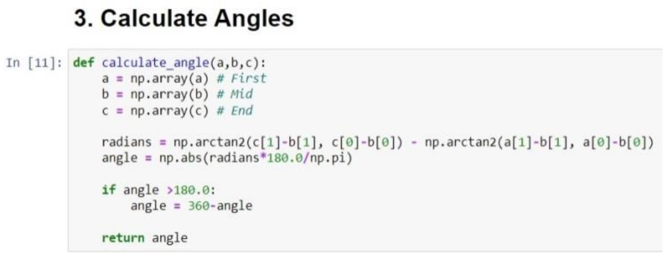
With BlazePose, we present a new topology of 33 human body keypoints, which is a superset of COCO, BlazeFace and BlazePalm topologies. This allows us to determine body semantics from pose prediction alone that is consistent with face and hand models.

B. An ML Pipeline for Pose Tracking

To measure the position, we use our two-step ML pipe certified detector-tracker. Using a detector, the pipeline begins by setting a profit interval (ROI) within the frame. The tracker predicts all 33 key points from this ROI. Note that in the case of video usage, the detector is used in the first draft only.

C. Person/pose Detection Model (BlazePose Detector)

The detector is inspired by our lightweight BlazeFace model, which is used in the MediaPipe Face Detection, as a personal detector. It clearly predicts two additional visual cues that strongly define the center of the human body, rotation and measurement as a circle. Inspired by Leonardo the Vitruvian man, we predict the central position of the human hip, the circumference of the human body, and the inclination angle of the shoulder joint and hip midpoints



3. Calculate Angles

```
In [11]: def calculate_angle(a,b,c):
             a = np.array(a) # First
             b = np.array(b) # Mid
             c = np.array(c) # End

             radians = np.arctan2(c[1]-b[1], c[0]-b[0]) - np.arctan2(a[1]-b[1], a[0]-b[0])
             angle = np.abs(radians*180.0/np.pi)

             if angle >180.0:
                 angle = 360-angle

             return angle
```

.

Fig. 3 Calculation of angles between the joints.

Therefore, we trained a face scanner, recommended by our sub-millisecond BlazeFace model, as a stand-alone detector. Note, this model only finds a person's location within a frame and cannot be used to identify individuals. In contrast to Face Mesh and MediaPipe Hand tracking pipelines, where we obtain ROI at predictable key points, by tracking a person's position we clearly predict two additional key points that strongly define the human body center, rotation and scale as a circle. Encouraged by Leonardo the Vitruvian man, we predict the central area of the human hip, the area around the peripheral, and the inclined angle of the line connecting the shoulder and hip midpoints. This results in consistent tracking even in the most complex cases.

IV. PROPOSED DESIGN

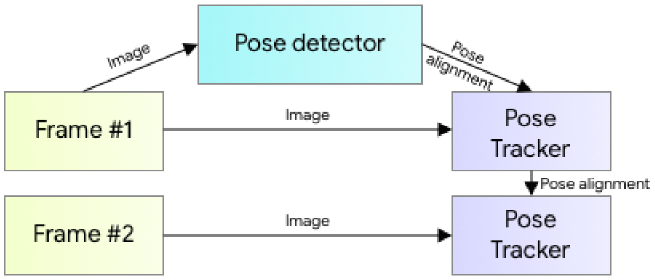


Fig. 4 Architecture of pose estimation model

Functionality Initially we use the code in the jupyter IDE. when the code is released, the camera starts working. It reads the links using the complete MEDIAPIPE and compares it to the body_language.pkl file previously defined in the previous training. Then it predicts a touch. Also, it predicts the accuracy of the position compared to the pre-trained touch using the numpy- argsmax () function imported from the Numpy library. In the model training mode, we apply the code and the webcam starts working. It reads links from our bodies using the complete work entered in the MediaPipe library and writes the coords.csv file. We name the class as the action verb that trains us. These links are then compiled using the flatten () function imported from the ML library for reading scikit.

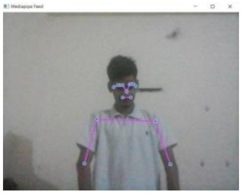
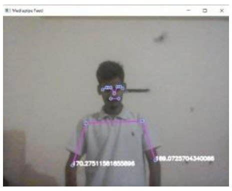
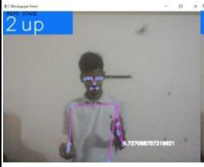
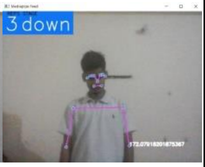
A. System Modules

(a) Install and import dependencies: In this step mediapipe, opencv, pandas and scikit-learn are installed and imported. (b) Make some detections: Here a webcam window is opened using cv2 and specifications like color and thickness of our face, hand and pose landmarks are mentioned/modified.(c) Capture landmarks and export to CSV: Here various coordinates of facial expressions and body gestures are captured and exported to a csv file named coords.csv. (d) Train custom model using scikit learn: Here the collected data is read and processed to train our machine learning classification model on it.The model is then evaluated and serialized. (e) Make detections with model: Now the code, when run, can make predictions of the user's gestures using the above trained

V. IMPLEMENTATION

Based on a person's posture, we can create a variety of applications, such as fitness or yoga trackers. For example, we introduce squats and push up figures, which can automatically calculate user numbers, or verify the quality of the exercises performed. Such usage conditions can be created using an additional partition network or even a simple dual-distance view algorithm, corresponding to a close position in a normal position.

Table 1. Real time analysis and detection results

Stages	Processing	Result
Stage 1	Detection of human skeleton body joints	
Stags 2	Joints angle detection and calculation	
Stage 3	Curl counter mechanism to evaluate number of reps.	  Mediapipe Feed REPS STAGE 3 down 172.07918201875367

VI. CONCLUSION AND FUTURE SCOPE

Finding and analyzing body language has received a lot of attention lately. Being able to see and analyse a client / customer facial expression helps businesses and advertising teams to get honest reviews and feedback. But facial expressions are a small part of body language. Body language contains other features such as gestures and posture. And body language plays a vital role in communication. For example in interviews, interviewees consider the body language of a person. By developing this project, a facilitator can be provided with a facilitator that helps them understand how the candidate responds to questions from different domains or is placed in different situations during the HR cycle. As this project supports real-time sign acquisition, sign-language acquisition can also be used. Not only that, through this project, the implementation of existing projects such as drowsiness detection, action detection etc. can be made easier with the best results.

VII. REFERENCES

[1] Human Body Pose Estimation and Applications. Published in: 2021 Innovations in Power and Advanced Computing Technologies (i-PACT).

[2] https://google.github.io/mediapipe/solutions/pose.html

[3] Human Pose Estimation from Single Image July(2018).Conference: 2018 International Conference on Audio, Language and Image Processing.

[4] Liangchen Song;Gang Yu;Junsong Yuan;Zicheng Liu; (2021). Human pose estimation and its application to action recognition

[5] Deep Learning-Based Human Pose Estimation: A Survey CE ZHENG∗ , University of Central Florida, USA WENHAN WU∗ , University of North Carolina at Charlotte, USA.

[6] DeepPose: Human Pose Estimation via Deep Neural Networks Alexander Toshev toshev@google.com Google Christian Szegedy szegedy@google.com Google

[7] AI-Based Yoga Pose Estimation for Android Application Girija Gireesh Chiddarwar, Abhishek Ranjane, Mugdha Chindhe, Rachana Deodhar, Palash Pune-411041,India

[8] Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. Real-time facial surface geometry from monocular video on mobile gpus. CoRR, abs/1907.06724, 2019. 1, 3

[9] Sven Kreiss, Lorenzo Bertoni, and Alexandre Alahi. Pifpaf: Composite fields for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 11977–11986, 2019.

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence ´ Zitnick. Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755. Springer, 2014.

[11] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In European conference on computer vision, pages 483–499. Springer, 2016.

[12] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5693–5703, 2019.